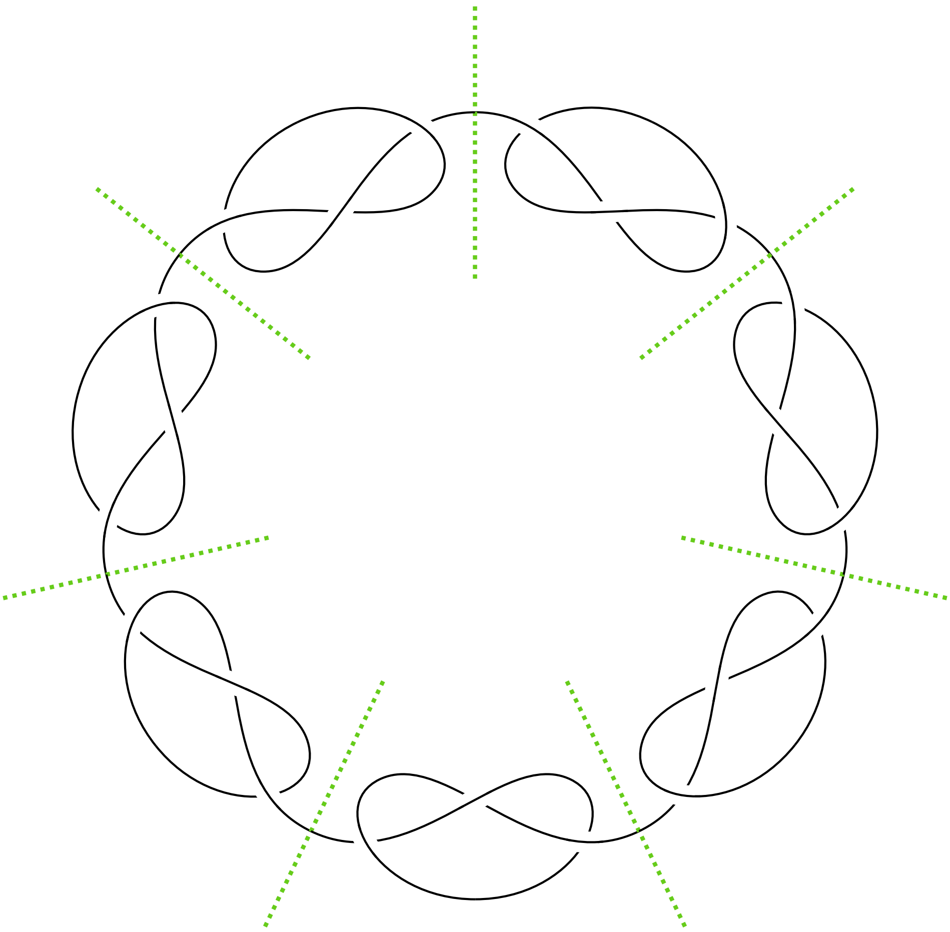
Connected sum decompositions of manifolds and submanifolds

Masters thesis in Mathematics
submitted by

Tobias Hirsch



supervised by

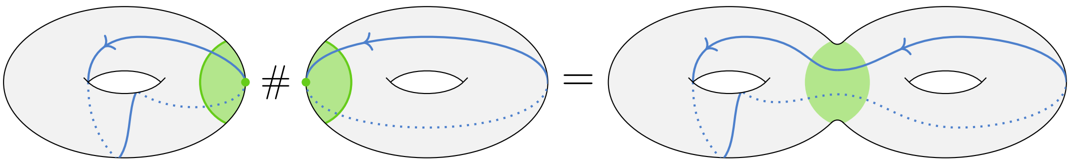
Prof. Stefan Friedl PhD

at the

University of Regensburg

Introduction

The connected sum operation is a fundamental tool in Geometric Topology. It takes two manifolds of the same dimension and assembles them into a new one by deleting a point in each and identifying the resulting ends. This provides a method for assembling complicated manifolds from simpler ones. Conversely, one can take a large manifold and attempt to decompose it into smaller – perhaps easier to understand – pieces.
In dimension 2, this leads to the classification of surfaces: A closed surface can be decomposed as a connected sum of tori or projective spaces, which cannot be decomposed any further. One might now wonder how this generalises to higher dimensions. Can every manifold still be written as a connected sum of manifolds that cannot be further decomposed, how unique is such a decomposition, and what are the indecomposable manifolds?
It is also possible to define a connected sum operation on manifolds with a distinguished submanifold – in particular knotted spheres in spheres – by performing the operation simultaneously on both the ambient manifold and the submanifold. Of course, the same questions then also arise in this setting. These questions will accompany us through this thesis.



connected sum of two manifolds with a distinguished submanifold

In the first chapter, we begin by precisely defining these connected sum operations, taking care to show how smooth structures arise. We then examine their effect on basic algebraic and geometric invariants like homology, cohomology the fundamental group and the universal covering.
The second chapter sees us develop a general algebraic framework within which to study our questions. This is given by monoids since the connected sum operations all define monoid structures on suitable sets. Here we fix a common terminology for this thesis adopting notions like '(unique) factorization', 'prime' and 'irreducible' from Ring Theory for our purposes. We further borrow ideas to build a toolkit for showing that a monoid allows factorization, mainly evolving around an analogue to a Euclidean ring. The key idea is that we can associate to an element of a monoid a 'complexity' in $\mathbb{N}_0$ that decreases when decomposing. As $\mathbb{N}_0$ is bounded below one can then not decompose forever and must reach irreducibles eventually.
As a test case, we apply our techniques to the monoids $\mathcal{A}$ – finitely generated abelian groups under direct sum – and $\mathcal{G}$ – finitely generated groups under free products. We show unique factorization in $\mathcal{A}$ and determine all the irreducibles, thereby recovering a proof of the Structure Theorem for finitely generated abelian groups. In $\mathcal{G}$ unique factorization is Grushko's Theorem. We briefly sketch Stallings' Theorem can be used to find some irreducibles, although this only gives a complete list in the torsion-free case.
We return to topology in chapter three to study factorization of manifolds. Using that homology and the fundamental group define monoid homomorphisms to $\mathcal{A}$ and $\mathcal{G}$, we use our results from chapter two to show existence of factorizations subject to a constraint arising from non-trivial homotopy spheres. We then prove that without restrictions on orientability, factorization is never unique. This is based on the observation that the twisted S^n-bundle over S^1 can be untwisted after a connected sum with a non-orientable manifold. For oriented manifolds, factorization is unique in dimension 2 and 3 – in dimension 2 this is the classification of surfaces, in dimension 3 it is the Kneser–Milnor Theorem. We conclude our study of manifolds by exhibiting some irreducible manifolds, most notably, pointing out that Riemannian manifolds of constant curvature are irreducible.

With the fourth chapter, we turn to knot theory. It is entirely dedicated to proving the existence of factorization – at least except for knots in S^4. The reason for this restriction is that in all other dimensions the unknotting conjecture is true: a knot is trivial if and only if its complement is homotopy equivalent to S^1 – which can be checked algebraically on the fundamental group and the Alexander modules. We therefore begin by introducing the Alexander modules of a knot. As they are not finitely generated as abelian groups, they do not define a monoid homomorphism to $\mathcal{A}$. Consequently, we must use their $\mathbb{Z}[t^{\pm 1}]$-module structure. Considering the rational and torsion parts separately we can find a suitable complexity on the $\mathbb{Z}[t^{\pm 1}]$-modules arising as Alexander modules. This crucially relies on the fact that the torsion subgroup of an Alexander module is always finite. It then remains to consider the fundamental group. Here our approach uses tools from Geometric Group Theory, namely, accessibility. We develop this in the setting of graphs of groups in the appendix.
In the final chapter, we discuss uniqueness of factorization of knots. In the classical dimension, factorization is unique by a theorem of H. Schubert. In higher odd dimensions this is no longer the case. We show this starting from the classification of simple knots in terms of their Seifert form. We translate this into a classification in terms of the Blanchfield form, where we can use Number Theory to find knots contradicting uniqueness. To disprove uniqueness in even dimensions, we spin this example to increase its dimension. This approach does of course again not work for knots in S^4 where the question of uniqueness remains open.
Throughout this thesis we work in the smooth category. The tools developed to show existence of factorizations – that is chapters two to four – are topological. Consequently, they work equally well in any other setting with a reasonably defined connected sum operation – for example, topological manifolds or locally flat knots. Some results may even be strengthened due to the different status of the Poincaré conjecture and the unknotting conjecture. The reason we did not work in all categories simultaneously is that this would greatly impair readability. Proving that the connected sum operation is well-defined in the topological setting also requires considerably more effort than in the smooth setting: for an account of the arising issues and how to overcome them compare [Fri25: Appendix 210.7] for topological manifolds and [Liv24] for locally flat knots.

The author is indebted to Prof. Stefan Friedl for continual support and guidance, not only while writing this thesis but throughout his formal study of Mathematics. He would like to thank Daniel Zach and Patrick Perras for valuable discussions. He is grateful to Madeleine Bauer for her support in the stressful phases of this project and help understanding [Hae61].

Table of Contents

1. The connected sum operation

1.1. The connected sum of manifolds

We fittingly begin by defining the connected sum of smooth manifolds:

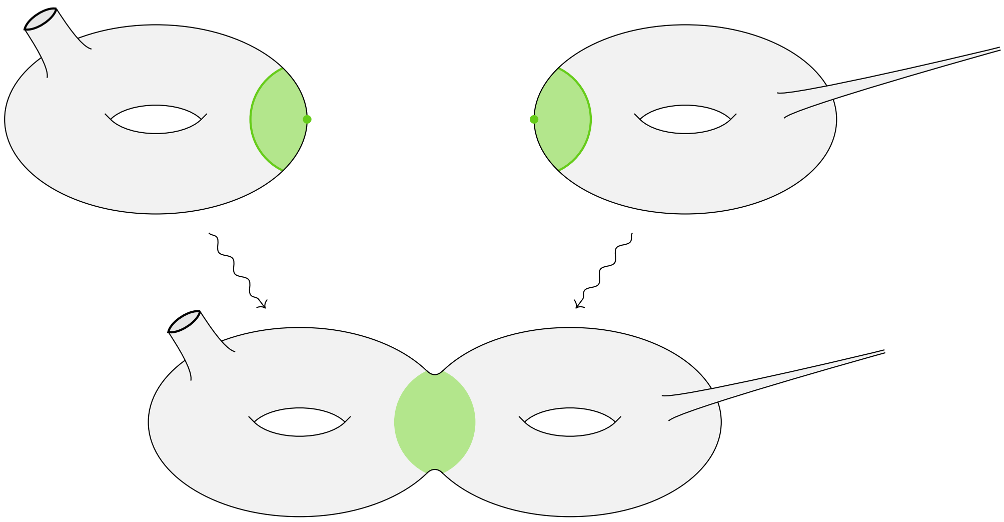
Definition. Let $n \geq 1$ and X, Y be non-empty connected n-dimensional smooth manifolds. Choose smooth embeddings $\varphi\colon \overline{B}^n \to X \setminus \partial X$, $\psi\colon \overline{B}^n \to Y \setminus \partial Y$. If X and Y are orientable, we assume that they are oriented and that precisely one of φ and ψ is orientation-preserving. The *connected sum of X and Y* is

$$X \# Y \coloneqq (X \setminus \varphi(0)) \sqcup (Y \setminus \psi(0)) /\!\sim$$

where $\sim$ is the equivalence relation generated by $\varphi(P) \sim \psi(\alpha(P))$ for all $P \in B^n \setminus \{0\}$ where

$$\begin{aligned} \alpha\colon B^n \setminus \{0\} &\to B^n \setminus \{0\} \\ P &\mapsto \frac{P}{\|P\|} - P \end{aligned}$$

is an orientation-reversing diffeomorphism.



The approach presented here is a bit different from that of [Fri25: Section 51.4]. It is closer to the approach of [Kos93: Section VI.1]. The advantage of this approach is that it makes the origin of the smooth structure readily apparent since it glues along a suitable open subset.The disadvantage is that, in practice, one would much rather deal with compact subsets as in the other approach. We will remedy this in Proposition 1.2 (4).
With a definition in place, the next question of course is whether the result depends on any choices. Here, there could be a dependency on the precise choice of smoothly embedded balls. The next theorem gives us the tools to show this is not the case:

Theorem 1.1. *Let M be a connected n-dimensional smooth manifold and $\varphi, \psi\colon \overline{B}^n \to M \setminus \partial M$ smooth embeddings. If M is orientable, we demand that it is oriented and φ and ψ are both orientation-preserving or both orientation-reversing. There exists a diffeotopy $F\colon M \times [0,1] \to M$ such that $F_0 = \mathrm{id}_M$ and $F_1 \circ \varphi = \psi$.*

Proof. see [Fri25: Theorem 42.10] ∎

With this we now prove some basic properties of the connected sum. In particular, well-definedness in the sense that it is independent of the choices made and again yields a smooth manifold:

Proposition 1.2. *Let $n \geq 1$ and X, Y be non-empty connected n-dimensional smooth manifolds. Choose smooth embeddings $\varphi\colon \overline{B}^n \to X$, $\psi\colon \overline{B}^n \to Y$. If X and Y are orientable, we assume that they are oriented and that precisely one of φ and ψ is orientation-preserving.*

(1) (a) *The connected sum $X\#Y$ is an n-dimensional topological manifold. It admits a unique smooth structure such that the inclusions $X \setminus \varphi(0) \hookrightarrow X\#Y$ and $Y \setminus \psi(0) \hookrightarrow X\#Y$ are smooth embeddings.*

(b) *The diffeomorphism class of $X\#Y$ only depends on the diffeomorphism classes of X and Y. In particular, it does not depend on the choice of φ and ψ and which of the two is orientation-preserving.*

(2) *If X and Y are oriented, $X\#Y$ admits a unique orientation such that the inclusions $X \setminus \varphi(0) \to X\#Y$ and $Y \setminus \psi(0) \to X\#Y$ are orientation-preserving. Otherwise, $X\#Y$ is non-orientable.*

(3) *We have $\partial(X\#Y) = \partial X \sqcup \partial Y$.*

(4) *Consider $\hat{X} \coloneqq X \setminus \varphi(B^n_{\frac{1}{2}})$ and $\hat{Y} \coloneqq Y \setminus \psi(B^n_{\frac{1}{2}})$. We have pushouts of topological spaces*

$$\begin{tikzcd} S^{n-1}_{\frac{1}{2}} \arrow[r, "\varphi"] \arrow[d, "\psi"'] & \hat{X} \arrow[d] \\ \hat{Y} \arrow[r] & X\#Y \end{tikzcd} \qquad \begin{tikzcd} S^{n-1}_{\frac{1}{2}} \arrow[r, "\varphi"] \arrow[d] & \hat{X} \arrow[d] \\ \overline{B}^n_{\frac{1}{2}} \arrow[r] & X \end{tikzcd} \qquad \begin{tikzcd} S^{n-1}_{\frac{1}{2}} \arrow[r, "\psi"] \arrow[d] & \hat{Y} \arrow[d] \\ \overline{B}^n_{\frac{1}{2}} \arrow[r] & Y \end{tikzcd}$$

(5) (a) *If X and Y are compact, so is $X\#Y$.*

(b) *If $n \geq 2$ or at least one of X and Y is closed, $X\#Y$ is also connected.*

Proof.

(1) (a) The two inclusions $X \setminus \varphi(0) \hookrightarrow X\#Y$ and $Y \setminus \psi(0) \hookrightarrow X\#Y$ are embeddings since they are open injections. We use this to view $X \setminus \varphi(0)$ and $Y \setminus \psi(0)$ as subspaces of $X\#Y$.
The space $X\#Y$ is Hausdorff since the equivalence relation is closed and the quotient projection is open. It is second-countable, since it is the union of $X \setminus \varphi(0)$ and $Y \setminus \psi(0)$ which are both second-countable.
The union of smooth atlases for $X \setminus \varphi(0)$ and $Y \setminus \psi(0)$ defines an atlas for $X\#Y$. This atlas is smooth, since the map

$$\psi \circ \alpha \circ \varphi^{-1}\colon \varphi(B^n \setminus \{0\}) \to \psi(B^n \setminus \{0\})$$

along which we glued the manifolds together is a diffeomorphism. By construction the inclusions are now smooth embeddings and the resulting smooth structure is the unique smooth structure in which the inclusions are smooth.

(b) To cleanly write down this argument, we upgrade our notation temporarily to take the smooth embeddings $\varphi\colon \overline{B}^n \to X \setminus \partial X$, $\psi\colon \overline{B}^n \to Y \setminus \partial Y$ into account:

$$X\#Y(\varphi, \psi) \coloneqq (X \setminus \varphi(0)) \sqcup (Y \setminus \psi(0)) \,/ \varphi(P) \sim \psi(\alpha(P)) \text{ for } P \in B^n \setminus \{0\}$$

Now let $\varphi'\colon \overline{B}^n \to X \setminus \partial X, \psi'\colon \overline{B}^n \to Y \setminus \partial Y$ also be smooth embeddings.
We consider the following cases separately:

case 1: X, Y oriented
If φ and φ' (and therefore also ψ and ψ') have the same orientability, it follows essentially immediately from Theorem 1.1 that $X\#Y(\varphi, \psi) \cong X\#Y(\varphi', \psi')$. So assume now this is not the case. Let $\rho\colon \overline{B}^n \to \overline{B}^n$ be a reflection in some hyperplane of $\mathbb{R}^n$. Then ρ is orientation-reversing implying that φ and $\varphi' \circ \rho$ have the same

orientability. The above then implies that $X\#Y(\varphi,\psi) \cong X\#Y(\varphi'\circ\rho,\psi'\circ\rho)$. But $X\#Y(\varphi'\circ\rho,\psi'\circ\rho) = X\#Y(\varphi',\psi')$ since

$$\psi\circ\alpha\circ\varphi^{-1} = \psi\circ\rho\circ\alpha\circ\rho^{-1}\circ\varphi^{-1}\colon \varphi(B^n\setminus\{0\}) \to \psi(B^n\setminus\{0\})$$

i.e. the equivalence relations defining them are literally the same.

case 2: X, Y non-orientable
Then $X\#Y(\varphi,\psi) \cong X\#Y(\varphi',\psi')$ follows essentially immediately from Theorem 1.1.

case 3: X orientable, Y non-orientable
If φ and φ' have the same orientability, the $X\#Y(\varphi,\psi) \cong X\#Y(\varphi',\psi')$ follows as usual from Theorem 1.1. Now suppose this is not the case. Then as in case 1 $X\#Y(\varphi,\psi) \cong X\#Y(\varphi'\circ\rho,\psi'\circ\rho) = X\#Y(\varphi',\psi')$.

case 4: Y orientable, X non-orientable
This case is analogous to case 3.

(2) Suppose X and Y are oriented. The diffeomorphism $\alpha\colon B^n\setminus\{0\}\to B^n\setminus\{0\}$ is orientation-reversing. Hence,

$$\psi\circ\alpha\circ\varphi^{-1}\colon \varphi(B^n\setminus\{0\}) \to \psi(B^n\setminus\{0\})$$

is an orientation-preserving diffeomorphism. It follows that all transition maps in the smooth atlas for $X\#Y$ are orientation-preserving, i.e. $X\#Y$ is orientable in the claimed way.
For the converse note that if $X\#Y$ is orientable, $X\setminus\varphi(0)$ and $Y\setminus\psi(0)$ also need to be orientable as they are codimension 0 submanifolds. Then X and Y are orientable, too. This follows as an orientation on $B^n\setminus\{0\}$ always extends to one on B^n.
(3) Follows directly from the definition of the atlas on $X\#Y$.
(4) By (1) we have smooth embeddings $X\setminus\varphi(0)\hookrightarrow X\#Y$ and $Y\setminus\psi(0)\hookrightarrow X\#Y$. Identifying the various spaces with their images under them, we have $\hat{X}\cap\hat{Y} = S^{n-1}_{\frac{1}{2}}$, so the first pushout follows from [Fri25: Lemma 6.37]. The other two can be proven similarly.
(5) (a) Follows directly from (4).
(b) If $n\geq 2$, $\mathbb{R}^n\setminus\{0\}$ is connected. Then $X\setminus\varphi(0)$ and $Y\setminus\psi(0)$ are connected implying the claim.
If $n=1$ and X or Y is closed the claim follows by the classification of 1-dimensional manifolds. ■

In the next proposition we study the effect of the connected sum operation on the usual algebraic invariants. Properly considering orientability makes it a bit unwieldy.

Proposition 1.3. *Let X, Y be non-empty connected closed n-dimensional smooth manifolds.*
(1) *If $n\geq 3$, $\pi_1(X\#Y)\cong\pi_1(X)*\pi_1(Y)$.*
(2) *Let R be a commutative ring. For $k\in\{1,\ldots,n-2\}$, $\mathrm{H}_k(X\#Y;R)\cong\mathrm{H}_k(X;R)\oplus\mathrm{H}_k(Y;R)$. If X or Y is R-oriented, this also holds for $k=n-1$. Otherwise we have a short exact sequence*

$$0\to R\to \mathrm{H}_{n-1}(X\#Y;R)\to\mathrm{H}_{n-1}(X;R)\oplus\mathrm{H}_{n-1}(Y;R)\to 0$$

(3) *Let R be a commutative ring. There is a ring homomorphism $\mathrm{H}^*(X\vee Y;R)\to\mathrm{H}^*(X\#Y;R)$ with the following properties:*
- *If X and Y are R-oriented, it is surjective and its kernel is the ideal generated by $[X]^*-[Y]^*$. The dual fundamental class $[X\#Y]^*$ is the image of $[X]^*$ and $[Y]^*$.*
- *If X is R-oriented but Y is not R-orientable, it is surjective and its kernel is the ideal generated by $[X]^*$, and conversely for X and Y switched.*
- *If both X and Y are not R-orientable, it is injective. It is surjective on H^k unless $k=n-1$ where we have a short exact sequence*

$$0\to\mathrm{H}^{n-1}(X\vee Y;R)\to\mathrm{H}^{n-1}(X\#Y;R)\to R\to 0$$

Proof. As in Proposition 1.2 (4) we construct pushouts

$$\begin{tikzcd} S^{n-1} \arrow[r] \arrow[d] & \hat{X} \arrow[d] \\ \hat{Y} \arrow[r] & X\#Y \end{tikzcd} \qquad \begin{tikzcd} S^{n-1} \arrow[r] \arrow[d] & \hat{X} \arrow[d] \\ \overline{B}^n \arrow[r] & X \end{tikzcd} \qquad \begin{tikzcd} S^{n-1} \arrow[r] \arrow[d] & \hat{Y} \arrow[d] \\ \overline{B}^n \arrow[r] & Y \end{tikzcd}$$

(1) S^{n-1} is simply connected for $n \geq 3$. By the Seifert–van Kampen Theorem it therefore follows from the last two pushouts that $\pi_1(\hat{X}) \cong \pi_1(X)$, $\pi_1(\hat{Y}) \cong \pi_1(Y)$. The first pushout then similarly implies the claim.

(2) Consider $S^{n-1} \hookrightarrow X\#Y$ as in the above pushout. Collapsing this S^{n-1} gives a quotient $X\#Y \to X \vee Y$. Hence, there is a long exact sequence

$$\mathrm{H}_k(S^{n-1};R) \to \mathrm{H}_k(X\#Y;R) \to \underbrace{\mathrm{H}_k(X\#Y,S^{n-1};R)}_{\cong \mathrm{H}_k(X;R)\oplus \mathrm{H}_k(Y;R)} \to \mathrm{H}_{k-1}(S^{n-1};R)$$

For $k \in \{1,\dots,n-2\}$ this directly gives the claim.
For $k = n-1$ first assume X is R-orientable. We have a commutative diagram

$$\begin{tikzcd} & \mathrm{H}_{n-1}(\hat{X}) \arrow[d] \\ \mathrm{H}_{n-1}(S^{n-1};R) \arrow[ur, "=0"] \arrow[r] & \mathrm{H}_{n-1}(X\#Y;R) \end{tikzcd}$$

The diagonal map is the inclusion of the boundary of a compact orientable manifold, hence it is trivial by [Fri25: Proposition 171.18]. It follows that the bottom map is also trivial, proving the claim in this case. The case where Y is orientable is of course analogous.
It only remains to consider the case where both X and Y are not R-orientable. Here,

$$\mathrm{H}_n(X\#Y,S^{n-1};R) \cong \mathrm{H}_n(X;R) \oplus \mathrm{H}_n(Y;R) = 0$$

finishing the proof by again considering the exact sequence.

(3) As in (2) we consider the quotient $S^{n-1} \hookrightarrow X\#Y \twoheadrightarrow X \vee Y$. The quotient map induces the ring homomorphism we want to consider. It is clearly bijective on degree 0. For $k \geq 1$ consider the long exact sequence

$$\mathrm{H}^{k-1}(S^{n-1}) \to \mathrm{H}^k(X \vee Y) \to \mathrm{H}^k(X\#Y) \to \mathrm{H}^k(S^{n-1})$$

which proves bijectivity unless $k = n-1, n$. For these two cases consider

$$0 \to \mathrm{H}^{n-1}(X \vee Y) \to \mathrm{H}^{n-1}(X\#Y) \xrightarrow{\Phi} \mathrm{H}^{n-1}(S^{n-1}) \xrightarrow{\Psi} \mathrm{H}^n(X \vee Y) \to \mathrm{H}^n(X\#Y) \to 0$$

First assume X and Y are R-oriented. We get induced R-orientations on $\hat{X}$ and $\hat{Y}$. Wlog. we can assume that the diffeomorphism $S^{n-1} \to \partial\hat{X}$ is R-orientation-preserving and the diffeomorphism $S^{n-1} \to \partial\hat{Y}$ is R-orientation-reversing. As in (2) we see that Φ is trivial, showing bijectivity for $k = n-1$. It remains to determine the image of Ψ. We have a commutative diagram

$$\begin{tikzcd} \mathrm{H}^{n-1}(\partial Y;R) = R\cdot[\partial\hat{Y}]^* \arrow[r, "(1)"] \arrow[d, "(-1)"'] & \mathrm{H}^{n-1}(Y,\partial Y;R) = R\cdot[Y,\partial Y]^* \arrow[r, "(1)"] & \mathrm{H}^n(Y;R) = R\cdot[Y]^* \arrow[d] \\ \mathrm{H}^{n-1}(S^{n-1};R) = R\cdot[S^{n-1}]^* \arrow[rr] & & \mathrm{H}^n(X \vee Y;R) \\ \mathrm{H}^{n-1}(\partial X;R) = R\cdot[\partial\hat{X}]^* \arrow[r, "(1)"] \arrow[u, "(1)"] & \mathrm{H}^{n-1}(X,\partial X;R) = R\cdot[X,\partial X]^* \arrow[r, "(1)"] & \mathrm{H}^n(X;R) = R\cdot[X]^* \arrow[u] \end{tikzcd}$$

The decorations on the maps are the matrices representing them in the given bases. This proves that $\operatorname{im}(\Phi) = ([X]^* - [Y]^*)$.
The cases where precisely one of X and Y is R-oriented is dealt with similarly. In the case where X and Y are not R-orientable, Ψ and all groups to the right of it are trivial, completing the proof. ■

We have seen in Proposition 1.2 that the connected sum of oriented manifolds is again a naturally oriented manifold. It might seem reasonable to expect that we can just drop all orientability requirements, with no cost but the resulting manifold no longer having a natural orientation. However, this is not the case:

Example 1.4. Choose an orientation on $\mathbb{CP}^2$ and let $\overline{\mathbb{CP}}^2$ denote the manifold with the opposite orientation. Let $x \in \mathrm{H}^2(\mathbb{CP}^2)$ be a generator. Wlog. we have $x^2 = [\mathbb{CP}^2]^* = -[\overline{\mathbb{CP}}^2]^*$ in $\mathrm{H}^*(\mathbb{CP}^2) = \mathrm{H}^*(\overline{\mathbb{CP}}^2)$. Applying Proposition 1.3 we get bases $\mathrm{H}^2(\mathbb{CP}^2\#\mathbb{CP}^2) = \mathbb{Z}\cdot\{a,b\}$, $\mathrm{H}^2(\mathbb{CP}^2\#\overline{\mathbb{CP}}^2) = \mathbb{Z}\cdot\{u,v\}$ such that

$$\begin{aligned} a\cup b = b\cup a &= 0 \qquad & a^2 &= b^2 \\ u\cup v = v\cup u &= 0 \qquad & u^2 &= -b^2 \end{aligned}$$

One can now directly compute that these ring structures are non-isomorphic – or more cleverly observe that this implies that $\mathbb{CP}^2\#\mathbb{CP}^2$ has signature 2 while $\mathbb{CP}^2\#\overline{\mathbb{CP}}^2$ has signature 0, see [Fri25: Proposition 212.18]. Hence, $\mathbb{CP}^2\#\mathbb{CP}^2$ and $\mathbb{CP}^2\#\overline{\mathbb{CP}}^2$ are not even homotopy equivalent, let alone diffeomorphic.

At a later point it will be useful to know the universal covering of a connected sum. Since our description will not be very concrete anyway, we do not bother to establish its natural smooth structure – although this can surely be done by essentially the same approach if one starts with the original definition given above. In the following, we will therefore only use the topological description:[1]

Construction 1.5. Let $n \geq 3$ and M be a non-empty connected n-dimensional smooth manifold and suppose $M = X\#Y$ for suitable X, Y. Let $\hat{X}$, $\hat{Y}$ be X, Y with the interior of a closed ball removed. By Proposition 1.2 (4) we have a pushout

$$\begin{tikzcd} S^{n-1} \arrow[r] \arrow[d] & \hat{X} \arrow[d] \\ \hat{Y} \arrow[r] & M \end{tikzcd}$$

where the top and left map are the inclusion of the respective boundaries. Fix a basepoint $x \in S^{n-1} \hookrightarrow M$ and let π_X, π_Y and π_M be the corresponding fundamental groups of $\hat{X}, \hat{Y}$ and M. By the Seifert–van Kampen Theorem $\pi_M = \pi_X * \pi_Y$, in particular, $\pi_X, \pi_Y \subseteq \pi_M$ are subgroups. Now let $\widetilde{X}, \widetilde{Y}$ and $\widetilde{M}$ be the universal coverings of $\hat{X}, \hat{Y}$ and M.
Before we continue this discussion, we draw the attention of the reader to the fairly technical [Fri25: Propositions 113.26, 113.27, 115.7] which we will use readily in the following. Combined these statements will provide all non-trivial arguments made.
There are embeddings $S^{n-1} \to \widetilde{X}$, $S^{n-1} \to \widetilde{Y}$ lying over $S^{n-1} \to X$, $S^{n-1} \to Y$. Then there are embeddings $\widetilde{X} \to \widetilde{M}$, $\widetilde{Y} \to \widetilde{M}$ lying over $\hat{X} \to M$, $\hat{Y} \to M$ that agree on S^{n-1}. We once and for all fix all these and now view $S^{n-1}, \widetilde{X}, \widetilde{Y} \subseteq \widetilde{M}$ as subspaces.
Consider the action of π_M on $\widetilde{M}$ by deck transformations. Letting it act on the subspace $\widetilde{X}$ cycles through the different lifts we could have chosen. Note that $g\widetilde{X} = g'\widetilde{X}$ if and only if g and g' are in the same coset of π_X in π_M and otherwise $g\widetilde{X} \cap g'\widetilde{X} = \varnothing$. The same of course also holds for $\widetilde{Y}$. Since S^{n-1} is simply connected, the $gS^{n-1}, g \in \pi_M$ are pairwise disjoint.

[1]The following is morally a special case of [Cap76: pp.80–82].

For $g \in \pi_M$ we have inclusions $gS^{n-1} \subseteq g\widetilde{X}$ and $gS^{n-1} \subseteq g\widetilde{Y}$. Taken together we get a commutative diagram

$$\begin{array}{ccc} \coprod_{g\in\pi_M} gS^{n-1} & \longrightarrow & \coprod_{g\in\pi_M/\pi_X} g\widetilde{X} \\ \downarrow & & \downarrow \\ \coprod_{g\in\pi_M/\pi_Y} g\widetilde{Y} & \longrightarrow & \widetilde{M} \end{array}$$

It follows from [Fri25: Lemma 6.37] that this is a pushout diagram.

We ignored the 2-dimensional case in the above, since S^1 not being simply connected would have complicated things unnecessarily. Note that this is not a big loss, since there are only two possible universal coverings of surfaces anyway:

Theorem 1.6. *Let M be a non-empty connected 2-dimensional smooth manifold without boundary and $\mathrm{H}_1(M) = 0$. Then M is diffeomorphic to S^2 or $\mathbb{R}^2$.*

Proof. If M is compact, it is diffeomorphic to S^2 by the classification of surfaces. So we only need to consider the case where M is non-compact.
By [Fri25: Proposition 35.2] we can let $\Sigma_1, \Sigma_2, \Sigma_3, \ldots$ be a compact exhaustion of M, i.e. the following hold for all $i \geq 1$

- Σ_i is compact and connected
- $\Sigma_i \subseteq M$ is a 2-dimensional smooth submanifold of M
- $\Sigma_i \subseteq \mathring{\Sigma}_{i+1}$
- $M = \bigcup_{i\geq 1} \mathring{M}_i$

The subset $C_i := M \setminus \mathring{\Sigma}_i$ is a 2-dimensional smooth submanifold and $\partial C_i = \partial\Sigma_i = \Sigma_i \cap C_i$ (see [Fri25: Proposition 22.48]). By the Mayer–Vietoris Theorem applied to $M = C_i \cup \Sigma_i$ we have an exact sequence

$$\underbrace{\mathrm{H}_2(\partial\Sigma_i)}_{=0} \to \mathrm{H}_2(\Sigma_i) \oplus \mathrm{H}_2(C_i) \to \underbrace{\mathrm{H}_2(M)}_{=0}$$

so $\mathrm{H}_2(\Sigma_i) = \mathrm{H}_2(C_i) = 0$ which implies that Σ_i and C_i have no closed components. If C_i has compact components, they therefore must have boundary. We can then add those compact components to Σ_i obtaining a new compact connected 2-dimensional smooth submanifold Σ_i'. By compactness, there exists $i > j$ such that $\Sigma_i' \subseteq \Sigma_j$. We can replace Σ_i by Σ_i' and drop $\Sigma_i, \ldots, \Sigma_{j-1}$ to obtain a new compact exhaustion of M. After doing this inductively for all $i \in \mathbb{N}$, we may assume that C_i never has compact components.
The top row in the following diagram is again the Mayer–Vietoris sequence of $M = C_i \cup \Sigma_i$. The vertical arrow is the natural projection. The diagonal arrow is induced by the boundary inclusion. It is a monomorphism by the long exact sequence of the pair $(C_i, \partial C_i)$, since every component of C_i is non-compact.

$$\begin{tikzcd} \underbrace{\mathrm{H}_2(M)}_{=0} \arrow[r] & \mathrm{H}_1(\partial C_i) \arrow[r] \arrow[dr, hook] & \mathrm{H}_1(\Sigma_i) \oplus \mathrm{H}_1(C_i) \arrow[r] \arrow[d, two heads] & \underbrace{\mathrm{H}_1(M)}_{=0} \\ & & \mathrm{H}_1(C_i) & \end{tikzcd}$$

It follows from the diagram that the vertical and diagonal arrows are in fact isomorphisms and therefore $\mathrm{H}_1(\Sigma_i) = 0$. By the classification of surfaces Σ_i is diffeomorphic to a disc. Since two smooth discs in $\overline{B}^2$ are related by a self-diffeomorphism of $\overline{B}^2$ (see [Fri25: Proposition 37.3]), there exist diffeomorphisms

$$\varphi_i \colon \Sigma_i \to \{x \in \mathbb{R}^2 \mid |x| \leq i\}$$

such that $\varphi_{i+1}|_{\Sigma_i} = \varphi_i$. Taken together they give a diffeomorphism from M to $\mathbb{R}^2$. ■

1.2. The connected sum of submanifolds

Next we want to generalize the connected sum to manifolds with a submanifold, performing the operation at the same time on the surrounding manifold and the submanifold:

Definition. Let $0 \leq k \leq n$. A (n,k)-*dimensional smooth manifold pair* is a pair (X, A) where X is an n-dimensional smooth manifold and $A \subseteq X$ is a k-dimensional proper smooth submanifold. Such a pair is *non-empty, connected, closed, oriented* or *orientable* if X and A are non-empty, connected, closed, oriented or orientable manifolds.
A map of pairs $\varphi\colon (X, A) \to (Y, B)$ is an *embedding* (resp. *diffeomorphism*) if $\varphi\colon X \to Y$ and $\varphi\colon A \to B$ are *embeddings* (resp. *diffeomorphisms*) and $\varphi(A) = \varphi(X) \cap B$.
If X and Y are oriented, an embedding $\varphi\colon (X, A) \to (Y, B)$ is *orientation-preserving* (resp. *orientation-reversing*) if $\varphi\colon X \to Y$ is orientation-preserving (resp. orientation-reversing) and, should A and B also be oriented, the restriction $\varphi\colon A \to B$ also is orientation-preserving (resp. orientation-reversing).

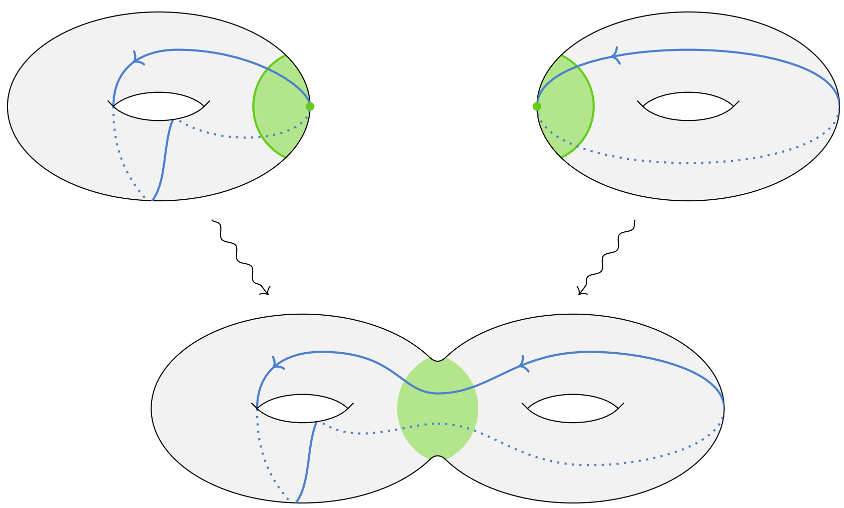
Definition. Let $1 \leq k < n$ and $(X, A), (Y, B)$ be non-empty connected (n, k)-dimensional smooth manifold pairs. Assume that X and Y are oriented. If A and B are orientable, assume they are also oriented. Let $\varphi\colon (\overline{B}^n, \overline{B}^k) \hookrightarrow (X \setminus \partial X, A \setminus \partial A), \psi\colon (\overline{B}^n, \overline{B}^k) \hookrightarrow (Y \setminus \partial Y, B \setminus \partial B)$ be smooth embeddings such that one of them is orientation-preserving and the other orientation-reversing. The *connected sum of* (X, A) *and* (Y, B) is

$$(X, A)\#(Y, B) \coloneqq ((X \setminus \varphi(0) \sqcup Y \setminus \psi(0))/{\sim}, (A \setminus \varphi(0) \sqcup B \setminus \psi(0))/{\sim})$$

where $\sim$ is the equivalence relation generated by $\varphi(P) \sim \psi(\alpha(P))$ for all $P \in B^n \setminus \{0\}$ where

$$\begin{aligned} \alpha\colon (B^n \setminus \{0\}, B^k \setminus \{0\}) &\to (B^n \setminus \{0\}, B^k \setminus \{0\}) \\ P &\mapsto \frac{P}{\|P\|} - P \end{aligned}$$

is an orientation-reversing diffeomorphism.



To show that this operation is well-defined, we need an analogue to Theorem 1.1:

Theorem 1.7. *Let $1 \leq k \leq n$ and (X, A) be a non-empty connected (n, k)-dimensional smooth manifold pair where X is oriented. If A is orientable, we assume it is also oriented. Let $\varphi, \psi\colon (\overline{B}^n, \overline{B}^k) \to (X \setminus \partial X, A \setminus \partial A)$ be smooth embeddings that are both orientation-preserving or both orientation-reversing. There exists a diffeotopy $F\colon X \times [0, 1] \to X$ such that $F_0 = \mathrm{id}_X$, $F_1 \circ \varphi = \psi$ and $F_t \coloneqq (X, A) \to (X, A)$ is a diffeomorphism of pairs for all $t \in [0, 1]$.*

Sketch of a proof: In principle this theorem is proven the same way as Theorem 1.1: One first proves that the smooth embeddings $\varphi, \psi\colon (\overline{B}^n, \overline{B}^k) \to (X \setminus \partial X, A \setminus \partial A)$ are smoothly isotopic through maps of pairs $(\overline{B}^n, \overline{B}^k) \to (X \setminus \partial X, A \setminus \partial A)$. The proof for this is not much more complicated than showing the analogous statements for smooth embeddings of balls $\overline{B}^n \to X$, see [Fri25: Theorem 42.15].
In the second step, one applies the Isotopy Extension Theorem to obtain the desired diffeotopy. For ordinary connected sum, one can just use the standard Isotopy Extension Theorem – in the present setting, it does however not necessarily produce a diffeotopy that respects the submanifold, so one needs to prove a more general version:

Claim. *Let (M, V) and (N, W) be smooth manifold pairs. Let $F\colon M \times [0,1] \to N \setminus \partial N$ be a smooth isotopy such that $F_t(V) \subseteq W$ for all $t \in [0,1]$. There exists a diffeotopy $G\colon N \times [0,1] \to N$ such that*

- $G_0 = \mathrm{id}$
- $G_t \circ F_0 = F_t$ *for all* $t \in [0,1]$
- $G_t(W) = W$ *for all* $t \in [0,1]$

Again, a proof of this can be obtained by suitably modifying the proof of the standard Isotopy Extension Theorem. One has to construct the relevant vector field to always be tangential to W – but this is not a big problem. [Fri25: Exercise 42.6] claims the same with more authority. ■

With this theorem the proof of well-definedness of the connected sum of manifolds immediately transfers to submanifolds:

Proposition 1.8. *Let $1 \leq k < n$ and (X, A), (Y, B) be non-empty connected (n, k)-dimensional smooth manifold pairs. Assume that X and Y are oriented. If A and B are orientable, assume they are also oriented.*

(1) *There are smooth embeddings $\varphi\colon (\overline{B}^n, \overline{B}^k) \hookrightarrow (X \setminus \partial X, A \setminus \partial A)$, $\psi\colon (\overline{B}^n, \overline{B}^k) \hookrightarrow (Y \setminus \partial Y, B \setminus \partial B)$ such that one of them is orientation-preserving and the other orientation-reversing.*
(2) *The connected sum $(X, A)\#(Y, B)$ exhibits $A\#B$ as a smooth submanifold of $X\#Y$, i.e. $(X, A)\#(Y, B)$ is a (n, k)-dimensional smooth manifold pair.*
(3) *The diffeomorphism type of $(X, A)\#(Y, B)$ only depends on the diffeomorphism types of (X, A) and (Y, B). In particular, it does not depend on the choice of φ and ψ and which of the two is orientation-preserving.*

Proof.
(1) Since $k < n$ this follows easily from the definition of a submanifold.
(2) The argument from Proposition 1.2 (1) about taking the union of atlases on $X \setminus \varphi(0)$ and $Y \setminus \psi(0)$ also works for finding the necessary submanifold charts.
(3) The proof of Proposition 1.2 (1) still applies after replacing each reference to Theorem 1.1 by Theorem 1.7. ■

This proof was a further reason why we chose to define the connected sum by gluing along open subsets. With this definition the argument was straightforward. If one had defined the connected sum by gluing along boundaries, it would certainly still be doable to prove this claim – but it would be much more of a technical nuisance than we are prepared to deal with.
Next, we want to determine how algebraic invariants behave under this connected sum operation. Of course, to extract such invariants from a submanifold, we have to turn to its complement. We begin with the following observation which mostly allows us to ignore basepoints:

Lemma 1.9. *Let $0 \leq k \leq n$ and (X, A) be a non-empty connected (n, k)-dimensional smooth manifold pair. If $k \leq n-2$, $X \setminus A$ is connected.*

Proof. see [Fri25: Corollary 47.11]. The argument is essentially the following:
Two points $P, Q \in X \setminus A$ are connected by a smooth path $\gamma\colon [0,1] \to X$ which can be made transversal to A. But since $\dim([0,1]) + \dim(A) < \dim(X)$, transversality here means disjoint. ■

It is now straightforward to describe how the complement changes under the connected sum operation and extract the change on algebraic invariants:

Proposition 1.10. *Let $0 \leq k \leq n$ and (X, A), (Y, B) be non-empty connected (n, k)-dimensional smooth manifold pairs. There exists a pushout*

$$\begin{tikzcd} B^n \setminus B^k \arrow[r] \arrow[d] & X \setminus A \arrow[d] \\ Y \setminus B \arrow[r] & X\#Y \setminus A\#B \end{tikzcd}$$

where the top and right map are induced by the embeddings used to define the connected sum.

(1) *If $k \leq n-2$,*

$$\pi_1(X\#Y \setminus A\#B) \cong \pi_1(X \setminus A) *_{\pi_1(B^n \setminus B^k)} \pi_1(Y \setminus B)$$

(2) *For $i \neq 0, n-k-1, n-k$,*

$$\mathrm{H}_i(X\#Y \setminus A\#B) \cong \mathrm{H}_i(X \setminus A) \oplus \mathrm{H}_i(Y \setminus B)$$

Proof. The pushout follows essentially immediately from the construction of the connected sum and [Fri25: Lemma 6.37].

(1) The spaces under consideration are connected by Lemma 1.9. Therefore we just need to apply the Seifert–van Kampen Theorem to the pushout.

(2) Applying the Mayer–Vietoris Theorem the decomposition of $X\#Y \setminus A\#B$ given by the pushout gives a long exact sequence

$$\mathrm{H}_i(B^n \setminus B^k) \to \mathrm{H}_i(X \setminus A) \oplus \mathrm{H}_i(Y \setminus B) \to \mathrm{H}_i(X\#Y \setminus A\#B) \to \mathrm{H}_{i-1}(B^n \setminus B^k)$$

Since $B^n \setminus B^k \simeq S^{n-k-1}$, this directly proves the claim. ∎

What happens to homology in degrees $n-k-1$ and $n-k$ is less clear. We will not investigate this further as we are most interested in the situation where the surrounding manifold is a sphere – in which case homology does not tell us anything new by Alexander duality:

Proposition 1.11. *Let $0 \leq k < n$ and (S^n, A) be a non-empty compact (n, k)-dimensional smooth manifold pair. For $i \geq 0$ there exists a natural isomorphism*

$$\mathrm{AD}_A \colon \tilde{\mathrm{H}}_i(S^n \setminus A) \cong \tilde{\mathrm{H}}^{n-i-1}(A)$$

Proof. see [Fri25: Theorem 204.7]. ∎

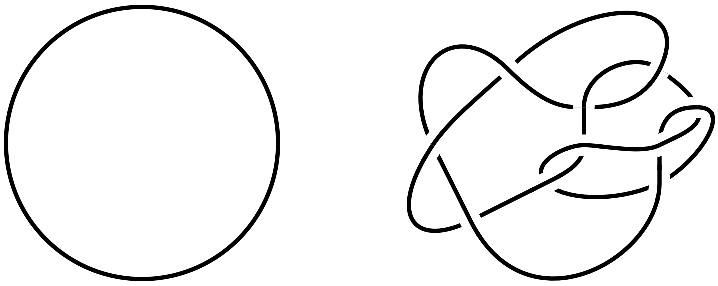
1.3. The connected sum of knots

In this section we specialize to the situation we are most interested in: codimension 2 submanifolds of spheres, i.e. knots:

Definition. Let $n \geq 1$. An *n-dimensional knot* is a smooth submanifold $K \subseteq S^{n+2}$ diffeomorphic to S^n.
Two n-dimensional knots $K, L \subseteq S^{n+2}$ are *smoothly isotopic* if there exists a *smooth isotopy* between them, i.e. a smooth map $F \colon K \times [0,1] \to S^{n+2}$ such that

- $F_0 \colon K \times \{0\} \to K$ and $F_1 \colon K \times \{1\} \to L$ are diffeomorphisms.
- For each $t \in [0,1]$, $F_t \colon K \times \{t\} \to S^{n+2}$ is a smooth embedding.

An n-dimensional knot K is *oriented* if it is oriented as a submanifold. For oriented knots to be smoothly isotopic we additionally demand that F_0 and F_1 are orientation-preserving.
An (oriented) n-dimensional knot is *trivial* or the *unknot* if it is the boundary of an (oriented) smooth submanifold $B \subseteq S^{n+2}$ diffeomorphic to $\overline{B}^{n+1}$.
A *classical knot* is a 1-dimensional knot.



The classical unknot and a classical knot drawn as usual in $\mathbb{R}^3 \subseteq \mathbb{R}^3 \cup \{\infty\} \cong S^3$

The relation of smooth isotopy is clearly reflexive and symmetric. However, naively gluing smooth isotopies together might not lead to a smooth map. So transitivity is a bit technical:

Lemma 1.12. *Let $n \geq 1$. Smooth isotopy defines an equivalence relation on (oriented) n-dimensional knots.*

Proof. see [Fri25: Proposition 38.5]. The argument is essentially the following:
It is clear that the relation is reflexive and symmetric. To show that it is transitive, find a smooth monotone map $\Phi\colon [0,1] \to [0,1]$ such that Φ is constant at 0 on $[0,0.1]$ and constant at 1 on $[0.9,1]$. Given a smooth isotopy $F\colon S^n \times [0,1] \to S^{n+2}$, we can precompose it with Φ in the second coordinate to obtain a smooth isotopy between the same knots that is 'horizontal' at the beginning and end. Two such isotopies can be glued together naively to prove transitivity. ∎

We also note that the equivalence class of trivial knots is unique:

Proposition 1.13. *Let $n \geq 1$. Let $K, L \subseteq S^{n+2}$ be trivial (oriented) n-dimensional knots. Then K and L are smoothly isotopic.*

Proof. see [Fri25: Theorem 42.13] ∎

Smooth isotopy is a priori a different equivalence relation than the diffeomorphisms of pairs considered in the last section. In the following theorem we see that they do in fact agree:[2]

Theorem 1.14. *Let $n \geq 1$ and $K, L \subseteq S^{n+2}$ be (oriented) n-dimensional knots. Then K and L are smoothly isotopic if and only if there exists an orientation-preserving diffeomorphism $\Phi\colon (S^{n+2}, K) \to (S^{n+2}, L)$.*

Proof. If the knots are smoothly isotopic, the existence of a suitable Φ follows directly from the Isotopy Extension Theorem. For the reverse direction, we begin with the oriented case:
Let $\Phi\colon (S^{n+2}, K) \to (S^{n+2}, L)$ be an orientation-preserving diffeomorphism. Since K is compact, there exists an orientation-preserving smooth embedding $\varphi\colon \overline{B}^{n+2} \to S^{n+2}$ such that $K \subseteq \varphi(\overline{B}^{n+2})$. Then $\Phi \circ \varphi\colon \overline{B}^{n+2} \to S^{n+2}$ is also an orientation-preserving smooth embedding. By Theorem 1.1 there exists a diffeotopy $F\colon S^{n+2} \times [0,1] \to S^{n+2}$ such that $F_0 = \mathrm{id}_{S^{n+2}}$ and $F_1 \circ \Phi \circ \varphi = \varphi$. Then Φ is diffeotopic to $\Psi := F_1 \circ \Phi$ which restricts to the identity on $\varphi(\overline{B}^{n+2})$. Let $\iota\colon K \to S^{n+2}$ be the inclusion. Since $\iota(K) = K \subseteq \varphi(\overline{B}^n)$, we have $\iota = \Psi \circ \iota$. By construction, this is smoothly isotopic to $\Phi \circ \iota\colon K \to L$, which is an orientation-preserving diffeomorphism.
This proves the oriented case. For the non-oriented case, choose orientations on K and L such that $\Phi|_K\colon K \to L$ is orientation-preserving, apply the oriented case and forget the orientations. ∎

We can now just port the definition of the connected sum on submanifolds to the current setting:

Definition. Let $n \geq 1$ and $K, L \subseteq S^{n+2}$ be oriented n-dimensional knots. Consider their connected sum as $(n+2, n)$-dimensional smooth manifold pairs $(S^{n+2}, K)\#(S^{n+2}, L)$. Choose an orientation-preserving diffeomorphism $\Phi\colon S^{n+2}\#S^{n+2} \to S^{n+2}$. The *connected sum of K and L* is $\Phi(K\#L) \subseteq S^{n+2}$.

[2] The argument is adapted from [Fri25: Proposition 95.11].

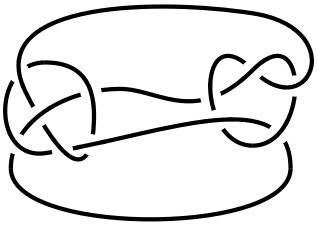
Proposition 1.15. *Let $n \geq 1$ and $K, L \subseteq S^{n+2}$ be oriented n-dimensional knots. Consider $(S^{n+2}, K)\#(S^{n+2}, L)$, i.e. their connected sum as $(n+2, n)$-dimensional smooth manifold pairs. Choose an orientation-preserving diffeomorphism $\Phi\colon S^{n+2}\#S^{n+2} \to S^{n+2}$.*

(1) *The connected sum $K\#L \subseteq S^{n+2}$ is an oriented n-dimensional knot.*
(2) *The connected sum of K and L only depends on the smooth isotopy class of K and L. In particular, it does not depend on how $(S^{n+2}, K)\#(S^{n+2}, L)$ is formed and the choice of Φ.*

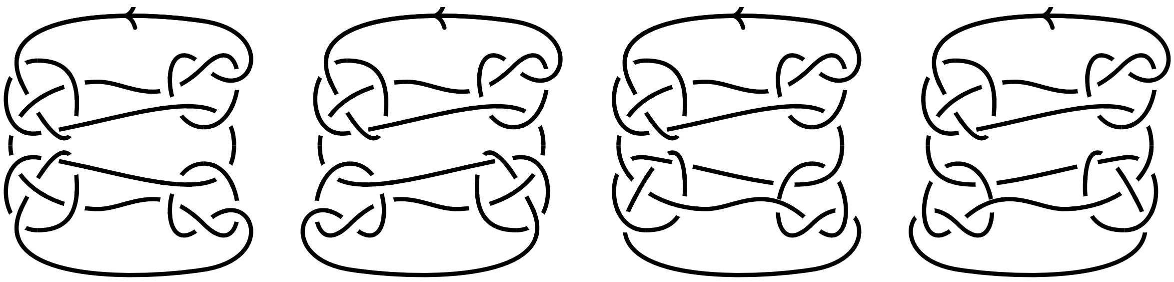
Proof.
(1) follows from Proposition 1.8
(2) By Theorem 1.14 the smooth isotopy class of K, L determines the pairs $(S^{n+2}, K), (S^{n+2}, L)$ up to diffeomorphism. By Proposition 1.8 this determines $(S^{n+2}, K)\#(S^{n+2}, L)$ up to diffeomorphism. Applying Theorem 1.14 to go back to knots yields the claim. ■

As with manifolds the precise orientation conventions are important. In particular, the connected sum cannot be defined on unoriented knots:

Example 1.16. Consider the Conway knot C below:



There are four possible choices of orientations for embeddings $(\overline{B}^3, \overline{B}^1) \to (S^3, C)$ – two each for the choice of orientation in surrounding S^3 and the knot. Fix one such embedding and use it to form the connected sum with all four embeddings, leading to the following four knots:



We will see in Theorem 5.22 that they represent four different smooth isotopy classes since the Conway knot is neither reversible nor amphichiral nor invertible (see [Fri25: Example p.2056]).

The next step is to study how algebraic invariants behave under this connected sum operation. Of course, we again want to turn to the complement for this. The following corollary to Theorem 1.14 shows that this is valid:

Corollary 1.17. *Let $n \geq 1$ and $K, L \subseteq S^{n+2}$ be smoothly isotopic (oriented) n-dimensional knots. There exists an orientation-preserving diffeomorphism $S^{n+2} \setminus K \cong S^{n+2} \setminus L$.*

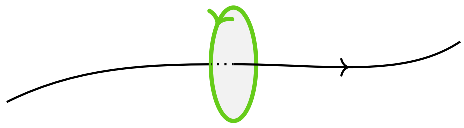
Proof. Restricting the diffeomorphism from Theorem 1.14 gives the desired result. ■

By construction, practically all the work in determining the effect of the connected sum operation on algebraic invariants of the complement has already been done in Proposition 1.10. Furthermore, we do not need to worry about homology by Proposition 1.11. We can however make the result on fundamental groups a bit more explicit by more carefully studying the fundamental group of knot complements. We will later need these results in any case.

We introduce the meridian of a knot and study its role in the fundamental group:

Definition. Let $n \geq 1$ and $K \subseteq S^{n+2}$ be an oriented n-dimensional knot. A *meridian of* K is a closed oriented 1-dimensional smooth submanifold $\mu \subseteq S^{n+2} \setminus K$ such that there exists a smooth embedding $\varphi\colon \overline{B}^2 \to S^{n+2} \setminus K$ with the following properties:
- $\varphi|_{S^1}\colon S^1 \to \mu$ is an orientation-preserving diffeomorphism
- $\varphi(B^2)$ intersects K transversally in a single point P
- A positive basis for $T_P\varphi(B^2)$ followed by a positive basis for T_PK is a positive basis for $T_PS^{n+2} = T_P\varphi(B^2) \oplus T_PK$

Example 1.18.



a meridian for a classical knot

Lemma 1.19. *Let $n \geq 1$ and $K \subseteq S^{n+2}$ be an oriented n-dimensional knot.*
(1) *There exists a meridian for K.*
(2) *Any two meridians of K are smoothly isotopic in $S^{n+2} \setminus K$.*

Proof.
(1) By definition of a submanifold there exists an orientation-preserving smooth embedding $\varphi\colon (\overline{B}^{n+2}, \overline{B}^n) \to (S^{n+2}, K)$. The smooth submanifold

$$D := \{(x_1, \ldots, x_{n+2}) \in B^{n+2} \mid x_1 = \cdots = x_n = 0\} \subseteq B^{n+2}$$

is diffeomorphic to $\overline{B}^2$ and $D \pitchfork \overline{B}^n = \{0\}$. We can orient D such that a positive basis for T_0D followed by a positive basis for $T_0\overline{B}^2$ is a positive basis for $T_0\overline{B}^{n+2}$. Then $\varphi(\partial D)$ is a meridian for K.
(2) Let $\mu, \mu' \subseteq S^{n+2} \setminus K$ be meridians of K with smooth embeddings $\varphi, \varphi'\colon \overline{B}^2 \to S^{n+2} \setminus K$ as in the definition of a meridian. Let P (resp. P') be the point where $\varphi(\overline{B}^2)$ (resp. $\varphi'(\overline{B}^2)$) intersects K. Let $\Phi, \Phi'\colon (\overline{B}^{n+2}, \overline{B}^n) \to (S^{n+2}, K)$ be smooth embeddings with $\Phi(0) = P$ and $\Phi'(0) = P'$. After radially shrinking we may assume that $\varphi(\overline{B}^2) \subseteq \Phi(B^{n+2})$ and $\varphi'(\overline{B}^2) \subseteq \Phi'(B^{n+2})$. By Theorem 1.7 we can assume wlog. that $\Phi = \Phi'$ and $P = P'$. The claim now follows from a mild adaptation of [Fri25: Proposition 37.3 (2)] (here we use the transversality assumptions). ■

The meridian of a knot defines an element in the fundamental group of its complement:

Lemma 1.20. *Let $n \geq 1$ and $K \subseteq S^{n+2}$ be an oriented n-dimensional knot. Let $x \in S^{n+2} \setminus K$. Choose a meridian $\mu \subseteq S^{n+2} \setminus K$ of K, an orientation-preserving diffeomorphism $\varphi\colon S^1 \to \mu$ and a continuous map $\gamma\colon [0,1] \to S^{n+2} \setminus K$ with $\gamma(0) = x$ and $\gamma(1) = \varphi(0,1)$. Then*

$$\gamma * (t \mapsto \varphi(\exp(2\pi\,\mathrm{i}\cdot t)) * \overline{\gamma}$$

is a loop based at x and therefore defines an element $\pi_1(S^{n+2} \setminus K)$. Up to conjugation, this element does not depend on the choice of μ, φ and γ.

Proof. This is a special case of [Fri25: Lemma 98.14]. ■

By convention if we write $[\mu] \in \pi_1(S^{n+2} \setminus K)$ for a meridian μ of K, we mean an element constructed in the manner of Lemma 1.20 – conveniently sweeping under the rug that this is only well-defined up to conjugation.

The meridian of a knot plays an important role in the algebra of the fundamental group of a knot complement:[3]

Proposition 1.21. *Let $K \subseteq S^{n+2}$ be an oriented n-dimensional knot with a meridian μ.*
(1) *The normal closure of $[\mu]$ in $\pi_1(S^{n+2} \setminus K)$ is the whole group.*
(2) *The abelianization $\pi_1(S^{n+2} \setminus K)_{\mathrm{ab}}$ is isomorphic to $\mathbb{Z}$ generated by the meridian.*

Proof.
(1) Let $\varphi\colon S^1 \to S^{n+2} \setminus K$ be a loop. Since S^{n+2} is simply connected, we can extend it to a continuous map $\varphi\colon \overline{B}^2 \to S^{n+2}$. We may assume that φ is smooth (see [Fri25: Theorem 33.1]) and transversal to K (see [Fri25: Theorem 47.9]). By [Fri25: Theorem 47.4]
$$\varphi(\overline{B}^2) \pitchfork K =: \{P_1, \dots, P_n\}$$
is a finite set. We can thereby find smooth embeddings $f_1, \dots, f_n\colon \overline{B}^2 \to \overline{B}^2$ with disjoint images and $f_i(0) = P_i$ such that
   - $(\varphi \circ f_i)(B^2)$ intersects K transversally in the single point P_i
   - A positive basis for $T_{P_i}(\varphi \circ f)(B^2)$ followed by a positive basis for $T_{P_i}K$ is a positive basis for $T_{P_i}S^{n+2} = T_{P_i}K \oplus T_{P_i}\varphi(B^2)$

   Then $(\varphi \circ f_i)(S^1)$ is a meridian of K and in $\pi_1(S^{n+2} \setminus K)$
$$[\varphi(S^1)] = \prod_{i=1}^{n} [\varphi \circ f_i]$$
The claim follows since by Lemma 1.20 all meridians are conjugate.
(2) It follows from (1) that $\pi_1(S^{n+2} \setminus K)_{\mathrm{ab}}$ is generated by the meridian. By the Hurewicz Theorem and Proposition 1.11 this group is isomorphic to $\mathbb{Z}$. ∎

With this we can now provide the promised clarification of Proposition 1.10 (1):

Proposition 1.22. *Let $K, L \subseteq S^{n+2}$ be oriented n-dimensional knots with meridians μ_K, μ_L. Then*
$$\pi_1(S^{n+2} \setminus K\#L) \cong \pi_1(S^{n+2} \setminus K) *_{[\mu_K]=[\mu_L]} \pi_1(S^{n+2} \setminus L)$$

Proof. This is essentially a special case of Proposition 1.10 (1). Note that the following group theoretic observation together with Lemma 1.20 shows that the amalgamated product on the right does not depend on the choice of $[\mu_K]$ and $[\mu_L]$:

Claim. *Let G, H be groups with elements $g \in G, h \in H$ and automorphisms $\varphi\colon G \to G$, $\psi\colon H \to H$. Then $G *_{g=h} H \cong G *_{\varphi(g)=\psi(h)} H$.*

The automorphisms φ and ψ induce a map $G *_{g=h} H \to G *_{\varphi(g)=\psi(h)} H$ by the universal property of a pushout. Their inverses induce the inverse to it. ∎

[3] The argument for (1) is presented in [DF87: Lemma 2.5].

2. Algebraic considerations

2.1. Factorization in monoids

Before studying connected sum decompositions, we develop a suitable algebraic framework. We start with the following basic definition:

Definition. A *monoid* is a set M equipped with a binary operation $\cdot\colon M \times M \to M$ such that
- $(a \cdot b) \cdot c = a \cdot (b \cdot c)$ for all $a, b, c \in M$
- there exists $e_M \in M$ with $e_M \cdot a = a = a \cdot e_M$ for all $a \in M$.

A monoid M is *abelian* if $a \cdot b = b \cdot a$ for all $a, b \in M$.
A *monoid homomorphism* $f\colon M \to N$ between monoids M and N is a map such that
- $f(a \cdot b) = f(a) \cdot f(b)$ for all $a, b \in M$
- $f(e_M) = e_N$

The *kernel* of a monoid homomorphism is $\ker(f) := f^{-1}(e_N)$.
As usual, we mostly omit the symbol for the operation.

A monoid does not carry much structure – therefore, there are lots of examples. In the following, we note a few, at the same time establishing some conventions:

Example 2.1.
- $\mathbb{N}_0$ is the monoid of non-negative integers under addition.[4]
- $\mathbb{N}$ is the monoid of positive integers under multiplication.
- $\mathbb{Z} \setminus \{0\}$ is the monoid of non-zero integers under multiplication.

In ring theory one studies decompositions in the form of prime factorizations. We adopt some of the relevant terminology for our purposes with the next two definitions:

Definition. Let M be an abelian monoid and $a, b \in M$. We say
- a *divides* b if there exists $c \in M$ such that $ac = b$.
- a is a *unit* if it divides the neutral element. The *group of units* is M^*.
- a and b are *associated* if there exists a unit $u \in M^*$ such that $au = b$.

Further, an element $m \in M$ is
- *prime* if m is not a unit and if m divides ab then m also divides a or b for all $a, b \in M$.
- *irreducible* if m is not a unit and $m = ab$ for some $a, b \in M$ then a or b is a unit.
- *cancellable* if $ma = mb$ implies $a = b$ for all $a, b \in M$.
- *weakly cancellable* if $ma = mb$ implies that a and b are associated for all $a, b \in M$.

Definition. An abelian monoid M allows factorization if every $m \in M \setminus M^*$ can be expressed as $m = a_1 \cdots a_n$ with $a_1, \ldots, a_n \in M$ irreducible. The monoid M is a *unique factorization monoid* if additionally $a_1 \cdots a_n = b_1 \cdots b_m$ with $a_1, \ldots, a_n, b_1, \ldots, b_m \in M$ irreducible implies that there exists a bijection $\sigma\colon \{1, \ldots, n\} \to \{1, \ldots, m\}$ such that a_i is associated to $b_{\sigma(i)}$ for all $i \in \{1, \ldots, n\}$.

Since we modelled our definitions after ring theory, ring theory also gives us an easy example:

Example 2.2. If R is an integral domain, $R \setminus \{0\}$ is an abelian monoid under multiplication. In it the notions unit, prime, irreducible mean precisely what they do in usual ring theory and every element is cancellable. Furthermore, $R \setminus \{0\}$ is a unique factorization monoid if and only if R is a unique factorization domain.

[4]Of course, this set is also a monoid under multiplication, but we will always consider it with respect to addition.

We see that all monoids arising from ring theory are special in that they allow cancellation. It can be easily seen that this will not always hold. Another ring theoretic result is therefore lost – prime no longer implies irreducible:

Example 2.3. Let M be the monoid with two elements where one element is neutral and the other squares to itself. The latter element is prime since if it divides a product it is equal to one of the factors. But it is a square and hence not irreducible.

The lack of cancellation allows monoids to have 'attracting elements'. This leads to a counterexample to another seemingly intuitive statement:

Example 2.4. Let M be an abelian monoid. We say *every divisor chain in M stabilizes* if for all $m_1, m_2, \dots \in M$

$$\underset{i \geq 1}{\forall}\ m_{i+1} \text{ divides } m_i \Rightarrow \underset{n \geq 1}{\exists}\ \underset{i \geq n}{\forall}\ m_i \text{ is associated to } m_{i+1}$$

- Intuitively, it might seem as if this implied that M allows factorization[5]. This is however not the case:
  Consider $\mathbb{N}_0$ as a monoid under *multiplication*. Every divisor chain in this monoid stabilizes: If a divisor chain contains a positive integer at some point, all elements after it are also positive and the chain classically stabilizes. The only other chain has all entries equal to 0. But this monoid does not allow factorization: Every factorization of 0 must contain 0, but 0 is not irreducible.
- The converse of this implication is also wrong, but the counterexample is somewhat more involved. It was found by A. Grams [Gra74] – after the claim had also been believed to be correct:
  Let $(p_n)_{n \geq 0}$ be the sequence of odd primes and consider

  $$M := \left\{ \sum_{i=1}^{n} \frac{1}{2^{a_i} p_{a_i}} \;\middle|\; n \in \mathbb{N}_0,\ \underset{i \in \{1,\dots,n\}}{\forall}\ a_i \geq 0 \right\}$$

  as a monoid under addition. By considering the p_a-adic valuations on $M \subseteq \mathbb{Q}$, one can directly see that $\frac{1}{2^a p_a}$ is an irreducible element of M for $a \geq 0$. Hence, M allows factorization. But $\frac{1}{2^a} = \frac{1}{2^a p_a} \cdot p_a \in M$. We therefore have a non-stabilizing divisor chain $(\frac{1}{2^a})_{a \geq 0}$ in M.

These examples should serve as a warning to be very careful when dealing with monoids. Thankfully, some properties do generalize from rings to monoids. In a unique factorization monoid, primes and irreducibles still agree:

Proposition 2.5. *Let M be an abelian monoid.*
(1) *If M is a unique factorization monoid, every element of M is weakly cancellable.*
(2) *If $m \in M$ is prime and weakly cancellable, m is irreducible.*
(3) *If M is a unique factorization monoid and $m \in M$ is irreducible, m is also prime.*

Proof.
(1) Let $m, a, b \in M$ with $ma = mb$. By factorizing m, a, b into irreducibles, we see that the irreducible factors of a and b are associated. So a and b are associated.
(2) Let $a, b \in M$ such that $m = ab$. Wlog. there exists $c \in M$ such that $a = mc$, i.e. $m = mcb$. Then cb is associated to the neutral element, which implies that b is a unit.
(3) Let $a, b \in M$ and m divide ab. Since m is irreducible, it needs to appear as an irreducible factor of ab and therefore by uniqueness also as an irreducible factor of a or b. Hence, m divides a or b. ∎

[5] as used, for example, in [BCF21: Corollary 2.5].

This raises the question whether the 'weakly' in Proposition 2.5 (1) could be omitted:

Example 2.6. The equivalence relation $x \sim -x$ for $x \neq \pm 1$ on $\mathbb{Z} \setminus \{0\}$ is compatible with the operation. Hence, we get an induced monoid structure on the quotient by it. This inherits unique factorization, but $[2] \cdot [-1] = [2] \cdot [1]$. Since $[-1] \neq [1]$ we deduce that $[2]$ is not cancellable.

In ring theory it can be rather hard to directly prove that a ring is factorial. The standard examples for factorial domains all arise as Euclidean rings. To further study factorization in monoids, we proceed similarly:

Definition. Let M be an abelian monoid. An *(additive) complexity function* on M is a monoid homomorphism $c\colon M \to \mathbb{N}_0$. We write $\widetilde{M} \coloneqq c^{-1}(0)$.
A *(multiplicative) complexity function* on M is a monoid homomorphism $c\colon M \to \mathbb{N}$. We write $\widetilde{M} \coloneqq c^{-1}(1)$.

The next proposition in particular says that we can combine any finite number of complexity functions on a monoid into a single multiplicative complexity function.

Proposition 2.7. *Let M be an abelian monoid.*

(1) *Let $c_1, \ldots, c_n\colon M \to \mathbb{N}_0$ be additive complexity functions. Then $c \coloneqq c_1 + \cdots + c_n\colon M \to \mathbb{N}_0$ is an additive complexity function and $c^{-1}(0) = c_1^{-1}(0) \cap \cdots \cap c_n^{-1}(0)$.*

(2) *Let $c_1, \ldots, c_n\colon M \to \mathbb{N}$ be multiplicative complexity functions. Then $c \coloneqq c_1 \cdots c_n\colon M \to \mathbb{N}$ is a multiplicative complexity function and $c^{-1}(1) = c_1^{-1}(1) \cap \cdots \cap c_n^{-1}(1)$.*

(3) *If $c\colon M \to \mathbb{N}_0$ is an additive complexity function,*

$$\begin{array}{rcl} \bar{c}\colon M & \to & \mathbb{N} \\ m & \mapsto & \alpha^{c(m)} \end{array}$$

is a multiplicative complexity function for all $\alpha \in \mathbb{N}$ and $c^{-1}(0) = \bar{c}^{-1}(1)$.

Proof. (1) and (2) are essentially clear and (3) is the functional equation of exponentiation. ∎

Note the asymmetry in this proposition: We can easily turn an additive complexity function into a multiplicative one, but not vice versa. The monoid $\mathbb{N}$ is the free monoid on countably many generators, so does of course contain the free monoid on one generator $\mathbb{N}_0$. The converse is not true since every submonoid of $\mathbb{N}_0$ is finitely generated[6]. We defined both types of function since, although additive complexity functions are easier to work with, some monoids only easily carry a multiplicative complexity function, as illustrated by the following example:

Example 2.8.

- The absolute value defines a multiplicative complexity function on $\mathbb{Z} \setminus \{0\}$.
- Cardinality gives a multiplicative complexity function on the monoid of finite abelian groups under direct sum.

Of course any type of complexity function is much less than a Euclidean function on a Euclidean domain – indeed, every monoid can be endowed with the trivial complexity function. If the $\widetilde{M}$ is too large, we cannot hope to learn much. Instead, we only obtain information relative to $\widetilde{M}$:

Definition. Let M be an abelian monoid. A submonoid $S \subseteq M$ is *closed under division* if $ab \in S$ implies $a, b \in S$ for all $a, b \in M$.

- An element $a \in M$ is *irreducible relative to S* if $a \notin S$ and $a = bc$ for $b, c \in M$ implies $b \in S$ or $c \in S$.
- Two elements $a, b \in M$ are *associated relative to S* if there exist $s, t \in S$ such that $as = b$ and $a = bt$.

The monoid M *allows factorization relative to S* if every $m \in M \setminus S$ can be written as $m = a_1 \cdots a_n$ with $a_1 \cdots a_n$ irreducible relative to S.
The monoid M is a *unique factorization monoid relative to S* if additionally $a_1 \cdots a_n = b_1 \cdots b_m$ with $a_1, \ldots, a_n, b_1, \ldots, b_m \in M$ irreducible relative to S implies that there exists a bijection $\sigma\colon \{1, \ldots, n\} \to \{1, \ldots, m\}$ such that a_i is associated to $b_{\sigma(i)}$ relative to S for all $i \in \{1, \ldots, n\}$.

[6]This is essentially a theorem of I. Schur, see [Bra42: Theorem 1 and the remark preceding it].

Proposition 2.9. *Let M be an abelian monoid with a complexity function c.*
(1) *$\widetilde{M}$ is closed under division.*
(2) *The primes of M are irreducible relative to $\widetilde{M}$.*
(3) *M allows factorization relative to $\widetilde{M}$.*

Proof. By Proposition 2.7 (3) it suffices to prove this for multiplicative complexity functions.
(1) Let $a, b \in \widetilde{M}$. Then $c(a) \cdot c(b) = c(ab) = 1$, so $a, b \in \widetilde{M}$ as $c(a), c(b) \geq 1$.
(2) Let $p \in M$ be prime and $p = ab$ for $a, b \in M$. Wlog. there exists $m \in M$ such that $a = pm$, i.e. $p = pmb$. Then $mb \in \widetilde{M}$ since every element of $\mathbb{N}$ is cancellable. Hence, $b \in \widetilde{M}$ by (1).
(3) We proceed inductively. The case for $m \in M$ with $c(m) = 1$ is clear. Assume we have proven the claim for all $m \in M$ with $c(m) \leq n$ for some $n \in \mathbb{N}_0$. Let $m \in M$ with $c(m) = n + 1$. If m is irreducible relative to $\widetilde{M}$, nothing needs to be proven, else we can factor $m = ab$ with $a, b \notin \widetilde{M}$. Then $c(a), c(b) \leq n$ and the claim follows by induction. ■

For a unique factorization monoid we can go the other way round and obtain a complexity function from the factorization using the Axiom of Choice:

Proposition 2.10. *Let M be a unique factorization monoid relative to a submonoid S closed under division. Then there exists an additive complexity function $c\colon M \to \mathbb{N}$ with $\widetilde{M} = S$.*

Proof. Let $I \subseteq M$ be a set of representative classes of the irreducibles of M relative to S up to being associated relative to S. Then every $m \in M$ can be written as

$$m = u \cdot \prod_{a \in I} a^{i_a(m)}$$

for $u \in S$ and unique $i_a(m) \in \mathbb{N}_0$, $a \in I$ which are almost all 0. The map

$$\begin{aligned} M &\to \mathbb{N}_0 \\ m &\mapsto \sum_{m \in I} i_a(m) \end{aligned}$$

is the desired complexity function. ■

Using this, we can now turn a multiplicative complexity function into an additive one. We will never make use of this observation but still want to point it out for completeness.

Lemma 2.11. *Let $c\colon M \to \mathbb{N}$ be a multiplicative complexity function and an abelian monoid M. There exists an additive complexity function $c'\colon M \to \mathbb{N}_0$ with $c^{-1}(0) = c'^{-1}(0)$.*

Proof. Since $\mathbb{Z}\setminus\{0\}$ is a unique factorization monoid, Proposition 2.10 gives an additive complexity function $f\colon \mathbb{Z} \setminus \{0\} \to \mathbb{N}_0$. Then $c' := f \circ c$ is the desired additive complexity function. ■

2.2. Factorization of groups

2.2.1. Abelian groups

Before we return to topology, we digress a bit into decomposition of groups. We will need the basics of this anyway in our considerations of manifolds. We start with abelian groups, finishing what we began in Example 2.8:

Definition. We denote by $\mathcal{A}$ the monoid of isomorphism classes of finitely generated abelian groups under direct sum. Let A be an abelian group.
- The *rank* of A is $\operatorname{rank}(A) := \dim_{\mathbb{Q}}(A \otimes \mathbb{Q})$.
- The *torsion subgroup* of A is $\operatorname{Tor}(A) := \{a \in A \mid \exists n \geq 1 : n \cdot a = 0\}$. The group A is *torsion* if $\operatorname{Tor}(A) = A$. The group A is *torsion-free* if $\operatorname{Tor}(A) = \{e\}$.
- The *torsion size* of A is $t(A) := \# \operatorname{Tor}(A)$.

The torsion subgroup can also be described in terms of tensoring with $\mathbb{Q}$:

Lemma 2.12. *Let A be an abelian group. The kernel of the natural map $A \to A \otimes \mathbb{Q}$ is $\operatorname{Tor}(A)$.*

Proof. Let $a \in A$ and consider the homomorphism $f\colon \mathbb{Z} \to A$ with $f(1) = a$ and resulting monomorphism $\tilde{f}\colon \mathbb{Z}/\ker(f) \hookrightarrow A$. Consider the commutative diagram

$$\begin{tikzcd} \mathbb{Z}/\ker(f) \arrow[r, hook, "\tilde{f}"] \arrow[d] & A \arrow[d, "i"] \\ (\mathbb{Z}/\ker(f)) \otimes \mathbb{Q} \arrow[r, hook, "\tilde{f}\otimes \mathrm{id}_{\mathbb{Q}}"] & A \otimes \mathbb{Q} \end{tikzcd}$$

The bottom map is a monomorphism as $\mathbb{Q}$ is a flat $\mathbb{Z}$-module (see [Lan02: Proposition XVI.3.2]). This implies the claim by the following observations:
- If $a \in \operatorname{Tor}(A)$, $\ker(f) = (n)$ for some $n \geq 1$ and $(\mathbb{Z}/\ker(f)) \otimes \mathbb{Q} = 0$.
- If $a \notin \operatorname{Tor}(A)$, f is injective and $(\mathbb{Z}/\ker(f)) \otimes \mathbb{Q} \cong \mathbb{Z} \otimes \mathbb{Q} \cong \mathbb{Q}$. ∎

This allows us to find a complexity function on $\mathcal{A}$ with $\widetilde{\mathcal{A}} = \{e\}$:

Lemma 2.13.
(1) *Rank defines an additive complexity function on $\mathcal{A}$ with $\widetilde{\mathcal{A}}$ given by torsion groups.*
(2) *Torsion size defines a multiplicative complexity function on $\mathcal{A}$ with $\widetilde{\mathcal{A}}$ given by torsion-free groups. In particular, the torsion-size is finite.*
(3) *The monoid $\mathcal{A}$ allows factorization.*

Proof.
(1) The additivity of the rank is standard algebra (see [Lan02: Proposition XVI.2.1]). The remainder of the claim follows from Lemma 2.12.
(2) The only non-trivial part is to show that $t(A)$ is finite for a finitely generated abelian group A. This follows from the following observations:
   - Subgroups of finitely generated abelian groups are finitely generated abelian, hence $\operatorname{Tor}(A)$ is a finitely generated abelian group (see [Lan02: Proposition X.1.1])
   - Finitely generated abelian torsion groups are finite: If $A = \langle a_1, \ldots, a_n \rangle$ is an abelian torsion group, there exists an epimorphism

$$\bigoplus_{i=1}^{n} \mathbb{Z}/\operatorname{ord}(a_i) \twoheadrightarrow A$$

(3) By (1), (2) and Proposition 2.7 there exists a complexity function on $\mathcal{A}$ with $\widetilde{\mathcal{A}}$ given by torsion-free torsion groups. The only such group is the trivial group, so (3) follows from Proposition 2.9. ∎

Of course, this is no surprise, since the well-known classification of finitely generated abelian groups gives the factorization in $\mathcal{A}$:

Theorem 2.14 (Classification of finitely generated abelian groups). *For a finitely generated abelian group A there exist $n \in \mathbb{N}_0$ and $q_1, \ldots, q_n$ powers of primes such that*

$$A \cong \mathbb{Z}^n \oplus \mathbb{Z}/q_1 \oplus \cdots \oplus \mathbb{Z}/q_n$$

Furthermore, $n = \operatorname{rank}(A)$ and $q_1, \ldots, q_n$ are determined up to permutation.

We dedicate the remainder of this section to reproving this theorem. Since we already know that $\mathcal{A}$ allows factorization, it remains to find all irreducible elements of $\mathcal{A}$ and prove the uniqueness statement. We proceed carefully to ensure we do not already employ the classification in our reasoning.

First, we note that the given decomposition actually consists of irreducibles:

Lemma 2.15. *The groups $\mathbb{Z}$ and $\mathbb{Z}/q$ where q is a power of a prime are irreducible in $\mathcal{A}$.*

Proof. Since $\operatorname{rank}(\mathbb{Z}) = 1$, $\mathbb{Z}$ is irreducible by Lemma 2.13 (1).
Let q be a power of a prime p. Suppose $\mathbb{Z}/q = A \oplus B$ where A, B are non-trivial subgroups of $\mathbb{Z}/q$. Then A and B are both cyclic with order a power of p and therefore each contain a subgroup of order p. But then $\mathbb{Z}/q$ would contain two subgroups of order p. Contradiction! ∎

It will be somewhat harder to show that these are the only irreducible elements of $\mathcal{A}$. We begin doing this with the following fact most often proven as a corollary to the classification:

Proposition 2.16. *A finitely generated torsion-free abelian group A is free abelian.*

Proof. Let $\{a_1, \dots, a_n\} \subseteq A$ be a minimal generating set in the sense that there is no smaller set at all that generates A. Suppose there exist $\lambda_1, \dots, \lambda_n \in \mathbb{Z}$ not all zero such that

$$\lambda_1 a_1 + \dots + \lambda_n a_n = 0$$

Then $d \coloneqq \gcd(\lambda_1, \dots, \lambda_n) \neq 0$ and $\frac{\lambda_1}{d} a_1 + \dots + \frac{\lambda_n}{d} a_n$ must still be 0 since it would otherwise be a non-trivial torsion element. Hence, wlog. $d = 1$.
For $\mu \in \mathbb{Z}$ and $i, j \in \{1, \dots, n\}$ distinct we replace a_i by $a_i' \coloneqq a_i + \mu a_j$ and the resulting set $\{a_1, \dots, a_{i-1}, a_i', a_{i+1}, \dots, a_n\}$ will still be a minimal generating set – but now the dependency relation is

$$\lambda_1 a_1 + \dots + \lambda_i a_i' + \dots + (\lambda_j - \mu\lambda_i) a_j + \dots + \lambda_n a_n = 0$$

By Euclid's algorithm we can iteratively assume that $\lambda_1 = \gcd(\lambda_1, \dots, \lambda_j)$ for $j = 2, \dots, n$. At the end $\lambda_1 = \gcd(\lambda_1, \dots, \lambda_n) = 1$. But this contradicts the minimality of $\{a_1, \dots, a_n\}$. Hence, $a_1, \dots, a_n \in A$ are linear independent. ∎

The next observation is that we can decompose an abelian group into a free and torsion part:

Lemma 2.17. *Let A be a finitely generated abelian group. Then $A/\operatorname{Tor}(A)$ is free abelian and $A \cong \operatorname{Tor}(A) \oplus A/\operatorname{Tor}(A)$.*

Proof. Observe that Lemma 2.12 implies that $\operatorname{Tor}(A) \subseteq A$ is a normal subgroup and $A/\operatorname{Tor}(A)$ embeds into a $\mathbb{Q}$-vector space. Hence, $A/\operatorname{Tor}(A)$ is torsion-free. By Proposition 2.16 $A/\operatorname{Tor}(A)$ is free abelian. We therefore have a split short exact sequence

$$0 \to \operatorname{Tor}(A) \to A \to A/\operatorname{Tor}(A) \to 0$$ ∎

With this established, we can split our search for irreducible abelian groups into two cases. One of them yields $\mathbb{Z}$ generating free abelian groups, the other one is finite groups. Here we can decompose further using the following:

Proposition 2.18 (Fitting decomposition[7]). *Let A be a finite abelian group and $f\colon A \to A$ a homomorphism. There exists $k \geq 1$ such that $A = \operatorname{im}(f^k) \oplus \ker(f^k)$.*

Proof. We have sequences of inclusions

$$\operatorname{im}(f) \supseteq \operatorname{im}(f^2) \supseteq \operatorname{im}(f^3) \supseteq \dots \qquad \text{and} \qquad \ker(f) \subseteq \ker(f^2) \subseteq \ker(f^3) \subseteq \dots$$

Since A is finite, these sequences stabilise from some $k \geq 1$ onward. Then $A = \operatorname{im}(f^k) \oplus \ker(f^k)$:
- Let $a \in A$. Then $f^k(a) = f^{2k}(a')$ for some $a' \in A$. Thus $f^k(a - f^k(a')) = 0$ and

$$a = f^k(a') + (a - f^k(a')) \in \operatorname{im}(f^k) + \ker(f^k)$$

- Let $a \in \operatorname{im}(f^k) \cap \ker(f^k)$. Then $a = f^k(b)$ for some $b \in A$. Then $0 = f^k(a) = f^{2k}(b)$ and therefore $b \in \ker(f^{2k}) = \ker(f^k)$. Hence, $a = f^k(b) = 0$. ∎

[7]This name has nothing to do with the English verb 'to fit' but comes from the 20th century German mathematician Hans Fitting.

Using this we can now further restrict which finite abelian groups can be irreducible:

Definition. A finite group A is a p-group if $\#A = p^k$ for a prime p and $k \in \mathbb{N}_0$.

Lemma 2.19. *An irreducible finite abelian group is a p-group.*

Proof. Let A be an irreducible finite abelian group. Then A is non-trivial, so $\#A$ has a prime factor, say p. Consider the homomorphism $f\colon A \to A$ given by multiplication by p. By Proposition 2.18 $A \cong \mathrm{im}(f^k) \oplus \ker(f^k)$ for some $k \geq 1$. Since A contains an element of order p by Cauchy's Theorem, $\ker(f^k) \neq \{0\}$. Since A is irreducible, $A = \ker(f^k)$ and the order of every element in A is a power of p. It follows again from Cauchy's Theorem that A is a p-group. ∎

It now remains to show that an irreducible abelian p-group is cyclic.

Lemma 2.20. *An irreducible abelian p-group is cyclic.*

Proof. Let A be an irreducible abelian p-group and $a \in A$ an element of maximal order. Let $n \in \mathbb{N}$ with $\mathrm{ord}(a) = p^n$. There exists a group monomorphism $\chi\colon \langle a \rangle \hookrightarrow S^1$ whose image is μ_{p^n}, the group of p^n-th roots of unity. It follows iteratively from the following claim that we can extend χ to a group homomorphism $\chi\colon A \to S^1$.

Claim. *Let G be a finite abelian group, $H \subsetneq G$ a proper subgroup and $\chi\colon H \to S^1$ a group homomorphism. For $g \in G \setminus H$ there exists an extension of χ to $\langle H, g \rangle$.*

Proof. Let $d \geq 1$ minimal with $dg \in H$, i.e. $d = [\langle H, g \rangle : H]$. Let $z \in S^1$ be such that $z^d = \chi(dg)$ and consider

$$\begin{aligned} \tilde{\chi}\colon \langle H, g \rangle &\to S^1 \\ h + ig &\to \chi(h) \cdot z^i \end{aligned}$$

By minimality of d, $z^i = 1$ for $i \in \mathbb{N}_0$ with $ig \in \langle H \rangle$, hence $\tilde{\chi}$ is well-defined. □

The image of χ is a finite subgroup of S^1 and thereby cyclic. Let $g \in A$ such that $\chi(A) = \langle \chi(g) \rangle$. Then $\chi(g)$ has order p^{n+m}, i.e. p^{n+m} divides the order of g in A and $p^{n+m} \leq p^n$ by definition of a, implying $m = 0$. Hence, the image of χ is still μ_{p^n} and generated by $\chi(a)$.
Let $x \in A$. Then $\chi(x) = \chi(ja)$ for some $j \in \mathbb{Z}$. Therefore $\chi(x - ja) = 0$ and

$$x = (x - ja) + (ja) \in \ker(\chi) + \langle a \rangle$$

As χ restricts to a monomorphism on $\langle a \rangle$, it follows that $A \cong \ker(\chi) \oplus \langle a \rangle$. Since A is irreducible, $\ker(\chi) = \{0\}$ and A is cyclic. ∎

With this we can now complete the proof of Theorem 2.14:

Proof of Theorem 2.14. In our terminology, Theorem 2.14 states that $\mathcal{A}$ is a unique factorization monoid with a single unit whose irreducible elements are precisely $\mathbb{Z}$ and $\mathbb{Z}/q$ for q a power of a prime.
By Lemma 2.13 $\mathcal{A}$ allows factorization and has only a single unit. By Lemma 2.15 $\mathbb{Z}$ and $\mathbb{Z}/q$ where q is a power of a prime are irreducible. If A is irreducible in $\mathcal{A}$, A must be free abelian or torsion by Lemma 2.17. In the first case, $A \cong \mathbb{Z}$. In the second case, A must be finite by Lemma 2.13 (2). It follows from Lemma 2.19 and Lemma 2.20 that $A \cong \mathbb{Z}/q$ where q is a power of a prime.
It remains to prove uniqueness. By standard commutative algebra, $n = \mathrm{rank}(A)$, so we only need to deal with the case that A is finite. For this, observe that for a prime p multiplication by p is an isomorphism of $\mathbb{Z}/\tilde{p}^k$ for $\tilde{p}$ a different prime and $k \in \mathbb{N}_0$, and that

$$\frac{\mathbb{Z}/p^k}{p\mathbb{Z}/p^k} \cong \mathbb{Z}/p^{k-1}$$

These observations imply that for all $i \geq 0$ there exists $e_i \in \mathbb{N}_0$ such that $p^i A/p^{i+1} A$ has cardinality p^{e_i} and that e_i is precisely the number of summands in the decomposition whose order is a power of p and at least p^{i+1}. ∎

2.2.2. Non-abelian groups

Having decomposed abelian groups, we turn our attention to possibly non-abelian groups. Here it seems fitting to consider free products instead of direct sums. We again begin by defining a suitable notion of complexity:

Definition. We denote by $\mathcal{G}$ the monoid of isomorphism classes of finitely generated groups with the operation given by free products.
Let G be a finitely generated group. The *rank* $d(G)$ of G is the minimal number of elements in a generating set for G.

For an abelian group, we have now defined two notions of rank: the one from the previous chapter, specifically defined for abelian groups, and the one defined here. Both are commonly referred to as the rank of a group in the literature. Unfortunately, they need not agree:

Proposition 2.21. *Let A be a finitely generated abelian group. Then $d(A) \geq \operatorname{rank}(A)$ with equality if and only if A is free abelian.*

Proof. We have an epimorphism from the free group on $d(A)$ generators onto A. Abelianizing and tensoring with $\mathbb{Q}$ gives a commutative diagram

$$\begin{tikzcd} \mathbb{Z}^{d(A)} \arrow[r, two heads] \arrow[d, hook] & A \arrow[d] \\ \mathbb{Q}^{d(A)} \arrow[r, two heads] & A \otimes \mathbb{Q} \end{tikzcd}$$

which shows $d(A) \geq \operatorname{rank}(A)$. If equality holds, the bottom map is an isomorphism. Then the top map also is one and A is free abelian. ∎

Unsurprisingly, the next step is proving that d does indeed define a complexity on $\mathcal{G}$. This would be quite a bit of work, hence we only refer to the literature:

Theorem 2.22 (Grushko–Neumann). *Rank defines an additive complexity function on $\mathcal{G}$ with $\widetilde{\mathcal{G}}$ given by the trivial group.*

Proof. see [DD89: Theorem 10.4] for a textbook account ∎

From this theorem we as usual get the following corollary:

Corollary 2.23 (Grushko Decomposition). *The monoid $\mathcal{G}$ allows factorization.*

Proof. By Theorem 2.22 this follows from applying Proposition 2.9 to $d(-)$. ∎

In fact, $\mathcal{M}$ is a unique factorization monoid. But proving this would elongate this detour from topology even more, so we only refer to [Kur34: Isomorphiesatz].
Comparing to what we achieved for abelian groups, it would also remain to find all irreducibles of $\mathcal{G}$, but this seems to be quite out of reach. Instead, we end this section by giving at least some examples of irreducibles:

Proposition 2.24. *A finitely generated group $G \neq \{e\}$ is irreducible in $\mathcal{G}$ if*
(1) *it is torsion, in particular, if G is finite*[8]*.*
(2) $\mathrm{Z}(G) := \{z \in G \mid \forall g \in G : gzg^{-1} = z\}$ *is non-trivial, in particular, if G is abelian.*
(3) *it is simple, i.e. has no non-trivial proper normal subgroups.*

Proof. Let A, B be non-trivial groups.
(1) After establishing a normal form for elements of a free product (see [Fri25: Section 85.1]) it is clear that $(ab)^n$ is non-trivial for all $a \in A, b \in B$ non-trivial and $n \geq 1$. Hence, $A * B$ is not torsion.

[8] The question whether finitely generated torsion groups are finite is the Burnside Problem. A counterexample was found by E. Golod and I. Shafarevich in 1964 (see [CD21] for an English account).

(2) Suppose $x \in Z(A * B)$ is non-trivial. After possibly replacing x by some conjugate of x, we may assume that its normal form begins with a non-trivial element of A. Then bx for $b \in B$ non-trivial has a normal form beginning with an element of B. Thus $bx \neq xb$. Contradiction!

(3) The normal closure of A in $A * B$ does not contain non-trivial elements of B: An element x of the normal closure of A can be written as

$$x = \prod_{i=1}^{n} b_i a_i b_i^{-1} = b_1 \cdot a_1 \cdot (b_1^{-1} b_2) \cdot a_2 \cdot (b_2^{-1} b_3) \cdots (b_{n-1}^{-1} b_n) \cdot a_n \cdot b_n^{-1}$$

with $a_1, \ldots, a_n \in A, b_1, \ldots, b_n \in B$. Assume that $n \geq 0$ is minimal such that such an expression for x exists. If the right side of the equality is not in normal form, then
- $a_i = e$ for some $i \in \{1, \ldots, n\}$ or
- $b_i = b_{i+1}$ for some $i \in \{1, \ldots, n\}$ implying

$$b_i a_i b_i^{-1} b_{i+1} a_{i+1} b_{i+1}^{-1} = b_i \cdot a_i a_{i+1} \cdot b_i^{-1}$$

Both cases contradict the minimality of n, so the right side is in normal form. Hence, $x \in B$ if and only if $n = 0$ and $x = e$. ∎

We can generalize these examples with tools from Geometric Group Theory, namely Stallings' Theorem on ends of groups. The required background is too far out of scope to establish here (see [Löh17: Chapter 8]), but we still want to provide the following sketch:

Proposition 2.25. *A reducible element of $\mathcal{G}$ has infinitely many ends or is $\mathbb{Z}/2 * \mathbb{Z}/2$. If it is torsion-free, the converse holds.*

Proof. Let A, B be non-trivial finitely generated groups. If A or B has order at least 3, $A * B$ has infinitely many ends by Stallings' Theorem on ends of groups (see [Löh17: Theorem 8.2.14]). Otherwise, $A \cong B \cong \mathbb{Z}/2$.
If G is a torsion-free group with infinitely many ends, Stallings' Theorem on ends of groups implies that it splits as an amalgamated free product or an HNN-extension over a finite cyclic subgroup. But the only such subgroup is the trivial group, so G splits as a free product[9]. ∎

The converse of Proposition 2.25 for groups with torsion does unfortunately not hold:

Example 2.26. The group $(\mathbb{Z}/6) *_{\mathbb{Z}/2} (\mathbb{Z}/6)$ has infinitely many ends by Stallings' Theorem on ends of groups. But its centre is non-trivial as it certainly contains the amalgam $\mathbb{Z}/2$. Hence it is irreducible by Proposition 2.24 (3).

By finding examples of groups with at most two ends, we can now give a long list of irreducibles:

Example 2.27. The following groups are irreducible in $\mathcal{G}$:
- Finite groups since they have no ends.
- Finitely generated abelian groups are virtually $\mathbb{Z}^n$ by Theorem 2.14. Hence, they have at most 2 ends.
- Virtually cyclic groups except $\mathbb{Z}/2 * \mathbb{Z}/2$ since infinite virtually cyclic groups are precisely the two-ended groups (see [Löh17: Theorem 8.2.14]).
- Products $G \times H$ where G and H are finitely generated and infinite as such groups have one end.
- Fundamental groups of closed manifolds with universal covering homeomorphic to $\mathbb{R}^n$ (in particular of hyperbolic manifolds) since by Riemannian Geometry and the Schwarz–Milnor Lemma (see [Löh17: Corollary 5.4.10]) these groups are quasi-isometric to $\mathbb{R}^n$ and therefore have at most 2 ends.
- Fundamental groups of complements of knots in S^3 by the Sphere Theorem and a Theorem of E. Specker (see [Pap57: Theorem 28.1]).

[9] Note that the HNN-extension over the trivial group is just the free product with $\mathbb{Z}$.

3. Monoids of manifolds

3.1. Existence of factorization

After these algebraic considerations, we are now finally ready to return to topology with the following definition:[10]

Definition. For $n \geq 1$ we consider[11]
- $\mathcal{M}_n^{\mathrm{or}}$, the set of orientation-preserving diffeomorphism classes of non-empty connected closed oriented n-dimensional smooth manifolds
- $\mathcal{M}_n^{\mathrm{no}}$, the set of diffeomorphism classes of non-empty connected closed non-orientable n-dimensional smooth manifolds and S^n
- $\mathcal{M}_n \coloneqq \mathcal{M}_n^{\mathrm{or}} \cup \mathcal{M}_n^{\mathrm{no}}$

Note that $\mathcal{M}_n$ is a bit weird: Its elements are equivalence classes of *oriented* and *non-orientable* manifolds with two being equivalent if either both are oriented and they are orientation-preserving diffeomorphic, or both are non-orientable and they are diffeomorphic. This peculiar equivalence relation is, of course, chosen since we have seen in Example 1.4 this is the set on which connected sum can be defined.
Our recent interest in monoids is motivated by the following proposition stating that connected sum defines a monoid structure:

Lemma 3.1. *Let $n \geq 1$ and X, Y and Z be non-empty connected closed (oriented) n-dimensional smooth manifolds.*
1. *There exists an (orientation-preserving) diffeomorphism $(X\#Y)\#Z \to X\#(Y\#Z)$.*
2. *There exists an (orientation-preserving) diffeomorphism $X\#Y \to Y\#X$.*
3. *There exists an (orientation-preserving) diffeomorphism $X \to X\#S^n$.*

In particular, the connected sum operation descends to a monoid structure on $\mathcal{M}_n$ with $\mathcal{M}_n^{\mathrm{or}}$ and $\mathcal{M}_n^{\mathrm{no}}$ as submonoids.[12]

Proof.
1. By Proposition 1.2 (5) $X\#Y$ and $Y\#Z$ are connected, so the statement is well-defined. The claim follows by letting the smooth embeddings of balls in Y used to define the connected sums have disjoint images.
2. This is clear. Note that when defining the connected sum we did not specify in the orientable case which of the embedded balls is orientation-preserving and which is orientation-reversing.
3. This can be done by choosing a 'nice' smooth embedding $\overline{B}^n \to S^n$. We are sure that the reader will not find it enlightening if we write down an explicit diffeomorphism. ■

[10] This section expands upon and clarifies certain aspects of [BCF21].

[11] For the concerned reader, we briefly sketch that these actually are sets: An n-dimensional closed smooth manifold can be covered by countably many open subsets diffeomorphic to $\mathbb{R}^n$. Hence, it is diffeomorphic to a manifold obtained by quotienting the countable disjoint union of $\mathbb{R}^n$'s by a suitable equivalence relation. Since equivalence relations on the countable disjoint union of $\mathbb{R}^n$'s form a set, diffeomorphism classes of smooth n-manifolds do, too.
In fact, there are only countably many diffeomorphism classes – but this is somewhat harder to prove, see [Fri25: Theorem 90.3].

[12] The categorically inclined reader is warned that this proposition does *not* say that the connected sum operation defines a symmetric monoidal product on a suitable category of manifolds – despite this lining up almost perfectly. The issue is that the connected sum operation is not defined on maps, so does not give the required bifunctor.

In low dimensions we can easily identify these monoids by appealing to classifications:

Example 3.2.

(1) We have $\mathcal{M}_1 = \mathcal{M}_1^{\mathrm{or}} = \mathcal{M}_1^{\mathrm{no}} = \{[S^1]\}$ by the classification of 1-dimensional manifolds. This is trivially a unique factorization monoid.

(2) By the classification of 2-dimensional manifolds the genus gives an isomorphism $\mathcal{M}_2^{\mathrm{or}} \to \mathbb{N}_0$. Similarly, $\mathcal{M}_2^{\mathrm{no}} \cong \mathbb{N}_0$. Put together, $\mathcal{M}_2$ is isomorphic to $\mathbb{N}_0 \sqcup \mathbb{N}$ with the monoid structure given by

$$ab := \begin{cases} a+b \in \mathbb{N}_0, & \text{if } a, b \in \mathbb{N}_0 \\ a+b \in \mathbb{N}, & \text{else} \end{cases}$$

Hence, $\mathcal{M}_2^{\mathrm{or}}$ and $\mathcal{M}_2^{\mathrm{no}}$ are unique factorization monoids – $\mathcal{M}_2$ is not, as exemplified by $\mathbb{RP}^2 \# \mathbb{RP}^2 \# \mathbb{RP}^2 \cong (S^1 \times S^1) \# \mathbb{RP}^2$.

In the absence of easy classifications, we will need a new approach to study $\mathcal{M}_n$ for $n \geq 3$. The next proposition relates $\mathcal{M}_n$ to monoids we already understand better:

Proposition 3.3. *Let $n \geq 3$.*

(1) *The fundamental group defines a monoid homomorphism $\mathcal{M}_n \to \mathcal{G}$.*

(2) *For $k \in \{1, \ldots, n-2\}$ the k-th homology group defines a monoid homomorphism $\mathcal{M}_n \to \mathcal{A}$.*

(3) *The $(n-1)$-st homology group with $\mathbb{F}_2$ coefficients defines a monoid homomorphism $\mathcal{M}_n \to \mathcal{A}$.*

Proof. Most of the work was already done in Proposition 1.3. We only need to show that all these invariants are finitely generated. This follows from compactness and the existence of CW-structures on smooth manifolds (see [Fri25: Proposition 165.9]). ∎

We have already found interesting complexity functions on $\mathcal{A}$ and $\mathcal{G}$. Precomposing them with these homomorphisms gives us a plethora of complexity functions on $\mathcal{M}_n$. Since all these complexity functions are derived from algebraic topology, we can at most hope to study decompositions up to manifolds that algebraically look like the neutral element S^n:

Definition. Let $n \geq 1$. An n-dimensional smooth manifold M is a *homotopy sphere* if $\pi_1(M) \cong \pi_1(S^n)$ and $\mathrm{H}_k(M) \cong \mathrm{H}_k(S^n)$ for all $k \geq 0$.

In fact, all manifolds that algebraically look like S^n are homotopy equivalent to S^n. We will not need this observation, but want to point it out for completeness. Hence, we only sketch a proof:

Theorem 3.4. *Let $n \geq 1$. An n-dimensional smooth manifold is a homotopy sphere if and only if it is homotopy equivalent to S^n.*

Sketch of a proof. Let M be an n-dimensional smooth manifold. The 'if'-direction is obvious. Hence, assume that M is a homotopy sphere. The case $n = 1$ follows from the classification, so we only need to consider $n \geq 2$. By the Hurewicz Theorem, $\pi_i(M) = 0$ for $i \in \{1, \ldots, n-1\}$ and $\pi_n(M) \cong \mathbb{Z}$. Let $f\colon S^n \to M$ be a generator of $\pi_n(M)$. Then f induces an isomorphism on all homology groups by the Hurewicz Theorem. Since S^n and M are simply connected, f must also induce an isomorphism on all homotopy groups (see [Fri25: Theorem 270.26]). Hence, f is a homotopy equivalence by Whitehead's Theorem. ∎

We now return to more elementary considerations, finishing our argument for the existence of factorization in $\mathcal{M}_n$:

Theorem 3.5. *Let $n \geq 1$. There exists a complexity function on $\mathcal{M}_n$ with $\widetilde{\mathcal{M}}_n$ given by homotopy spheres. This implies that*

(1) *The submonoid of $\mathcal{M}_n$ given by homotopy spheres is closed under division, in particular, it contains all units.*

(2) *The monoid $\mathcal{M}_n$ allows factorization relative to homotopy spheres.*[13]

The same holds for $\mathcal{M}_n^{\mathrm{or}}$ and $\mathcal{M}_n^{\mathrm{no}}$.

[13] Note that this statement subtly differs from the corresponding statement [BCF21: Proposition 1.2] in the paper on which this section is based.

Proof. The case $n = 1, 2$ was dealt with in Example 3.2. Hence, we may assume $n \geq 3$. Combining Proposition 3.3 with Theorem 2.22 and Lemma 2.13 gives complexity functions on $\mathcal{M}_n$ with the respective $\widetilde{\mathcal{M}}_n$ given by

– manifolds M with $\pi_1(M)$ trivial.
– manifolds M with $\mathrm{H}_k(M)$ trivial for $k \in \{1, \dots, n-2\}$.
– M with $\mathrm{H}_{n-1}(M; \mathbb{F}_2)$ trivial.

By Proposition 2.7 there exists a complexity function on $\mathcal{M}_n$ such that manifolds in $\widetilde{\mathcal{M}}_n$ have all three of these properties. Such manifolds are homotopy spheres:
Let M be a non-empty connected closed n-dimensional smooth manifold having the three properties. Since M is simply connected, M is orientable (see [Fri25: Corollary 82.16]). It follows that $\mathrm{H}_{n-1}(M)$ is free abelian (see [Fri25: Theorem 179.4]). By the Universal Coefficient Theorem, we have a monomorphism $\mathrm{H}_{n-1}(M) \otimes \mathbb{F}_2 \hookrightarrow \mathrm{H}_{n-1}(M; \mathbb{F}_2) = 0$, hence $\mathrm{H}_{n-1}(M) = 0$. This suffices to show that M is a homotopy sphere.
Statements (1) and (2) follow from Proposition 2.9.
We obtain complexity functions on $\mathcal{M}_n^{\mathrm{or}}$ and $\mathcal{M}_n^{\mathrm{no}}$ by composing with the inclusions to $\mathcal{M}_n$. This shows the claim also holds for these monoids. ∎

This raises the question of whether we can do any better. The following highly non-trivial theorem shows that in most dimensions we cannot:

Theorem 3.6.
(1) *For $n \neq 4$, the units of $\mathcal{M}_n$ and $\mathcal{M}_n^{\mathrm{or}}$ are precisely homotopy spheres*
(2) *For $n \geq 1$, $\mathcal{M}_n^{\mathrm{no}}$ has no units other than S^n.*

Proof. By Theorem 3.5 (1) all of $\mathcal{M}_n$ units are homotopy spheres. For the converse we consider the various dimensions separately:

Case 1: $n = 1, 2$
We have already dealt with this in Example 3.2

Case 2: $n = 3, 5$
In these dimensions the Poincaré conjecture is true. This gives the stronger statement that the neutral element is the only homotopy sphere. In dimension 3 this is due to G. Perelman building on work by R. Hamilton (see [MT07] for a textbook account), in dimension 5 see [Sma61].

Case 3: $n \geq 6$
A homotopy sphere is simply connected and therefore orientable by [Fri25: Corollary 82.16]. Let M be an oriented n-dimensional homotopy sphere and $D \subsetneq M$ a smooth submanifold diffeomorphic to $\overline{B}^n$. Let $\overline{M}$ be M with the orientation reversed. The boundary of $H := (M \setminus \mathring{D}) \times [0, 1]$ is diffeomorphic to $M \# \overline{M}$. Removing the interior of a smooth submanifold of $M \# \overline{M}$ diffeomorphic to $\overline{B}^n$ yields an h-cobordism between $M \# \overline{M}$ and S^n. By the h-cobordism Theorem this implies that $M \# \overline{M} \cong S^n$

Since homotopy spheres are simply connected, they are orientable by [Fri25: Corollary 82.16]. Hence, all homotopy spheres are in $\mathcal{M}_n^{\mathrm{or}}$, and the above proof still applies. For the same reason, $\mathcal{M}_n^{\mathrm{no}}$ cannot have units other than the manually added neutral element S^n. ∎

For completeness, we collect the results from the previous two theorems in the following corollary:

Corollary 3.7.
(1) *For $n \neq 4$, the monoids $\mathcal{M}_n$ and $\mathcal{M}_n^{\mathrm{or}}$ allow factorization.*
(2) *For $n \geq 1$, the monoid $\mathcal{M}_n^{\mathrm{no}}$ allows factorization.*

Proof. combine Theorem 3.5 (2) and Theorem 3.6. ∎

Almost all results we will obtain in the remainder of this chapter are 'relative to homotopy spheres'. For any of them, we can apply Theorem 3.6 to show that 'relative to homotopy spheres' can be omitted for $n \neq 4$ and non-orientable manifolds. We refrain from repeating this explicitly to keep the statements readable.

3.2. Uniqueness of factorization

A natural question to consider now is whether the factorization in $\mathcal{M}_n$ is unique. Since we have already seen in Example 3.2 that this is not even the case in dimension 2, we should expect this not to be the case. Indeed, we show in this section that the counterexample in dimension 2 directly generalizes to all higher dimensions:[14]

Definition. Let $n \geq 2$. View S^{n-1} as the unit sphere in $\mathbb{R}^n$ and let r be the reflection in the first coordinate. Define

$$S^1 \tilde{\times} S^{n-1} := \mathrm{Tor}(S^{n-1}, r) := S^{n-1} \times [0,1]/\sim$$

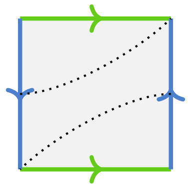
where $\sim$ is the equivalence relation generated by $(x,0) \sim (r(x),1)$ for all $x \in S^{n-1}$.

Lemma 3.8. *Let $n \geq 2$*

(1) *The topological space $S^1 \tilde{\times} S^{n-1}$ is a closed non-orientable n-dimensional smooth manifold such that $S^{n-1} \times [0,1] \twoheadrightarrow S^1 \tilde{\times} S^{n-1}$ is smooth.*
(2) *We have $\pi_1(S^1 \tilde{\times} S^{n-1}) \cong \mathbb{Z}$ and*
   – *for $n = 2$:* $\mathrm{H}_k(S^1 \tilde{\times} S^{n-1}) \cong \begin{cases} \mathbb{Z} \oplus \mathbb{Z}/2, & \text{if } k = 1 \\ \mathbb{Z}, & \text{if } k = 0 \\ 0, & \text{else} \end{cases}$
   – *for $n \geq 3$:* $\mathrm{H}_k(S^1 \tilde{\times} S^{n-1}) \cong \begin{cases} \mathbb{Z}/2, & \text{if } k = n-1 \\ \mathbb{Z}, & \text{if } k = 0,1 \\ 0, & \text{else} \end{cases}$
(3) *For $n = 2$, $S^1 \tilde{\times} S^{n-1} \cong \mathbb{RP}^2 \# \mathbb{RP}^2$.*
   For $n \geq 3$, $S^1 \tilde{\times} S^{n-1}$ is irreducible relative to homotopy spheres in $\mathcal{M}_n$ and $\mathcal{M}_n^{\mathrm{no}}$.

Proof.

(1) see [Fri25: Proposition 51.4]
(2) see [Fri25: Propositions 80.13, 139.6]
(3) For $n = 2$, $S^1 \tilde{\times} S^1$ is given by identifying the sides of a square as indicated below



Cutting along the dotted S^1 decomposes this surface into two Möbius bands. As $\mathbb{RP}^2$ can be formed by attaching a disc to a Möbius band, we have $S^1 \tilde{\times} S^1 \cong \mathbb{RP}^2 \# \mathbb{RP}^2$.
For $n \geq 3$, let X, Y be non-empty connected closed n-dimensional smooth manifolds such that $S^1 \tilde{\times} S^{n-1} \cong X \# Y$. Since $S^1 \tilde{\times} S^{n-1}$ is non-orientable, X or Y needs to be non-orientable by Proposition 1.2 (2), wlog. X. By Proposition 1.3 $\pi_1(X) * \pi_1(Y) \cong \pi_1(S^1 \tilde{\times} S^{n-1}) \cong \mathbb{Z}$. By Proposition 2.24 (2), precisely one of $\pi_1(X)$ and $\pi_1(Y)$ is trivial and the other is isomorphic to $\mathbb{Z}$. By [Fri25: Corollary 82.16] a non-orientable manifold must have non-trivial fundamental group, so $\pi_1(X) \cong \mathbb{Z}$ and $\pi_1(Y) = 0$. It follows from the Hurewicz Theorem that $\mathrm{H}_1(X) \cong \mathbb{Z}$ and $\mathrm{H}_1(Y) = 0$. Furthermore, applying [Fri25: Corollary 82.16] again yields that Y is orientable. Now Proposition 1.3 says that for $k \in \{1, \ldots, n-1\}$

$$\mathrm{H}_k(X) \oplus \mathrm{H}_k(Y) \cong \mathrm{H}_k(S^1 \tilde{\times} S^{n-1})$$

Hence, $\mathrm{H}_k(Y) = 0$ for $k \in \{1, \ldots, n-2\}$. For $k = n-1$, it follows from [Fri25: Theorem 179.4] that $\mathrm{H}_k(Y) = 0$, too. This shows that Y is a homotopy sphere. ■

[14] The idea of the presented construction originates in [Hem76: Lemmas 3.16, 3.17].

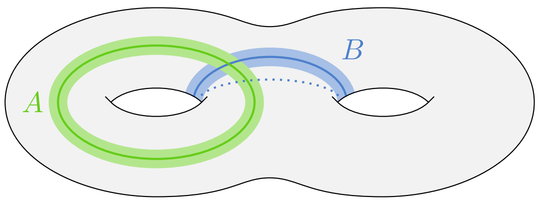
In Example 3.2 we have therefore seen that $\mathbb{RP}^2 \# S^1 \tilde{\times} S^1 \cong \mathbb{RP}^2 \# S^1 \times S^1$. We now generalize this behaviour to all dimensions. First, we need a geometric way to detect $S^1 \times S^{n-1}$ and $S^1 \tilde{\times} S^{n-1}$ factors in a given manifold:

Proposition 3.9. *Let $n \geq 2$ and M be a non-empty connected closed n-dimensional smooth manifold. Suppose there exists a smooth embedding $\iota\colon S^{n-1} \hookrightarrow M$ such that $M \setminus \iota(S^{n-1})$ is connected and there is a smooth embedding $\tau\colon S^{n-1} \times [-1,1] \hookrightarrow M$ with $\tau|_{S^{n-1}\times\{0\}} = \iota$*[15]. *Consider*

$$N := (M \setminus \tau(S^{n-1} \times (-1,1))) \cup_{\tau|_{S^{n-1}\times\{\pm 1\}}} \overline{B}^n \times \{\pm 1\}$$

(1) *If M is orientable, assume it is oriented and orient N such that the orientations agree on $M \setminus \tau(S^{n-1} \times [-1,1])$. Then $M \cong N \# S^1 \times S^{n-1}$.*
(2) *If M is non-orientable, $M \cong N \# S^1 \tilde{\times} S^{n-1}$.*

Proof. By [Fri25: Proposition 51.4] N is a smooth manifold with a smooth structure unique up to diffeomorphism that can be oriented as described. Since $M \setminus \iota(S^{n-1})$ is connected, there exists a smooth embedding $\jmath\colon S^1 \to M$ such that ι and $\jmath$ intersect transversally in a single point. We choose a tubular neighbourhood $B \to \jmath(S^1)$ that is transverse to the tubular neighbourhood $A := \tau(S^{n-1} \times [-1,1]) \to \iota(S^{n-1})$ in the sense of [Fri25: Theorem 54.32]. The intersection $\partial A \cap \partial B \subseteq M$ is a smooth submanifold wlog. diffeomorphic to the disjoint union of two S^{n-2}'s. After smoothing in a tubular neighbourhood of it we may enlarge B in such a way that $R := A \cup B \subseteq M$ is a smooth submanifold.[16]



It is diffeomorphic to $S^{n-1} \times [-1,1]$ with a 1-handle attached. Its boundary is therefore diffeomorphic to the disjoint union of two $\overline{B}^{n-1}$'s with $S^{n-1} \times [-1,1]$ attached between them, i.e. $\partial R \cong S^{n-1}$.
By construction $\tau(S^{n-1} \times [-1,1]) \subseteq R$, so we can consider

$$D := (R \setminus \tau(S^{n-1} \times (-1,1))) \cup_{\tau|_{S^{n-1}\times\{\pm 1\}}} \overline{B}^n \times \{\pm 1\}$$

Then D is given by two $\overline{B}^n$'s with a 1-handle in between, i.e. $D \cong \overline{B}^n$. By [Fri25: Proposition 22.48] $M \setminus \mathring{R}$ is also a codimension 0 submanifold of M with boundary equal to ∂R. Clearly,

$$N = (M \setminus \mathring{R}) \cup_{\partial R} D$$

By [Fri25: Proposition 51.4] we have a uniquely determined smooth manifold

$$X := R \cup_{\partial R} D$$

Since $D \cong \overline{B}^n$, $M \cong N \# X$.

(1) If M is orientable, the 1-handle attached to $S^{n-1} \times [-1,1]$ to form R is also orientable. It follows that $X \cong S^1 \times S^{n-1}$.
(2) For non-orientable M we may assume that $\jmath$ is an orientation-reversing loop (see [Fri25: Chapter 42]). Then the 1-handle attached to $S^{n-1} \times [-1,1]$ is non-orientable and therefore $X \cong S^1 \tilde{\times} S^{n-1}$. ∎

[15] For $n \geq 3$, τ always exists, since a tubular neighbourhood of $\iota(S^{n-1})$ exists by [Fri25: Theorem 54.11] and must be given by a product as any $[-1,1]$-bundle over S^k is trivial for $k \geq 2$. If M is orientable, τ also exists by [Fri25: Theorem 55.4]

[16] Of course, the tubular neighbourhoods are then no longer transverse.

With this established we immediately get the promised non-uniqueness as a corollary:

Corollary 3.10. *Let $n \geq 2$.*

1. *Let N be a non-empty connected closed non-orientable n-dimensional smooth manifold. Then $N\#S^1 \times S^{n-1} \cong N\#S^1 \mathbin{\tilde{\times}} S^{n-1}$.*
2. *No element of $\mathcal{M}_n^{\mathrm{no}}$ is weakly cancellable in $\mathcal{M}_n$. In particular, $\mathcal{M}_n$ is not a unique factorization monoid.*

Proof.

1. Choose $x \in S^1$ such that $\{x\} \times S^{n-1}$ is not affected by the n-ball used to construct $N\#S^1 \times S^{n-1}$. We then apply Proposition 3.9 (2) to $S^{n-1} \hookrightarrow \{x\} \times S^{n-1} \subseteq N\#S^1 \times S^{n-1}$. The same argument as in the proof of Proposition 3.9 shows that the N considered here is diffeomorphic to the N obtained from Proposition 3.9 (2) proving the claim.
2. By (1) and Proposition 2.5 (1) it only remains to show that $S^1 \times S^{n-1}$ and $S^1 \mathbin{\tilde{\times}} S^{n-1}$ are not associated. Every unit of $\mathcal{M}_n$ is a homotopy sphere by Theorem 3.5 (1) and every homotopy sphere is orientable by [Fri25: Corollary 82.16]. Therefore, any element associated to $S^1 \times S^{n-1}$ needs to be orientable – but $S^1 \mathbin{\tilde{\times}} S^{n-1}$ is not by Lemma 3.8. ■

It is perhaps not too surprising that by mixing orientable and non-orientable manifolds we can generate weird behaviour. Alas restricting the orientability only saves us in low dimensions: We have seen in Example 3.2 that uniqueness holds in $\mathcal{M}_2^{\mathrm{or}}$ and $\mathcal{M}_2^{\mathrm{no}}$ and the following theorem gives uniqueness in dimension 3:

Theorem 3.11 (Kneser-Milnor Decomposition).

1. *The monoids $\mathcal{M}_3^{\mathrm{or}}$ and $\mathcal{M}_3^{\mathrm{no}}$ are unique factorization monoids.*
2. *The monoid $\mathcal{M}_3$ allows factorization which is unique if one demands that $S^1 \times S^2$ never appears in factorizations of non-orientable manifolds.*

Proof. For orientable manifolds existence of factorization was proven by H. Kneser [Kne29][17] and uniqueness by J. Milnor [Mil62]. The generalization to non-orientable manifolds involves ideas starting of Proposition 3.9, see [Hem76: Theorem 1.21] noting [Tra87]. ■

In higher dimensions, uniqueness of factorization also fails under orientation conditions:

Theorem 3.12. *Let $n \geq 4$. The manifold $S^2 \times S^{n-2}$ is not weakly cancellable in $\mathcal{M}_n^{\mathrm{or}}$. In particular, $\mathcal{M}_n^{\mathrm{or}}$ is not a unique factorization monoid.*

Proof. see [BCF21: Theorem 1.3] ■

3.3. Irreducible manifolds

Having established decomposition into irreducible manifolds a natural question to ask is which manifolds actually are irreducible. The next statement shows that there are infinitely many:

Proposition 3.13. *For $n \geq 3$, the monoids $\mathcal{M}_n$, $\mathcal{M}_n^{\mathrm{or}}$ and $\mathcal{M}_n^{\mathrm{no}}$ are not finitely generated. In particular, they contain infinitely many elements irreducible relative to homotopy spheres.*

Proof. The 'in particular' statement follows from the first part by Theorem 3.5 (2). For the first part, we show that $\mathcal{M}_n$, $\mathcal{M}_n^{\mathrm{or}}$ and $\mathcal{M}_n^{\mathrm{no}}$ have an epimorphism to a non-finitely generated submonoid of $\mathbb{N}$:
Taking the first homology group defines a monoid homomorphism $\mathcal{M}_n \to \mathcal{A}$ by Proposition 3.3. Consider its composition with the torsion size from Lemma 2.13 (2) to get a monoid homomorphism $\varphi_n \colon \mathcal{M}_n \to \mathbb{N}$. For every prime p there exists a connected closed oriented 3-dimensional

[17] Note that this proof dates from 1929 and therefore cannot rely on the 2003 resolution of the 3-dimensional Poincaré conjecture as we did. Instead, it uses a far more clever argument: By Alexanders Theorem [Ale24], there no units in $\mathcal{M}_3$ but S^3. H. Kneser now shows that the number of factors in a factorization of a manifold M is bounded above in terms of the number of elements of $\mathrm{H}_1(M;\mathbb{F}_2)$ – and therefore must be finite.

smooth manifold L_p with first homology group $\mathbb{Z}/p$ (such as a lens space). Then $\varphi_3(L_p) = p$. Similarly, $\varphi_n(L_p \times S^{n-3}) = p$ for $n > 3$. Hence, φ_n and its restriction to $\mathcal{M}_n^{\mathrm{or}}$ are surjective. For non-orientable manifolds consider the connected sum of the above examples with $\mathbb{R}\mathrm{P}^2 \times S^{n-2}$. This shows that the image of φ_n restricted to $\mathcal{M}_n^{\mathrm{no}}$ always contains the submonoid of $\mathbb{N}$ generated by $\{2p \mid p \text{ prime}\}$. ∎

The structure of the cohomology ring of a connected sum directly gives us our first examples of irreducible manifolds:

Proposition 3.14.
1. *For $n \geq 1$, $\mathbb{C}\mathrm{P}^n$ is irreducible in $\mathcal{M}_{2n}$ and $\mathcal{M}_{2n}^{\mathrm{or}}$ relative to homotopy spheres.*
2. *For $n, m \geq 1$, $S^n \times S^m$ is irreducible in $\mathcal{M}_{n+m}$ and $\mathcal{M}_{n+m}^{\mathrm{or}}$ relative to homotopy spheres.*

Proof.
1. Suppose $\mathbb{C}\mathrm{P}^n \cong X \# Y$ for X, Y non-empty connected closed $2n$-dimensional smooth manifolds. Since $\mathbb{C}\mathrm{P}^n$ is orientable, both X and Y need to be oriented by Proposition 1.2 (2). By Proposition 1.3 X and Y are simply connected and there exists a ring epimorphism $\mathrm{H}^*(X \vee Y; \mathbb{Z}) \twoheadrightarrow \mathrm{H}^*(X \# Y; \mathbb{Z})$ whose kernel is the ideal generated by $[X]^* - [Y]^*$. Let $a \in \mathrm{H}^2(\mathbb{C}\mathrm{P}^n; \mathbb{Z})$ be a generator of $\mathrm{H}^*(\mathbb{C}\mathrm{P}^n; \mathbb{Z})$, i.e. $\mathrm{H}^*(\mathbb{C}\mathrm{P}^n; \mathbb{Z}) \cong \mathbb{Z}[a]/(a^{n+1})$. For $k \neq 0, 2n$ we have isomorphisms
$$\mathrm{H}^k(X) \oplus \mathrm{H}^k(Y) \cong \mathrm{H}^k(X \vee Y) \cong \mathrm{H}^k(\mathbb{C}\mathrm{P}^n) = \begin{cases} \mathbb{Z} \cdot a^i, & \text{if } k = 2i \text{ is even} \\ 0, & \text{if } k \text{ is odd} \end{cases}$$
It follows that we can assume wlog. that $\mathrm{H}^2(X) \cong \mathbb{Z}$ and $\mathrm{H}^2(Y) = 0$ and continuing upwards by using the ring structure that $\mathrm{H}^k(X) \cong \mathrm{H}^k(\mathbb{C}\mathrm{P}^n)$ and $\mathrm{H}^k(Y) = 0$. Hence Y is a homotopy sphere.
2. Suppose $S^n \times S^m \cong X \# Y$ for X, Y non-empty connected closed $(n+m)$-dimensional smooth manifolds. Since $S^n \times S^m$ is orientable, both X and Y need to be oriented by Proposition 1.2 (2). By Proposition 1.3 there exists a ring epimorphism $\mathrm{H}^*(X \vee Y; \mathbb{Z}) \twoheadrightarrow \mathrm{H}^*(X \# Y; \mathbb{Z})$ whose kernel is the ideal generated by $[X]^* - [Y]^*$. We begin by showing that $\mathrm{H}^k(Y) = 0$ for $k \neq 0, n+m$. We only consider the case where $n \neq m$, the case $n = m$ is similar. By the Künneth formula
$$\mathrm{H}^k(S^n \times S^m) = \begin{cases} \mathbb{Z} \cdot [S^n \times S^m]^*, & \text{if } k = n+m \\ \mathbb{Z} \cdot [S^n]^*, & \text{if } k = n \\ \mathbb{Z} \cdot [S^m]^*, & \text{if } k = m \\ \mathbb{Z}, & \text{if } k = 0 \\ 0, & \text{else} \end{cases}$$
with the cup-product given by $[S^n]^* \cup [S^m]^* = [S^n \times S^m]^*$. For $k \neq 0, n+m$, Proposition 1.3 gives isomorphisms
$$\mathrm{H}^k(X) \oplus \mathrm{H}^k(Y) \cong \mathrm{H}^k(X \vee Y) \cong \mathrm{H}^k(S^n \times S^m)$$
It follows that we can assume wlog. that $\mathrm{H}^n(X) = \mathbb{Z} \cdot a$ and $\mathrm{H}^n(Y) = 0$ where a corresponds to $[S^n]^*$ under the above isomorphism. For $k = m$ we similarly have either $\mathrm{H}^m(X) = \mathbb{Z} \cdot b$ and $\mathrm{H}^m(Y) = 0$ or $\mathrm{H}^m(X) = 0$ and $\mathrm{H}^m(Y) = \mathbb{Z} \cdot b$ where b corresponds to $[S^m]^*$. The second possibility contradicts the known cup-product structure since then $x \cup y$ would be trivial in $\mathrm{H}^*(X \vee Y)$.
It remains to show that $\pi_1(Y)$ is trivial: If $n = m = 1$, we know the claim by Example 3.2. Otherwise, we can apply Proposition 1.3 to deduce $\pi_1(S^n \times S^m) \cong \pi_1(X) * \pi_1(Y)$. If $n, m \neq 1$ it immediately follows that $\pi_1(Y) = \{e\}$. Else, $\pi_1(X) * \pi_1(Y) \cong \mathbb{Z}$. By Proposition 2.24 (2) we deduce that precisely one of $\pi_1(X)$ and $\pi_1(Y)$ is trivial and the other one is isomorphic to $\mathbb{Z}$. Since $\mathrm{H}^1(Y) = 0$ the Hurewicz Theorem implies that $\pi_1(Y)$ cannot be $\mathbb{Z}$. ∎

For the next examples we first consider the following theorem:

Theorem 3.15. *For $n \geq 3$, a non-empty connected closed n-dimensional smooth manifold is irreducible in $\mathcal{M}_n$ relative to homotopy spheres if*

(1) *it has contractible universal covering and torsion-free fundamental group*[18].

(2) *its universal covering is a homotopy sphere and its fundamental group is non-trivial.*

Proof. Let M be a non-empty connected closed n-dimensional smooth manifold and suppose $M \cong X \# Y$. As in Proposition 1.2 (4) we construct pushouts

$$\begin{array}{ccc} S^{n-1} & \longrightarrow & \hat{X} \\ \downarrow & & \downarrow \\ \hat{Y} & \longrightarrow & X\#Y \end{array} \qquad \begin{array}{ccc} S^{n-1} & \longrightarrow & \hat{X} \\ \downarrow & & \downarrow \\ \overline{B}^n & \longrightarrow & X \end{array} \qquad \begin{array}{ccc} S^{n-1} & \longrightarrow & \hat{Y} \\ \downarrow & & \downarrow \\ \overline{B}^n & \longrightarrow & Y \end{array}$$

Set $\pi_M := \pi_1(M), \pi_X := \pi_1(\hat{X}), \pi_Y := \pi_1(\hat{Y})$ and let $\widetilde{X}$ and $\widetilde{Y}$ be the universal coverings of $\hat{X}$ and $\hat{Y}$. Since $n \geq 3$, S^{n-1} is simply connected, hence by the Seifert–van Kampen Theorem $\pi_M \cong \pi_X * \pi_Y$. In Construction 1.5 we derived a pushout

$$\begin{array}{ccc} \coprod\limits_{\#\pi_M} S^{n-1} & \longrightarrow & \coprod\limits_{[\pi_M:\pi_X]} \widetilde{X} \\ \downarrow & & \downarrow \\ \coprod\limits_{[\pi_M:\pi_Y]} \widetilde{Y} & \longrightarrow & \widetilde{M} \end{array}$$

where the top and right map are the inclusion of the boundary.

(1) First assume that $\widetilde{M}$ is contractible and π_M torsion-free. Notice that π_M cannot be trivial since in this case M would be orientable by [Fri25: Corollary 82.16] and therefore $0 = \mathrm{H}_n(\widetilde{M}) = \mathrm{H}_n(M) \cong \mathbb{Z}$. Therefore, M is not a homotopy sphere.
We deduce from the Mayer–Vietoris sequence corresponding to the decomposition induced by the above pushout that $\mathrm{H}_i(\widetilde{X}) = \mathrm{H}_i(\widetilde{Y}) = 0$ for $i \neq 0, n-1$ and that the inclusions of the boundary induce an isomorphism

$$(*) \qquad \mathrm{H}_{n-1}\Big(\coprod_{\#\pi_M} S^{n-1}\Big) \to \mathrm{H}_{n-1}\Big(\coprod_{[\pi_M:\pi_X]} \widetilde{X}\Big) \oplus \mathrm{H}_{n-1}\Big(\coprod_{[\pi_M:\pi_Y]} \widetilde{Y}\Big)$$

Since $\pi_M \cong \pi_X * \pi_Y$ is torsion-free, it follows that π_X and π_Y are infinite or trivial.
Suppose both are infinite. In this case the coverings $\widetilde{X} \to X$ and $\widetilde{Y} \to Y$ have infinite degree and $\widetilde{X}, \widetilde{Y}$ are non-compact, i.e. $\mathrm{H}_n(\widetilde{X}, \partial\widetilde{X}) = \mathrm{H}_n(\widetilde{Y}, \partial\widetilde{Y}) = 0$. It follows that the inclusions $\partial\widetilde{X} \to \widetilde{X}$ and $\partial\widetilde{Y} \to \widetilde{Y}$ induce monomorphisms on H_{n-1}. But this contradicts $(*)$. Hence, precisely one of π_X, π_Y is trivial, wlog. $\pi_X = 1$, i.e. $\hat{X} = \widetilde{X}$. By $(*)$ and the above considerations $\mathrm{H}_i(\hat{X}) = 0$ for $i \geq 1$. By applying the Mayer–Vietoris sequence and the Seifert–van Kampen Theorem to the second pushout above it follows that X is a homotopy sphere.

(2) Now assume that $\widetilde{M}$ is a homotopy sphere and π_M is non-trivial. Then M is not a homotopy sphere. Further $\widetilde{M}$ is compact, so the universal covering $\widetilde{M} \to M$ has finite degree, i.e. π_M is finite. By Proposition 2.24 (1) π_M is irreducible in $\mathcal{G}$. Hence, wlog. $\pi_X = 1$ and $\pi_Y \cong \pi_M$.
We again consider the Mayer–Vietoris sequence resulting from the pushout describing $\widetilde{M}$. It follows that $\mathrm{H}_i(\widetilde{X}) = \mathrm{H}_i(\widetilde{Y}) = 0$ for $i \neq 0, n-1, n$. Since $\widetilde{X}$ and $\widetilde{Y}$ are manifolds with non-empty boundary, this also holds for $i = n$.

[18] The reader well versed in group cohomology knows of course that the torsion-free assumption is superfluous since a manifold with contractible universal covering is a finite dimensional classifying space for its fundamental group (see [Bro82: Corollary VIII.2.5] for details).

For $n-1$ consider the following diagram:

$$\begin{tikzcd}
\mathrm{H}_{n-1}\Big(\bigsqcup_{\#\pi_M} S^{n-1}\Big) \arrow[r, two heads] \arrow[dr] & \mathrm{H}_{n-1}\Big(\bigsqcup_{\#\pi_M} \widetilde{X}\Big) \oplus \mathrm{H}_{n-1}(\widetilde{Y}) \arrow[d, two heads] \\
 & \mathrm{H}_{n-1}\Big(\bigsqcup_{\#\pi_M} \widetilde{X}\Big)
\end{tikzcd}$$

The top arrow is surjective from the long exact sequence, the vertical arrow is the natural projection. The diagonal is induced by a disjoint union of boundary inclusions $S^{n-1} = \partial\widetilde{X} \to \widetilde{X}$. Hence, it is trivial by [Fri25: Proposition 171.18]. Then $\mathrm{H}_{n-1}(\widetilde{X}) = 0$. Hence, $\mathrm{H}_i(\hat{X}) = \mathrm{H}_i(\widetilde{X}) = 0$ for $i \geq 1$. As in (1) it follows that X is a homology sphere. ■

There is a more geometric way to obtain results similar to the above theorem: Using the notation from the proof, let $M \cong X\#Y$ be a manifold with universal covering $\mathbb{R}^n$ or S^n. The sphere S^{n-1} along which $\hat{X}$ and $\hat{Y}$ are glued in M lifts to the universal covering. If $n \neq 4$ we can apply the Generalized Schoenflies Theorem (see [Fri25: Theorem 156.7]) to deduce that this lift bounds a ball. Covering space considerations then imply that this ball must be $\widetilde{X}$ or $\widetilde{Y}$. Hence, $\hat{X}$ or $\hat{Y}$ is a ball and X or Y is a homotopy sphere.
With this in mind, one might hope for even more – namely that $\pi_{n-1}(M) = 0$ alone could imply that M is irreducible. This is not the case, but counterexamples only arise as rational homology spheres with finite fundamental group, see [Rub97].

Example 3.16. For $n \geq 3$, Theorem 3.15 shows that the following are irreducible in $\mathcal{M}_n$:
- Riemannian manifolds of constant sectional curvature as they have universal covering homeomorphic to S^n (for spherical manifolds) or $\mathbb{R}^n$ (for hyperbolic and Euclidean manifolds). see [Lee97: in particular Corollary 11.13] for definitions and details.
- Lens spaces as they are quotients of S^3.
- $\mathbb{RP}^n$ as it has universal covering S^n.

Having found a list of examples of irreducible manifolds, the question arises if we found all of them. In high dimensions, the answer is almost certainly no. By Example 3.2 the answer is yes in the 2-dimensional setting. It turns out that the answer is also yes in dimension 3 – at least when it comes to orientable manifolds:

Theorem 3.17. *The irreducible elements of $\mathcal{M}_3^{\mathrm{or}}$ are precisely*
- $S^1 \times S^2$
- *quotients of* S^3
- *manifolds with contractible universal covering*

Sketch of a proof. By Theorem 3.15 and Proposition 3.14 the given elements are irreducible.
Let M be an irreducible element of $\mathcal{M}_3^{\mathrm{or}}$. It follows that every smooth embedding $S^2 \hookrightarrow M$ that cuts M into two pieces must extend smoothly over $\overline{B}^3$ as one of those pieces.
First suppose $\pi_2(M)$ is non-trivial. By the Sphere Theorem there exists a smooth embedding $S^2 \hookrightarrow M$ representing a non-trivial element of $\pi_2(M)$. In particular, it does not extend smoothly over $\overline{B}^3$, and therefore cannot cut M into two pieces. It follows from Proposition 3.9 that M is diffeomorphic to $S^1 \times S^2$.
We can now restrict our attention to $\pi_2(M) = 0$. Consider the universal covering $\widetilde{M} \to M$. If $\pi_1(M)$ is finite, $\widetilde{M}$ is closed and simply connected. By the Hurewicz Theorem and Poincaré duality $\widetilde{M}$ is a homotopy sphere. The aforementioned resolution of the 3-dimensional Poincaré conjecture now implies that $\widetilde{M}$ is diffeomorphic to S^3 and M thereby a quotient of S^3.
If $\pi_1(M)$ is infinite, $\widetilde{M}$ is non-compact. Since $\pi_2(\widetilde{M}) \cong \pi_2(M) = 0$, we deduce from the Hurewicz Theorem that $\mathrm{H}_2(\widetilde{M}) = 0$. Since $\widetilde{M}$ is a non-compact 3-dimensional manifold, $\mathrm{H}_k(\widetilde{M}) = 0$ for $k \geq 3$. Iteratively applying the Hurewicz Theorem now shows that $\pi_k(\widetilde{M}) = 0$ for $k \geq 1$. The Whitehead Theorem then implies that $\widetilde{M}$ is contractible. ■

4. Monoids of knots: Existence of factorization

4.1. Other codimensions

In the last chapter, we decomposed manifolds using the connected sum operation. In Chapter 1 we introduced this operation not only for manifolds but also for knots and submanifolds. In this chapter we will study their decomposition.
Before restricting to the codimension 2 case, we want to explain why its factorization properties are the most interesting. Therefore, we begin by sketching more general theory:

Definition. For $1 \leq k < n$ we consider

- $\mathcal{M}^{\mathrm{or}}_{n,k}$, the set of orientation-preserving diffeomorphism classes of non-empty connected closed oriented (n,k)-dimensional smooth manifold pairs (X, A).
- $\mathcal{M}^{\mathrm{no}}_{n,k}$, the set of diffeomorphism classes of non-empty connected closed (n,k)-dimensional smooth manifold pairs (X, A) with X oriented and A non-orientable, together with $(S^n, S^k \times \{0\})$.
- $\mathcal{M}_{n,k} \coloneqq \mathcal{M}^{\mathrm{or}}_{n,k} \cup \mathcal{M}^{\mathrm{no}}_{n,k}$

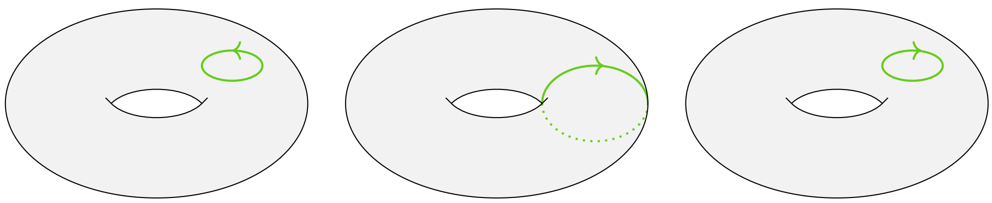
Analogously to the last chapter, these sets carry a monoid structure:

Lemma 4.1. *Let $1 \leq k < n$. The connected sum of manifold pairs defines a monoid structure on $\mathcal{M}_{n,k}$ with $\mathcal{M}^{\mathrm{or}}_{n,k}$ and $\mathcal{M}^{\mathrm{no}}_{n,k}$ as submonoids with the neutral element given by $(S^n, S^k \times \{0\})$.*

Proof. This is proved in essentially the same way as Lemma 3.1. ∎

In the lowest dimension, we again know this monoid from a classification:

Example 4.2. It follows from the classification of curves on surfaces (see [Fri25: Theorem 94.7]) that $\mathcal{M}^{\mathrm{or}}_{2,1}$ allows factorization with the irreducibles given by



The factorization of a separating curve is unique. For every genus $g \geq 1$ there exists a unique non-separating curve $\gamma \subseteq \Sigma_g$. The factorization of (Σ_g, γ) has g factors, at least one of which is the middle irreducible but the others can be arbitrary.

Having found new monoids, we of course want to consider their factorization properties: Luckily, these monoids come with homomorphisms to monoids we already understand quite well:

Proposition 4.3. *Let $1 \leq k < n$.*

(1) *We have monoid homomorphisms*

$$\begin{array}{rcl} \mathcal{M}_{n,k} & \to & \mathcal{M}_n \\ (X, A) & \mapsto & X \end{array} \quad \text{and} \quad \begin{array}{rcl} \mathcal{M}_{n,k} & \to & \mathcal{M}_k \\ (X, A) & \mapsto & A \end{array}$$

(2) *The monoids $\mathcal{M}_{n,k}$, $\mathcal{M}^{\mathrm{or}}_{n,k}$ and $\mathcal{M}^{\mathrm{no}}_{n,k}$ allow factorization relative to pairs of homotopy spheres, i.e. the submonoid $\{(X, A) \in \mathcal{M}_{n,k} \mid X, A \text{ homotopy spheres}\}$ which is closed under division.*

Proof.
(1) Follows from Proposition 1.8 (2).
(2) Precompose the complexity function on $\mathcal{M}_n$ and $\mathcal{M}_k$ from Theorem 3.5 with the homomorphisms from (1) to obtain complexity functions on $\mathcal{M}_{n,k}$. Combining them using Proposition 2.7 yields complexity functions on $\mathcal{M}_{n,k}$ with $\widetilde{\mathcal{M}}_{n,k}$ given by pairs of homotopy spheres. The claim follows from Proposition 2.9.
The argument for $\mathcal{M}^{\mathrm{or}}_{n,k}$ and $\mathcal{M}^{\mathrm{no}}_{n,k}$ is analogous. ∎

As for manifolds, results surrounding the work of S. Smale on h-cobordisms again yield that in most dimensions pairs of homotopy spheres are invertible – unless the codimension is 2:

Theorem 4.4. *Let $1 \leq k < n$. Then*

$$\mathcal{M}^*_{n,k} \subseteq \{(X, A) \in \mathcal{M}_{n,k} \mid X, A \textit{ homotopy spheres}\}$$

This is an equality if $k \neq 3, 4$ and
- *$n - k \geq 3$ or*
- *$n - k = 1$*

In these cases, $\mathcal{M}_{n,k}$ and $\mathcal{M}^{\mathrm{or}}_{n,k}$ allow factorization. The monoid $\mathcal{M}^{\mathrm{no}}_{n,k}$ always allows factorization.

Proof. The given inclusion follows from Proposition 4.3 (2). For the reverse inclusion, we consider the different cases separately, further distinguishing between high and low dimensions:

Case 1: $n - k \geq 3$, $k \geq 5$
This follows from a relative version of the h-cobordism Theorem, see [Sma62: Theorem 1.4] and [Hae62].

Case 2: $n - k \geq 3$, $k = 1, 2$
By [Hae61] all elements of $\mathcal{M}_{n,k}$ are equivalent to the neutral element.

Case 3: $n - k = 1$, $k \geq 5$
Let $(X, A) \in \mathcal{M}_{n,k}$ with X and A homotopy spheres. We can deduce as in Theorem 3.6 that $(X, A)\#(\overline{X}, \overline{A})$ is diffeomorphic to a smoothly embedded S^k in S^n. Then $(X, A)\#(\overline{X}, \overline{A})$ is the neutral element by the Generalized Schoenflies Theorem (see [Fri25: Theorem 156.7]).

Case 4: $n - k = 1$, $k = 1, 2$
Let $(X, A) \in \mathcal{M}_{n,k}$ with X and A homotopy spheres. By the resolution of the Poincaré conjecture in the relevant dimensions, (X, A) is a smoothly embedded S^k in S^n and therefore the neutral element by the Generalized Schoenflies Theorem.

The statements about monoids allowing factorization now follow as in Corollary 3.7. ∎

We have now established that in most dimensions with codimension not 2, ‘knots’ form a group under connected sum. In codimension 2, this is far from being true – in fact, we will see in Theorem 4.20 (2) that the only invertible element is the unknot.
To further study the codimension 2 case, we restrict ourselves to the classical setting of knotted spheres in spheres and replace our equivalence relation again by isotopy as described in Section 1.3:

Definition. For $n \geq 1$ let $\mathcal{K}_n$ be the set of smooth isotopy classes of oriented n-dimensional knots.

Lemma 4.5. *The connected sum operation defines a monoid structure on $\mathcal{K}_n$ with the neutral element given by the unknot.*

Proof. By Theorem 1.14 this follows from Lemma 4.1. ∎

4.2. The universal abelian covering

Before returning to this monoid, we develop some more knot theory. In the last chapter, we have successfully applied algebraic invariants of manifolds to deduce factorization properties. If we now want to apply algebraic invariants to knots, we will – as previously – have to turn to knot complements. Unfortunately, we cannot naively apply the same techniques again:

- By Proposition 1.22, the fundamental group of a connected sum of knots is an amalgamated free product of the groups of the summands. So it does not define a homomorphism to $\mathcal{G}$.[19]
- By Proposition 1.11, the homology groups of the complement of a knot do not tell us anything new at all.

Hence, we will need a new space to apply our invariants to:

Definition. Let X be a connected smooth manifold. The covering of X corresponding to the commutator subgroup $[\pi_1(X), \pi_1(X)] \subseteq \pi_1(X)$ is the *universal abelian covering* $\widetilde{X}_{\mathrm{ab}} \to X$.

Lemma 4.6. *Let X be a connected smooth manifold. The universal abelian covering $\widetilde{X}_{\mathrm{ab}} \to X$ exists and its group of deck transformations is naturally isomorphic to $\pi_1(X)_{\mathrm{ab}}$.*

Proof. see [Fri25: Proposition 114.10] ■

For knot complements, we can give a very explicit description of the universal abelian covering:

Construction 4.7. Let $K \subseteq S^{n+2}$ be an oriented n-dimensional knot. Let $\Sigma \subseteq S^{n+2}$ be a connected compact oriented $(n+1)$-dimensional smooth submanifold such that $K = \partial\Sigma$ with the boundary orientation[20]. By [Fri25: Theorem 52.2] there exists a bicollar $\beta\colon (\Sigma\setminus\partial\Sigma)\times(-1,1) \to S^{n+2}\setminus K$, i.e. an orientation-preserving smooth embedding such that $\beta|_{(\Sigma\setminus\partial\Sigma)\times\{0\}}$ is the identity. Set

$$N \coloneqq \beta((\Sigma\setminus\partial\Sigma)\times(-1,1)) \qquad N^+ \coloneqq \beta((\Sigma\setminus\partial\Sigma)\times(0,1)) \qquad N^- \coloneqq \beta((\Sigma\setminus\partial\Sigma)\times(-1,0))$$

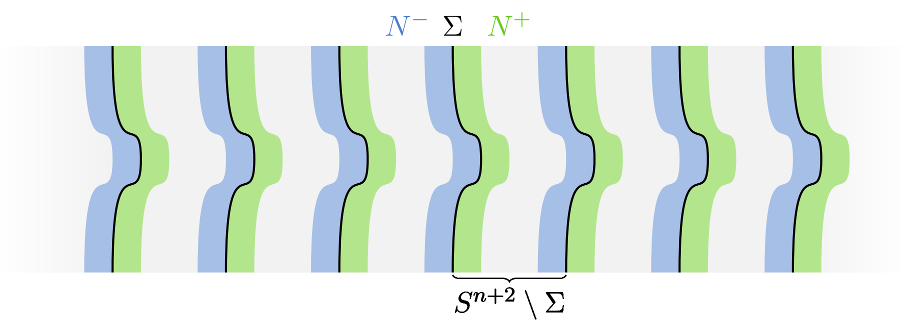
and form the 'disjoint unions' $(N\setminus\Sigma)\times\mathbb{Z}$, $N\times\mathbb{Z}$ and $(S^{n+2}\setminus\Sigma)\times\mathbb{Z}$. Consider the map

$$\begin{aligned} (N\setminus\Sigma)\times\mathbb{Z} &\to (S^{n+2}\setminus\Sigma)\times\mathbb{Z} \\ (n,i) &\mapsto \begin{cases} (n,i), & \text{if } n\in N^+ \\ (n,i-1), & \text{if } n\in N^- \end{cases} \end{aligned}$$

in the pushout

$$\begin{tikzcd} (N\setminus\Sigma)\times\mathbb{Z} \arrow[r] \arrow[d] & (S^{n+2}\setminus\Sigma)\times\mathbb{Z} \arrow[d] \\ N\times\mathbb{Z} \arrow[r] & X \end{tikzcd}$$

where the left map is the inclusion.



[19] Also note that by [Mae77] that the naive hope that $d(A *_{\mathbb{Z}} B) = d(A) + d(B) - 1$ is incorrect – even when restricting to knot groups glued along meridians.

[20] Such a Σ can always be found, see [Fri25: Propositions 266.11, 220.10].

There is an obvious $\mathbb{Z}$-action on the three spaces and they glue together to a $\mathbb{Z}$-action on X. The quotient of this action is the pushout of the quotients, i.e. $S^{n+2} \setminus K$ by [Fri25: Lemma 6.37]. This gives a covering $f\colon X \to S^{n+2} \setminus K$ and $f_*(\pi_1(X)) \subseteq \pi_1(S^{n+2} \setminus K)$ is a normal subgroup with quotient $\mathbb{Z}$ (see [Fri25: Corollary 114.12]). By the universal property of the abelianization there exists an epimorphism

$$\pi_1(S^{n+2} \setminus K)_{\mathrm{ab}} \twoheadrightarrow \pi_1(S^{n+2} \setminus K)/f_*(\pi_1(X)) \cong \mathbb{Z}$$

By Proposition 1.21 $\pi_1(S^{n+2}\setminus K)_{\mathrm{ab}} \cong \mathbb{Z}$ so this map is an isomorphism, $f_*(\pi_1(X)) \subseteq \pi_1(S^{n+2}\setminus K)$ is the commutator subgroup and $X \to S^{n+2} \setminus K$ the universal abelian covering.

In the next proposition we determine the effect of the connected sum operation on the universal abelian covering of knot complements:

Proposition 4.8. *Let $n \geq 1$ and $K, L \subseteq S^{n+2}$ be oriented n-dimensional knots. We have a commutative cube*

$$\begin{tikzcd}
\mathbb{R} \times B^{n+1} \arrow[rr, "\widetilde{\varphi}"] \arrow[dr] \arrow[dd, "\widetilde{\psi}"'] & & (\widetilde{S^{n+2} \setminus K})_{\mathrm{ab}} \arrow[dr] \arrow[dd] & \\
& B^{n+2} \setminus B^n \arrow[rr, "\varphi"] \arrow[dd, "\psi"] & & S^{n+2} \setminus K \arrow[dd] \\
(\widetilde{S^{n+2} \setminus L})_{\mathrm{ab}} \arrow[rr] \arrow[dr] & & (\widetilde{S^{n+2} \setminus K\#L})_{\mathrm{ab}} \arrow[dr] & \\
& S^{n+2} \setminus L \arrow[rr] & & S^{n+2} \setminus K\#L
\end{tikzcd}$$

where
- *the front and back faces are pushouts,*
- *the diagonal maps are the universal abelian coverings,*
- *the maps φ and ψ are the embeddings from the definition of a connected sum and*
- *the maps $\widetilde{\varphi}$ and $\widetilde{\psi}$ are the lifts of φ and ψ along the universal abelian coverings. They are embeddings.*

The induced maps on deck transformation groups between the four universal abelian coverings are isomorphisms.

Proof. The front face is the pushout from Proposition 1.10. Use it to view all spaces in it as subspaces of $S^{n+2} \setminus K\#L$. Let $p\colon (\widetilde{S^{n+2} \setminus L})_{\mathrm{ab}} \to S^{n+2} \setminus K\#L$ be the universal abelian covering. By elementary set theory

$$(\widetilde{S^{n+2} \setminus L})_{\mathrm{ab}} = p^{-1}(S^{n+2} \setminus K) \cup p^{-1}(S^{n+2} \setminus L)$$

and $p^{-1}(S^{n+2} \setminus K) \cap p^{-1}(S^{n+2} \setminus L) = p^{-1}(B^{n+2} \setminus B^n)$. Let $X \in \{S^{n+2} \setminus K, S^{n+2} \setminus L, B^{n+2} \setminus B^n\}$ We have a commutative diagram

$$\begin{tikzcd}
\pi_1(X) \arrow[r] \arrow[d, two heads] & \pi_1(S^{n+2} \setminus K\#L) \arrow[d, two heads] \\
\pi_1(X)_{\mathrm{ab}} \arrow[r, "\cong"] & \pi_1(S^{n+2} \setminus K\#L)_{\mathrm{ab}}
\end{tikzcd}$$

where the bottom map is an isomorphism by Proposition 1.22 and Proposition 1.21. Hence, [Fri25: Proposition 115.7] implies that
- $p\colon p^{-1}(X) \to X$ is a connected covering corresponding to the subgroup

$$\pi_1(X) \cap [\pi_1(S^{n+2} \setminus K\#L) : \pi_1(S^{n+2} \setminus K\#L)] \underset{\substack{\uparrow \\ \text{above diagram}}}{=} [\pi_1(X) : \pi_1(X)]$$

  Hence, $p\colon p^{-1}(X) \to X$ is the universal abelian covering
- the inclusions induce an isomorphism $\mathrm{Deck}(p|_{p^{-1}(X)}) \cong \mathrm{Deck}(p)$

The claim now follows from [Fri25: Lemma 6.37]. ∎

Corollary 4.9. *Let $n \geq 1$ and $K, L \subseteq S^{n+2}$ be oriented n-dimensional knots.*

(1) $\pi_1((\widetilde{S^{n+2} \setminus K\#L})_{\mathrm{ab}}) \cong \pi_1((\widetilde{S^{n+2} \setminus K})_{\mathrm{ab}}) * \pi_1((\widetilde{S^{n+2} \setminus L})_{\mathrm{ab}})$

(2) *for $i \geq 1$,* $\mathrm{H}_i((\widetilde{S^{n+2} \setminus K\#L})_{\mathrm{ab}}) \cong \mathrm{H}_i((\widetilde{S^{n+2} \setminus K})_{\mathrm{ab}}) \oplus \mathrm{H}_i((\widetilde{S^{n+2} \setminus L})_{\mathrm{ab}})$

Proof. Since $\mathbb{R} \times B^{n+1}$ is contractible this follows by applying the Seifert–van Kampen Theorem and the Mayer–Vietoris sequence to Proposition 4.8. ■

The significance of studying the universal abelian covering comes from the observation that in most dimensions, the homotopy type of the knot complement – and therefore also the universal abelian covering – can detect the unknot:

Theorem 4.10. *Let $n \geq 1$ and $K \subseteq S^{n+2}$ be an n-dimensional knot. The following are equivalent:*

(i) $S^{n+2} \setminus K \simeq S^1$

(ii) $(\widetilde{S^{n+2} \setminus K})_{\mathrm{ab}}$ *is contractible*

(iii) $\pi_1(S^{n+2} \setminus K) \cong \mathbb{Z}$ *and* $\mathrm{H}_i((\widetilde{S^{n+2} \setminus K})_{\mathrm{ab}}) = 0$ *for $i \geq 1$.*

These are furthermore implied by

(iv) *K is trivial*

If $n \neq 2$, all four are equivalent.

Proof.

(i) ⇒ (iii): In this case, $\pi_1(S^{n+2} \setminus K) \cong \pi_1(S^1)$ is abelian so the universal abelian covering is the universal covering and therefore simply connected. Since coverings induce isomorphisms on higher homotopy groups (see [Fri25: Proposition 120.22]), it follows that all homotopy groups of $(\widetilde{S^{n+2} \setminus K})_{\mathrm{ab}}$ are trivial. Then (iii) follows from the Hurewicz Theorem.

(iii) ⇒ (i): Conversely, (iii) implies in the same manner that the higher homotopy groups of $S^{n+2} \setminus K$ are trivial and $\pi_1(S^{n+2} \setminus K)$ is abelian. By Lemma 1.20 the inclusion of a meridian then induces an isomorphism on all homotopy groups. The claim follows from Whitehead's Theorem.

(ii) ⇒ (iii): If $(\widetilde{S^{n+2} \setminus K})_{\mathrm{ab}}$ is contractible, its homology and fundamental group vanish. Then statement (iii) follows from Proposition 1.21.

(iii) ⇒ (ii): Since $\pi_1(S^{n+2} \setminus K) \cong \mathbb{Z}$ is abelian, the universal abelian covering of $S^{n+2} \setminus K$ is the universal covering. The Hurewicz Theorem implies that all its homotopy groups are trivial and (ii) therefore follows from Whitehead's Theorem.

(iv) ⇒ (i): Consider the n-dimensional unknot

$$K \coloneqq \{(x_0, \ldots, x_{n+2}) \in S^{n+2} \subseteq \mathbb{R}^{n+3} \mid x_{n+1}, x_{n+2} = 0\}$$

The map

$$\begin{aligned} B^{n+1} \times S^1 &\to S^{n+2} \setminus K \\ (x, y) &\mapsto \left(x, \sqrt{1 - \|x\|^2} \cdot y\right) \end{aligned}$$

is a diffeomorphism.

(i) ⇒ (iv): As expected, this is a non-trivial theorem. The classical case $n = 1$ follows from Dehn's Lemma, see [Pap57: Theorem 28.1].
The case $n \geq 4$ was proven by J. Levine [Lev65: Theorem 3].
For $n = 3$ the claim was proven independently by J. Shaneson [Sha68: Theorem 1.1] and C. Wall [Wal65: Corollary 3.1][21]. ■

Whether the homotopy type of the complement can detect the unknot in dimension 2 is still an open problem. It is known topologically by work of M. Freedman [FQ90: Theorem 11.7A].

[21] Unfortunately, neither of these references contains a complete proof: J. Shaneson relies on an unpublished result from his thesis, C. Wall on a conjecture which he only resolved later. A proof of the result can be found in [Wal99: Section 16].

4.3. Alexander modules

On first glance it seems like we are in very good shape: Corollary 4.9 shows that the fundamental group and homology of the universal abelian covering of knot complements behave well under connected sum and Theorem 4.10 shows that these invariants can mostly detect the unknot. So one might hope that we can now just apply the same techniques as for manifolds to obtain existence of factorizations. Alas, this is not quite possible: In Proposition 3.3 we crucially needed to show that the algebraic invariants are finitely generated, which we did with a compactness argument. But the universal abelian covering of a knot complement has infinite degree and is therefore non-compact, so the invariants need not be finitely generated.[22] We therefore need to use the additional structure of the universal abelian covering – namely, the action by deck transformations – to obtain finiteness results:

Definition. For a commutative ring R, let $R[t^{\pm 1}]$ be the ring of *Laurent polynomials over* R.

Lemma 4.11. *Let X be a topological space with a $\mathbb{Z}g$-action and $i \geq 0$. Letting t act as*

$$\begin{aligned} \mathrm{C}_i(X) &\to \mathrm{C}_i(X) \\ \sigma\colon \Delta^i \to X &\mapsto (g \cdot \sigma)\colon \Delta^i \to X \end{aligned}$$

and extending to $\mathbb{Z}[t^{\pm 1}]$ defines a $\mathbb{Z}[t^{\pm 1}]$-module structure on $\mathrm{C}_i(X)$ that descends to $\mathrm{H}_i(X)$.

Proof. The module structure on $\mathrm{C}_i(X)$ is well-defined since $\mathrm{C}_i(X)$ is freely generated by i-simplices as an abelian group. To see that it descends to $\mathrm{H}_i(X)$, note that for $\tau \in \mathrm{C}_{i+1}(X)$, $g \cdot \partial\tau = \partial(g \cdot \tau)$. ■

This $\mathbb{Z}[t^{\pm 1}]$-module structure not only depends on the $\mathbb{Z}$-action on X, but also on a fixed choice of generator for $\mathbb{Z}$. Thankfully, in the setting we are most interested in there is a canonical choice of generator:

Definition. Let $n \geq 1$ and $K \subseteq S^{n+2}$ be an oriented n-dimensional knot with a meridian μ. Lemma 4.6 and Proposition 1.21 give a $\mathbb{Z}\mu$ action on $(\widetilde{S^{n+2} \setminus K})_{\mathrm{ab}}$. For $i \geq 0$, we consider $\mathrm{C}_i((\widetilde{S^{n+2} \setminus K})_{\mathrm{ab}})$ and $\mathrm{H}_i((\widetilde{S^{n+2} \setminus K})_{\mathrm{ab}})$ with the $\mathbb{Z}[t^{\pm 1}]$-module structure resulting from Lemma 4.11. These $\mathbb{Z}[t^{\pm 1}]$-modules $\mathrm{H}_i((\widetilde{S^{n+2} \setminus K})_{\mathrm{ab}})$ are the *Alexander modules* of K.

With this module structure, the homology of the universal abelian covering of a knot is indeed finitely generated:

Lemma 4.12. *Let $n \geq 1$, $K \subseteq S^{n+2}$ be an oriented n-dimensional knot and $i \geq 0$.*

(1) *The $\mathbb{Z}[t^{\pm 1}]$-module structure on $\mathrm{H}_i((\widetilde{S^{n+2} \setminus K})_{\mathrm{ab}})$ does not depend on the choice of meridian.*

(2) *The $\mathbb{Z}[t^{\pm 1}]$-module $\mathrm{H}_i((\widetilde{S^{n+2} \setminus K})_{\mathrm{ab}})$ is finitely generated.*

Proof.

(1) This follows from Lemma 1.20 and Lemma 4.6.

(2) Let μ be a meridian of K. In Construction 4.7 we found a pushout

$$\begin{tikzcd} (N \setminus \Sigma) \times \mathbb{Z} \arrow[r] \arrow[d] & (S^{n+2} \setminus \Sigma) \times \mathbb{Z} \arrow[d] \\ N \times \mathbb{Z} \arrow[r] & (\widetilde{S^{n+2} \setminus K})_{\mathrm{ab}} \end{tikzcd}$$

and that this pushout respects the $\mathbb{Z}\mu$-action on the four spaces. We consider the homology groups of the four spaces as $\mathbb{Z}[t^{\pm 1}]$-modules. The continuous maps between them then induce $\mathbb{Z}[t^{\pm 1}]$-module homomorphisms on homology.

[22] For a classical knot $K \subseteq S^3$ this is the case if and only if K is fibred by work of J. Stallings [Sta61: Theorem 2].

Observe that $N \setminus \Sigma \simeq \Sigma \sqcup \Sigma$ and $N \simeq \Sigma$, hence their homology groups are finitely generated over $\mathbb{Z}$. By [Fri25: Proposition 165.11], this also holds for $S^{n+2} \setminus \Sigma$. Therefore the homology groups of $(N \setminus \Sigma) \times \mathbb{Z}$, $N \times \mathbb{Z}$ and $(S^{n+2} \setminus \Sigma) \times \mathbb{Z}$ are finitely generated as a $\mathbb{Z}[t^{\pm 1}]$-modules. Now consider the Mayer–Vietoris sequence corresponding to the decomposition given by the pushout. Note that the maps arising in this sequence are homomorphisms of $\mathbb{Z}[t^{\pm 1}]$-modules since the module structure was defined on chain level. The claim follows since $\mathbb{Z}[t^{\pm 1}]$ is noetherian (see [Lan02: Exercise II.4, Propositions X.1.1, 2, 6]). ■

With a suitable finiteness condition established, we are one step closer to adapting our argument for the existence of factorizations from manifolds to knots. The next input we need is a complexity functions on the monoid of modules over $\mathbb{Z}[t^{\pm 1}]$. We begin with the following general observation:

Proposition 4.13. *Let R be a principal ideal domain and $\mathcal{M}$ the monoid of finitely generated R-modules under direct sum. There exists a complexity function $c \colon \mathcal{M} \to \mathbb{N}_0$ with $\widetilde{M} = \{0\}$.*

Proof. By the Structure Theorem of finitely generated modules over principal ideal domains (see [Lan02: Theorems III.7.3, 5]), the monoid $\mathcal{M}$ is a unique factorization monoid. The claim follows from Proposition 2.10. ■

Unfortunately, $\mathbb{Z}[t^{\pm 1}]$ is not a principal ideal domain – but its rational extension $\mathbb{Q}[t^{\pm 1}]$ is. In light of Lemma 2.12 this leaves us to deal with the torsion part. We start with a seemingly unrelated definition:[23]

Definition. A $\mathbb{Z}[t^{\pm 1}]$-module is *of type K* if multiplication by $t-1$ induces an automorphism. Denote by $\mathcal{M}_K$ the monoid of finitely generated $\mathbb{Z}[t^{\pm 1}]$-modules of type K with the operation given by direct sum.

Proposition 4.14. *Let $n, i \geq 1$. The map*

$$\begin{aligned} \mathcal{K}_n &\to \mathcal{M}_K \\ [K] &\mapsto \mathrm{H}_i(\widetilde{(S^{n+2} \setminus K)}_{\mathrm{ab}}) \end{aligned}$$

is a well-defined monoid homomorphism.

Proof. We begin by showing the map is well-defined. Let $K \subseteq S^{n+2}$ be an n-dimensional knot. By Lemma 4.12 (2), $\mathrm{H}_i(\widetilde{(S^{n+2} \setminus K)}_{\mathrm{ab}})$ is finitely generated, so we need to show it is of type K: We have a short exact sequence of chain complexes

$$0 \to \mathrm{C}_*(\widetilde{(S^{n+2} \setminus K)}_{\mathrm{ab}}) \xrightarrow{\cdot(t-1)} \mathrm{C}_*(\widetilde{(S^{n+2} \setminus K)}_{\mathrm{ab}}) \xrightarrow{p_*} \mathrm{C}_*(S^{n+2} \setminus K) \to 0$$

since

- multiplication by $t-1$ defines a monomorphism $\mathrm{C}_i(\widetilde{(S^{n+2} \setminus K)}_{\mathrm{ab}}) \to \mathrm{C}_i(\widetilde{(S^{n+2} \setminus K)}_{\mathrm{ab}})$ as this $\mathbb{Z}[t^{\pm 1}]$-module is free and $\mathbb{Z}[t^{\pm 1}]$ is an integral domain,
- the homomorphism p_* is surjective as simplices are simply connected and can thereby lifted along any covering,
- and exactness in the middle follows from the following observations: For a simplex $\sigma \colon \Delta^i \to \widetilde{X}$, $p \circ (t \cdot \sigma) = p \circ \sigma \colon \Delta^i \to X$ showing $\operatorname{im}(\cdot(t-1)) \subseteq \ker(p_*)$.

  For the converse, observe that every $x \in \mathrm{C}_i(\widetilde{(S^{n+2} \setminus K)}_{\mathrm{ab}})$ can be written as

  $$x = \sum_{a=1}^{m} \sum_{i \in \mathbb{Z}} \lambda_{a,i} \cdot t^i \cdot \sigma_a$$

  for $m \geq 0$, $\sigma_1, \ldots, \sigma_m \colon \Delta^i \to X$ with $p \circ \sigma_1, \ldots, p \circ \sigma_m$ pairwise distinct and $\lambda_{a,i} \in \mathbb{Z}$ for $a \in \{1, \ldots, m\}$, $i \in \mathbb{Z}$ almost all 0.
  If $x \in \ker(p_*)$

  $$\sum_{i \in \mathbb{Z}} \lambda_{a,i} = 0$$

  for all $a \in \{1, \ldots, m\}$ which implies that $x \in \operatorname{im}(\cdot(t-1))$.

[23]The presented proof is from [CKS04: Lemma 3.11].

From this short exact sequence of chain complexes, we obtain a long exact sequence in homology

$$\mathrm{H}_{i+1}(S^{n+2}\setminus K) \to \mathrm{H}_i(\widetilde{(S^{n+2}\setminus K)_{\mathrm{ab}}}) \xrightarrow{\cdot(t-1)} \mathrm{H}_i(\widetilde{(S^{n+2}\setminus K)_{\mathrm{ab}}}) \to \mathrm{H}_i(S^{n+2}\setminus K)$$

For $i \geq 2$ the claim now follows directly from Proposition 1.11. For $i = 1$, we have

$$\underbrace{\mathrm{H}_2(S^{n+2}\setminus K)}_{=0} \to \mathrm{H}_1(\widetilde{(S^{n+2}\setminus K)_{\mathrm{ab}}}) \xrightarrow{\cdot(t-1)} \mathrm{H}_1(\widetilde{(S^{n+2}\setminus K)_{\mathrm{ab}}}) \to \underbrace{\mathrm{H}_1(S^{n+2}\setminus K)}_{\cong\mathbb{Z}} \to$$
$$\to \underbrace{\mathrm{H}_0(\widetilde{(S^{n+2}\setminus K)_{\mathrm{ab}}})}_{\cong\mathbb{Z}} \to \underbrace{\mathrm{H}_0(\widetilde{(S^{n+2}\setminus K)_{\mathrm{ab}}})}_{\cong\mathbb{Z}} \xrightarrow{p_*} \underbrace{\mathrm{H}_0(S^{n+2}\setminus K)}_{\cong\mathbb{Z}} \to 0$$

Since p_* is an isomorphism, the connecting homomorphism also must be one, implying the claim. With this we have shown that the map is well-defined. By Theorem 4.10, the unknot gets mapped to 0. Applying the Mayer–Vietoris sequence to Proposition 4.8 yields that it is a homomorphism. ■

The torsion subgroup of a $\mathbb{Z}[t^{\pm 1}]$-module of type K is always finite:[24]

Lemma 4.15. *The torsion subgroup of a finitely generated $\mathbb{Z}[t^{\pm 1}]$-module of type K is finite. Its cardinality defines a multiplicative complexity function $\mathcal{M}_K \to \mathbb{N}$.*

Proof. Let M be a finitely generated $\mathbb{Z}[t^{\pm 1}]$-module of type K. The submodule $\mathrm{Tor}(M) \subseteq M$ is also finitely generated, since $\mathbb{Z}[t^{\pm 1}]$ is noetherian (see [Lan02: Exercise II.4, Proposition X.1.6]). Further, $\mathrm{Tor}(M)$ is of type K, since multiplication by $t-1$ is an automorphism of M and hence perserves the order of elements. Hence, it suffices to prove the claim for torsion M:
Then M is finitely generated over $\mathbb{Z}[t^{\pm 1}]$ and torsion, hence there exists $\lambda \geq 1$ with $\lambda \cdot M = 0$. We proceed inductively on the number of prime factors of λ:
If λ is prime, we can regard M as a module over the principal ideal domain $(\mathbb{Z}/\lambda)[t^{\pm 1}]$ (see [Lan02: Exercise II.4, Theorem IV.1.2]). By the Structure Theorem of finitely generated modules over principal ideal domains (see [Lan02: Theorems III.7.3,5]), M is the sum of cyclic $(\mathbb{Z}/\lambda)[t^{\pm 1}]$-modules. Multiplication by $t-1$ respects this decomposition. But any cyclic $(\mathbb{Z}/\lambda)[t^{\pm 1}]$-module is finite or free – with the latter being excluded since $\cdot(t-1)$ is an automorphism of M but not of $(\mathbb{Z}/\lambda)[t^{\pm 1}]$.
Now suppose that λ has more than one prime factor and let p be one of them. By induction pM and M/pM are finite, so M must be finite. ■

With this we can complete adapting our argument from manifolds to knots and obtain the following:

Proposition 4.16.
(1) *The monoid $\mathcal{M}_K$ allows factorization.*
(2) *Let $n \geq 1$. The monoid $\mathcal{K}_n$ allows factorization relative to*

$$\{K \in \mathcal{K}_n \mid \forall i \geq 1 : \mathrm{H}_i(\widetilde{(S^{n+2}\setminus K)_{\mathrm{ab}}}) = 0\}$$

Proof.
(1) Let $\mathcal{M}_{\mathbb{Q}[t^{\pm 1}]}$ be the monoid of finitely generated $\mathbb{Q}[t^{\pm 1}]$-modules under direct sum. We have a monoid homomorphism

$$\begin{array}{rccc} \varphi\colon & \mathcal{M}_K & \to & \mathcal{M}_{\mathbb{Q}[t^{\pm 1}]} \\ & M & \mapsto & M \otimes \mathbb{Q} \end{array}$$

By Lemma 2.12, the kernel of φ is given by torsion groups. It follows from Proposition 4.13 that there exists a complexity function on $\mathcal{M}_K$ with $\widetilde{\mathcal{M}_K}$ given by torsion groups ($\mathbb{Q}[t^{\pm 1}]$ is a principal ideal domain by [Lan02: Exercise II.4, Theorem IV.1.2]). Combining this complexity function with the complexity function from Lemma 4.15 using Proposition 2.7, yields a complexity function on $\mathcal{M}_K$ with $\widetilde{\mathcal{M}_K} = \{0\}$. The claim follows from Proposition 2.9.

[24] The argument presented here is from [Lev77: Lemma 3.1].

(2) Let for $i \geq 1$

$$\begin{aligned} \varphi_i\colon \mathcal{K}_n &\to \mathcal{M}_K \\ K &\mapsto \mathrm{H}_i(\widetilde{(S^{n+2} \setminus K)}_{\mathrm{ab}}) \end{aligned}$$

be the monoid homomorphism from Proposition 4.14. Composing these with the complexity function from (1) yields complexity functions on $\mathcal{K}_n$. Combining them using Proposition 2.7 yields a complexity function on $\mathcal{K}_n$ with $\widetilde{\mathcal{K}_n}$ given by

$$\{K \in \mathcal{K}_n \mid \forall i \in \{1, \ldots, n+1\} : \mathrm{H}_i(\widetilde{(S^{n+2} \setminus K)}_{\mathrm{ab}}) = 0\}$$

Since $\widetilde{(S^{n+2} \setminus K)}_{\mathrm{ab}}$ is a non-compact $(n+1)$-dimensional manifold this is equal to the submonoid from the claim – which therefore follows from Proposition 2.9. ■

4.4. The fundamental group

Comparing Proposition 4.16 (2) with Theorem 4.10 there is still room for optimization: We proved that $\mathcal{K}_n$ allows factorization relative to knots where the homology of the universal abelian covering is the same as for the unknot, but to prove that $\mathcal{K}_n$ allows factorization, we will also need to control its fundamental group.

As in the last section, $G := \pi_1(\widetilde{(S^{n+2} \setminus K)}_{\mathrm{ab}})$ is not finitely generated. But it again comes with a $\mathbb{Z}\mu$-action via deck transformations. This does not define a $\mathbb{Z}[t^{\pm 1}]$-module structure on G, as G is non-abelian – but it is still finitely generated in the sense that

$$d_{\mathbb{Z}}(G) := \{\#S \mid S \subseteq G,\ \mathbb{Z} \cdot S \text{ generates } G\}$$

is finite. Now one might hope that

$$d_{\mathbb{Z}}(\pi_1(\widetilde{(S^{n+2} \setminus K\#L)}_{\mathrm{ab}})) = d_{\mathbb{Z}}(\pi_1(\widetilde{(S^{n+2} \setminus K)}_{\mathrm{ab}})) + d_{\mathbb{Z}}(\pi_1(\widetilde{(S^{n+2} \setminus L)}_{\mathrm{ab}}))$$

Alas, this is not the case: The counterexample from [Mae77] we discussed in Footnote 19 also contradicts this. With this in mind, there likely is no complexity function on $\mathcal{K}_n$ describing the fundamental group good enough. Luckily, we can get away with a weaker notion:

Definition. Let M be an abelian monoid and $S \subseteq M$ a submonoid that is closed under division. A map of sets $\varphi\colon M \to \mathbb{N}_0$ *bounds factors relative to* S if for all $m \in M$:

$$\underset{n \in \mathbb{N}_0}{\forall} \quad \underset{m_1, \ldots, m_n \in M \setminus S}{\forall} \quad m = m_1 \cdots m_n \quad \Rightarrow \quad n \leq \varphi(m)$$

Proposition 4.17. *Let M be an abelian monoid and $S \subseteq M$ a submonoid that is closed under division. Suppose there is a factor bounding map $\varphi\colon M \to \mathbb{N}_0$ relative to S. Then M allows factorization relative to S.*

Proof. Let $m \in M$. By definition

$$\max\{n \in \mathbb{N}_0 \mid \exists m_1, \ldots, m_n \in M \setminus S : m = m_1 \cdots m_n\} \leq \varphi(m)$$

In particular, this maximum is finite and realised by some $n \in \mathbb{N}_0$, $m_1, \ldots, m_n \in M$ with $m = m_1 \cdots m_n$. Then $m_1, \ldots, m_n$ need to be irreducible relative to S. ■

A complexity function induces a factor bounding map and we can combine such maps:

Proposition 4.18. *Let M be an abelian monoid.*

(1) *A complexity function c on M bounds factors relative to $\widetilde{M}$.*

(2) *Let $\varphi_1, \ldots, \varphi_k\colon M \to \mathbb{N}_0$ bound factors relative to $S_1, \ldots, S_n$. Then*

$$\varphi := \frac{1}{k}(\varphi_1 + \cdots + \varphi_k)\colon M \to \mathbb{N}_0$$

bounds factors relative to $S := S_1 \cap \cdots \cap S_k$.

Proof.

(1) Let $m_1, \dots, m_n \in M \setminus \widetilde{M}$. Then
$$c(m_1 \cdots m_n) \geq c(m_1) + \cdots + c(m_n) \geq n$$
since the product of natural numbers not equal to 1 is at least their their sum.

(2) Let $m_1, \dots, m_n \in M \setminus S$. Then
$$\varphi(m_1 \cdots m_n) = \frac{1}{k} \sum_{i=1}^{k} \varphi_i(m_1 \cdots m_n) \geq \frac{1}{k} \sum_{i=1}^{k} n = n$$
proving the claim. ∎

The core mathematical question is now where a factor bounding map on $\mathcal{K}_n$ relative to knots whose complement has fundamental group isomorphic to $\mathbb{Z}$ comes from. Algebraically the question is to bound the number of factors arising when decomposing the fundamental group of a knot complement as an amalgamated free product. Thankfully, Bass–Serre theory and accessibility provide a framework to obtain such bounds. We develop this in Appendix A.
With it we can prove the following statement:

Lemma 4.19. *Let $K \subseteq S^{n+2}$ be an oriented n-dimensional knot. There exists $k \in \mathbb{N}$ such that $K \cong K_1 \# \dots \# K_m$ with $\pi_1(S^{n+2} \setminus K_i) \ncong \mathbb{Z}$ for all $i \in \{1, \dots, m\}$ implies $m \leq k$.*

Proof. Construct a graph of groups
$$\begin{tikzcd} \overset{\pi_1(S^{n+2} \setminus K_1)}{\bullet} \arrow[r, dash, "\mathbb{Z}"'] & \overset{\pi_1(S^{n+2} \setminus K_2)}{\bullet} \arrow[r, dash] & \cdots \arrow[r, dash] & \overset{\pi_1(S^{n+2} \setminus K_{n-1})}{\bullet} \arrow[r, dash, "\mathbb{Z}"'] & \overset{\pi_1(K_n)}{\bullet} \end{tikzcd}$$
where the homomorphism from an edge to a vertex groups always is the inclusion of a meridian. These are monomorphisms by Proposition 1.21. By Proposition 1.22 and the inductive argument of Lemma A.8 (2) $\pi_1(S^{n+2} \setminus K)$ is isomorphic to the fundamental group of this graph of groups and the edge groups include to subgroups generated by meridians.
Hence, it remains to check the hypotheses of Theorem A.24:

- (i) and (iv) follow from Proposition 1.21. It also implies that the graph of groups is reduced since $\pi_1(S^{n+2} \setminus K_i) \ncong \mathbb{Z}$ for all $i \in \{1, \dots, m\}$.
- (ii) is a consequence of Lemma 1.20 and the observation that the edge groups include to meridians in $\pi_1(S^{n+2} \setminus K)$.
- (iii) is a special case of [Fri25: Proposition 165.11]. ∎

This provides the final missing piece for proving that $\mathcal{K}_n$ allows factorization:

Theorem 4.20. *Let $n \geq 1$.*

(1) *The monoid $\mathcal{K}_n$ allows factorization relative to knots with complement homotopy equivalent to S^1. In particular, $\mathcal{K}_n$ allows factorization for $n \neq 2$.*

(2) *For $n \neq 2$, the only unit of $\mathcal{K}_n$ is the unknot.*

Proof.

(1) By Lemma 4.19 there exists a map $\mathcal{K}_n \to \mathbb{N}_0$ bounding factors relative to $K \in \mathcal{K}_n$ with $\pi_1(S^{n+2} \setminus K) \cong \mathbb{Z}$. By Proposition 4.16 (2) there exists a complexity function on $\mathcal{K}_n$ relative to
$$\{K \in \mathcal{K}_n \mid \forall i \geq 1 : \mathrm{H}_i((\widetilde{S^{n+2} \setminus K})_{\mathrm{ab}}) = 0\}$$
Combining them using Proposition 4.18 yields by Proposition 4.17 that $\mathcal{K}_n$ allows factorization relative to
$$\{K \in \mathcal{K}_n \mid \forall i \geq 1 : \mathrm{H}_i((\widetilde{S^{n+2} \setminus K})_{\mathrm{ab}}) = 0,\ \pi_1(S^{n+2} \setminus K) \cong \mathbb{Z}\}$$
This proves the claim by Theorem 4.10.

(2) By (1) the subset above is closed under division. In particular, it contains all units and the claim again follows from Theorem 4.10. ∎

5. Monoids of knots: Uniqueness of factorization

5.1. The Seifert form and simple knots

In the last chapter, we established the existence of a factorization of knots. A natural next question is to wonder whether it is unique. We will need new invariants to study this. At first we restrict ourselves to odd dimensions where we can adapt Seifert surfaces and forms from the classical dimension:[25]

Definition. Let $q \geq 1$ and $K \subseteq S^{2q+1}$ be an oriented $(2q-1)$-dimensional knot. A *Seifert manifold* for K is a connected compact oriented smooth submanifold $\Sigma \subseteq S^{2q+1}$ with $\partial\Sigma = K$ as oriented manifolds. Let $\beta\colon (\Sigma \setminus \partial\Sigma) \times [-1,1] \to S^{2q+1} \setminus K$ be a bicollar, i.e. an orientation-preserving smooth embedding such that $\beta|_{(\Sigma\setminus\partial\Sigma)\times\{0\}}$ is the identity. Set $\beta^+ := \beta_1\colon \Sigma \setminus \partial\Sigma \to S^{n+2} \setminus \Sigma$. The *Seifert form of K with respect to* Σ is the bilinear form

$$\begin{array}{rcl} \mathrm{FH}_q(\Sigma) \times \mathrm{FH}_q(\Sigma) & \to & \mathbb{Z} \\ (a,b) & \mapsto & \langle (\mathrm{AD}_\Sigma \circ (\beta^+)_* \circ (i_*)^{-1})(b), a \rangle \end{array}$$

where

- $\mathrm{FH}_q(\Sigma) := \mathrm{H}_q(\Sigma)/\operatorname{Tor}(\mathrm{H}_q(\Sigma))$ is the free part of $\mathrm{H}_q(\Sigma)$ (compare Lemma 2.17)
- $\mathrm{AD}_\Sigma\colon \mathrm{H}_q(S^{n+2} \setminus \Sigma) \to \mathrm{H}^q(\Sigma)$ is the Alexander Duality isomorphism from Proposition 1.11
- $i\colon \Sigma \setminus \partial\Sigma \hookrightarrow \Sigma$ is the inclusion. It is a homotopy equivalence (see [Fri25: Corollary 32.8])
- $\langle -, - \rangle$ is the Kronecker pairing (see [Fri25: Lemma 185.7])

A *Seifert matrix* of K is an integer matrix representing the Seifert form in some basis of $\mathrm{FH}_q(\Sigma)$.

Lemma 5.1. *Let $q \geq 1$ and $K \subseteq S^{2q+1}$ be an oriented $(2q-1)$-dimensional knot.*

1. *There exists a Seifert manifold Σ for K and a bicollar $\beta\colon (\Sigma \setminus \partial\Sigma) \times [-1,1] \to S^{2q+1} \setminus K$.*
2. *The Seifert form of a K with respect to a Seifert manifold Σ does not depend on the choice of bicollar.*

Proof.

1. see [Fri25: Propositions 266.11, 220.10] and [Fri25: Theorem 52.2]
2. This follows from the uniqueness of bicollars, see [Fri25: Proposition 52.4] ■

At least in the easiest case, it is straightforward to determine this form:

Example 5.2. Let $q \geq 1$ and $U \subseteq S^{2q+1}$ be the $(2q-1)$-dimensional unknot.

- Then U has a Seifert manifold diffeomorphic to $\overline{B}^{2q}$. The corresponding Seifert form is the unique bilinear form $\{0\} \times \{0\} \to \mathbb{Z}$.
- Take the standard embedding of a product of spheres

$$\begin{array}{rcl} S^q \times S^q & \to & S^{2q+1} \\ (x,y) & \mapsto & \frac{1}{\sqrt{2}}(x,y) \end{array}$$

  and remove the interior of a $\overline{B}^{2q}$ from its image. This gives another Seifert manifold for U. Calculating the Seifert form using [Fri25: Propositions 266.13, 241.13, Theorem 206.10] yields that

$$\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$$

  also is a Seifert matrix for U.

[25]This section we follow various papers of J. Levine, primarily [Lev70; Lev77].

Thus we see that – as for classical knots – the Seifert form can only be an invariant up to S-equivalence:

Definition. Let A be a square integer matrix. Then

$$\begin{pmatrix} A & * & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix} \qquad \text{and} \qquad \begin{pmatrix} A & 0 & 0 \\ * & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$$

with arbitrary $*$ are *elementary enlargements* of A. The equivalence relation on square integer matrices generated by congruence and elementary enlargements is *S-equivalence.*
Two bilinear forms on finitely generated free abelian groups are *S-equivalent* if matrices representing them are S-equivalent.

Proposition 5.3. *Let $q \geq 1$ and $K, L \subseteq S^{2q+1}$ be oriented $(2q-1)$-dimensional knots with Seifert manifolds Σ_K, Σ_L. If K and L are smoothly isotopic, the Seifert forms of K with respect to Σ_K and L with respect to Σ_L are S-equivalent.*

Proof. see [Lev70: Theorem 1] [26] ■

To apply Seifert matrices in our discussions of factorization, we need that they form a monoid homomorphism to a suitable monoid:

Definition. Let $\mathcal{S}$ be the monoid of S-equivalence classes of square integer matrices under direct sum.

Proposition 5.4. *Let $q \geq 1$. The Seifert matrix defines a monoid homomorphism $\mathcal{K}_{2q-1} \to \mathcal{S}$.*

Proof. The map is well-defined by Proposition 5.3. The unknot gets mapped to the neutral element by Example 5.2. It remains to prove additivity:
Let $K, K' \subseteq S^{2q+1}$ be $(2q-1)$-dimensional knots bounding Seifert manifolds $\Sigma, \Sigma' \subseteq S^{2q+1}$. Choose smooth embeddings $\varphi\colon (\overline{B}^{2q+1}, \overline{B}^{2q-1}) \to (S^{2q+1}, K)$, $\varphi'\colon (\overline{B}^{2q+1}, \overline{B}^{2q-1}) \to (S^{2q+1}, K')$ such that one of them is orientation-preserving and the other orientation-reversing. We can assume that $\varphi\colon \overline{B}^{2q+1} \to S^{2q+1}$ restricts to a diffeomorphism

$$\varphi\colon \overline{B}^{2q+1} \cap \left(\mathbb{R}^{2q-1} \times \mathbb{R}_{\geq 0} \times \{0\}\right) \to \Sigma \cap \varphi(\overline{B}^{2q+1})$$

and the analogous statement for φ'.[27] Let $\iota\colon S^{2q+1} \setminus \varphi(0) \to S^{2q+1} \# S^{2q+1}$ be the inclusion and analogously consider ι'. Pick an orientation-preserving diffeomorphism $\Phi\colon S^{2q+1} \# S^{2q+1} \to S^{2q+1}$. Then

$$\Sigma^{\#} := (\Phi \circ \iota)(\Sigma \setminus \varphi(0)) \cup (\Phi \circ \iota')(\Sigma' \setminus \varphi'(0))$$

is a Seifert manifold for $K \# K'$.
Let $\beta\colon (\Sigma \setminus \partial\Sigma) \times [-1,1] \to S^{2q+1}$, $\beta'\colon (\Sigma' \setminus \partial\Sigma') \times [-1,1] \to S^{2q+1}$ be bicollars. We can then use [Fri25: Theorem 54.30] to find a bicollar $\beta^{\#}\colon (\Sigma^{\#} \setminus \partial\Sigma^{\#}) \times [-1,1] \to S^{2q+1}$ such that $\beta^{\#} \circ \Phi \circ \iota = \Phi \circ \iota \circ \beta$ outside of a ball around $\varphi(0)$ and analogously for β'.
It is now a straightforward – if somewhat lengthy – calculation using the Mayer–Vietoris sequence to show that if A, A' are Seifert matrices for K, K' resulting from Σ, Σ', $A \oplus A'$ is a Seifert matrix for $K \# K'$ resulting from $\Sigma^{\#}$. ■

Of course, a monoid homomorphism does not help much in our quest to study uniqueness of factorization in $\mathcal{K}_n$. We now work to restrict domain and codomain in such a way that the homomorphism becomes an isomorphism.

[26] Note that the argument presented there does not go through the arduous process of describing how precisely two Seifert manifolds of a given knot are related. Whether it is actually simpler may however be debated. It certainly has the caveat of needing the invariance of the Alexander polynomial as an input.

[27] For lack of a better reference, this is a direct consequence of [Fri25: Theorem 54.33].

The restriction on the image comes from the observation that – as in the classical dimension – we can recover the intersection from of the Seifert manifold from the Seifert form:

Proposition 5.5. *Let $q \geq 1$ and $K \subseteq S^{2q+1}$ be an oriented $(2q-1)$-dimensional knot with Seifert manifold Σ. Let $\Phi\colon \mathrm{FH}_q(\Sigma) \times \mathrm{FH}_q(\Sigma) \to \mathbb{Z}$ be the corresponding Seifert form. The intersection number of $a, b \in \mathrm{FH}_q(\Sigma)$ is*

$$a \cdot_\Sigma b = \Phi(a,b) + (-1)^q \cdot \Phi(b,a)$$

Sketch of a proof. In principle, this works as in classical knot theory. The key observation is that

$$a \cdot_\Sigma b = \beta_*([-1,1] \times b) \cdot_{S^{2q+1}} a = \langle (\mathrm{AD}_\Sigma \circ (\beta^+)_*(b), a\rangle + (-1)^q \cdot \langle \mathrm{AD}_\Sigma \circ (\beta^-)_*(b), a\rangle$$

which follows from the definition of the Alexander duality isomorphism if one is patient enough – which we are not. Instead we refer to [Lev66: Paragraph 2.5] claiming the same with more authority. ∎

Corollary 5.6. *Let $q \geq 1$ and A be a Seifert matrix for a simple oriented $(2q-1)$-dimensional knot $K \subseteq S^{2q+1}$.*
 (1) *The matrix $A + (-1)^q A^T$ is invertible over $\mathbb{Z}$.*
 (2) *If $q = 2$, the signature of $A + A^T$ is divisible by 16.*

Proof. By Proposition 5.5 $A + (-1)^q A^T$ represents the intersection form of a $2q$-dimensional manifold Σ with boundary $\partial\Sigma \cong S^{2q-1}$. Let $\hat{\Sigma}$ be a smooth manifold obtained from Σ by attaching $\overline{B}^{2q}$ to $\partial\Sigma$ (see [Fri25: Proposition 51.4]). Then $\hat{\Sigma}$ is a closed $2q$-dimensional smooth manifold. By the Mayer–Vietoris sequence and naturality of the intersection form, the inclusion $\Sigma \hookrightarrow \hat{\Sigma}$ induces an isometry of intersection forms.
 (1) By Poincaré duality the intersection form of a closed even-dimensional manifold is non-singular (see [Fri25: Proposition 210.13 (3)]), i.e. $A + (-1)^q A^T$ is invertible over $\mathbb{Z}$.
 (2) Let $\tau\colon \mathrm{T}\Sigma \to \Sigma$ be the tangent bundle and $\epsilon\colon \Sigma \times \mathbb{R} \to \Sigma$ be the trivial line bundle. Then $\tau \oplus \epsilon^2 \cong \mathrm{T}\,\mathbb{R}^6$ is a trivial bundle, since codimension one submanifolds have trivial normal bundle by [Fri25: Theorem 55.4] and $\Sigma \subseteq S^5 \subseteq \mathbb{R}^6$ is a sequence of codimension one submanifolds. Hence, the second Stiefel-Whitney class of Σ is 0 by [Fri25: Theorem 257.16]. It follows from the Mayer–Vietoris sequence that the inclusion induced map $\mathrm{H}^2(\hat{\Sigma}) \to \mathrm{H}^2(\Sigma)$ is an isomorphism, so the naturality of Stiefel-Whitney classes [Fri25: Proposition 257.12] implies that the second Stiefel-Whitney class of $\hat{\Sigma}$ is 0, too. Hence, $\hat{\Sigma}$ is spin and Rokhlin's Theorem (see [Fri25: Theorem 214.7]) yields that its signature, i.e. the signature of $A + A^T$, is divisible by 16. ∎

It will turn out that these conditions are sufficient for describing which matrices arise as Seifert matrices.
The Seifert form is certainly not enough to describe the smooth isotopy class of an arbitrary knot. Hence, we restrict to a subclass of knots where most invariants vanish by definition:

Definition. Let $q \geq 1$. A $(2q-1)$-dimensional knot $K \subseteq S^{2q+1}$ is *simple* if $\pi_i(S^{n+2} \setminus K) \cong \pi_i(S^1)$ for $i \in \{1, \ldots, q-1\}$.

By Theorem 4.10 this amounts to defining a knot to be simple if the first $q-1$ homotopy groups of its complement are the same as for the unknot. With the next theorem we can establish that requiring this for the first q homotopy groups would force the knot to be trivial.

Theorem 5.7. *Let $q \geq 1$ and $K \subseteq S^{2q+1}$ be a $(2q-1)$-dimensional knot. For $k \in \{1, \ldots, q\}$ the following are equivalent:*
 (i) *$\pi_i(S^{2q+1} \setminus K) \cong \pi_i(S^1)$ for $i \in \{1, \ldots, k\}$*
 (ii) *K bounds a Seifert manifold $\Sigma \subseteq S^{2q+1}$ with $\pi_i(\Sigma) = 0$ for $i \in \{1, \ldots, k\}$*
 (iii) *$\pi_1(S^{2q+1} \setminus K) \cong \mathbb{Z}$ and $\mathrm{H}_i(\widetilde{(S^{n+2} \setminus K)}_{\mathrm{ab}}) = 0$ for $i \in \{1, \ldots, k\}$*

Proof.

(i) $\Rightarrow$ (ii) This is the content of [Lev65: Theorem 2].

(ii) $\Rightarrow$ (iii) By the HNN-Seifert–van Kampen Theorem (see [Fri25: Theorem 125.27])

$$\pi_1(S^{2q+1}\setminus K)\cong \pi_1(S^{2q+1}\setminus\Sigma)*_{\pi_1(\Sigma\setminus\partial\Sigma)} \underset{\substack{\uparrow\\ \pi_1(\Sigma\setminus\partial\Sigma)=\{e\}}}{\cong} \pi_1(S^{2q+1}\setminus\Sigma)*\mathbb{Z}\mu$$

where μ is a meridian of K. Hence, Proposition 1.21 implies that $\pi_1(S^{2q+1}\setminus K)=\mathbb{Z}\mu$ (also compare Proposition 2.24 (3)). Hence, $\widetilde{(S^{n+2}\setminus K)}_{\mathrm{ab}}$ is simply connected and thereby $\mathrm{H}_1(\widetilde{(S^{n+2}\setminus K)}_{\mathrm{ab}})=0$ by the Hurewicz Theorem. In Construction 4.7 we found a pushout

$$\begin{tikzcd}
((\Sigma\setminus\partial\Sigma)\times((-1,0)\cup(0,1)))\times\mathbb{Z} \arrow[r, hook] \arrow[d, hook] & (S^{2q+1}\setminus\Sigma)\times\mathbb{Z} \arrow[d, hook] \\
((\Sigma\setminus\partial\Sigma)\times(-1,1))\times\mathbb{Z} \arrow[r, hook] & \widetilde{(S^{2q+1}\setminus K)}_{\mathrm{ab}}
\end{tikzcd}$$

The Hurewicz Theorem and the Mayer–Vietoris sequence now imply that for $i\in\{2,\dots,k\}$

$$\mathrm{H}_i(\widetilde{(S^{2q+1}\setminus K)}_{\mathrm{ab}})\cong\bigoplus_{\mathbb{Z}}\mathrm{H}_i(S^{2q+1}\setminus\Sigma)$$

This is trivial as

$$\mathrm{H}_i(S^{2q+1}\setminus\Sigma)\overset{\substack{\text{Proposition 1.11}\\ \downarrow}}{\cong}\mathrm{H}^{2q-i}(\Sigma)\overset{\substack{\text{Poincaré Duality}\\ \downarrow}}{\cong}\mathrm{H}_i(\Sigma,\partial\Sigma)\overset{\substack{\text{Long exact sequence of }(\Sigma,\partial\Sigma)\\ \downarrow}}{=}0$$

(iii) $\Rightarrow$ (i) This is a consequence of the Hurewicz Theorem and the observation that coverings induce isomorphisms on higher homotopy groups (see [Fri25: Proposition 120.22]). ■

Corollary 5.8. *Let $q\geq 1$ and $K\subseteq S^{2q+1}$ be a $(2q-1)$-dimensional knot. Then K is unknotted if and only if $\pi_k(S^{2q+1}\setminus K)\cong\pi_k(S^1)$ for $k\in\{1,\dots,q\}$.*

Proof. The 'only if'-implication follows from Theorem 4.10. For the 'if'-part, apply Theorem 5.7 to obtain a Seifert manifold $\Sigma\subseteq S^{2q+1}$ with $\pi_i(\Sigma)=0$ for $i\in\{1,\dots,q\}$. For $i\in\{q+1,\dots,2q-1\}$

$$\mathrm{H}_i(\Sigma)\overset{\substack{\text{Poincaré Duality}\\ \downarrow}}{\cong}\mathrm{H}^{2q-i}(\Sigma,\partial\Sigma)\overset{\substack{\text{Long exact sequence of }(\Sigma,\partial\Sigma)\\ \downarrow}}{\cong}\mathrm{H}^{2q-i}(\Sigma)\underset{\substack{\uparrow\\ \text{Universal Coefficient Theorem}}}{\cong}\mathrm{Ext}(\mathrm{H}_{2q-i-1}(\Sigma),\mathbb{Z})\oplus\mathrm{Hom}(\mathrm{H}_{2q-i}(\Sigma),\mathbb{Z})\overset{\substack{\text{Hurewicz Theorem}\\ \downarrow}}{=}0$$

Since $\partial\Sigma\neq\varnothing$, also $\mathrm{H}_{2q}(\Sigma)=0$. Essentially the same argument as in Theorem 5.7 shows that $\mathrm{H}_i(\widetilde{(S^{n+2}\setminus K)}_{\mathrm{ab}})=0$ for $i\geq 1$. Then K is trivial by Theorem 4.10. ■

By work of J. Levine, the homomorphism $\mathcal{K}_{2q-1}\to\mathcal{S}$ becomes a monomorphism when restricted to simple knots with the image completely described by the condition in Corollary 5.6:

Theorem 5.9. *Let $q\geq 1$.*

(1) *Let A be a Seifert matrix for a simple oriented $(2q-1)$-dimensional knot $K\subseteq S^{2q+1}$.*
   (a) *The matrix $A+(-1)^qA^T$ is invertible over $\mathbb{Z}$.*
   (b) *If $q=2$, the signature of $A+A^T$ is divisible by 16.*

(2) *Conversely, any square integer matrix A is satisfying* (a) *and* (b) *is S-equivalent to a Seifert matrix of a simple oriented $(2q-1)$-dimensional knot.*

(3) *For $q\neq 1$, two simple oriented $(2q-1)$-dimensional knots are isotopic if and only if their Seifert forms are S-equivalent.*

Proof.

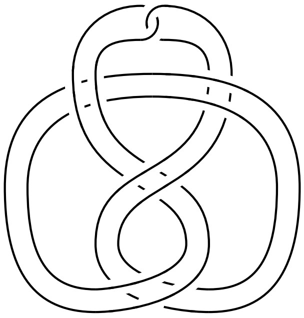
(1) see Corollary 5.6
(2) see [Lev70: Theorem 2]
(3) see [Lev70: Theorem 3] ■

In the classical dimension every knot is simple, but the Seifert matrix is not a complete invariant:

Example 5.10. The Whitehead double of any non-trivial knot is non-trivial but has

$$\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$$

as a Seifert matrix and is thereby S-equivalent to the unknot (compare [Lic97: Theorem 6.15]).



Whitehead double of the figure-8 knot

In fact, it will follow from Proposition 5.16 and Theorem 5.12 that any classical knot with trivial Alexander polynomial, i.e. $\det(tA - A^T) = 1$ for a Seifert matrix A, is S-equivalent to the unknot.

5.2. The Blanchfield form

In theory, we have now algebraically described the isotopy classes of simple knots. In practice, the problem is still rather hard since the number theoretic problem of determining whether two matrices are S-equivalent is non-standard. Thankfully, we can upgrade the Seifert form to a form which actually is an invariant up to isotopy – at the expense of it being defined over $\mathbb{Z}[t^{\pm 1}]$-modules. We begin with a general algebraic definition:[28]

Definition. Let R be a commutative ring with an *involution*, i.e. a ring homomorphism $R \to R,\ x \mapsto \overline{x}$ whose square is the identity. Let M, N be R-modules. Let $\overline{\mathrm{Hom}_R}(M, N)$ be the R-module with the same underlying abelian group as $\mathrm{Hom}_R(M, N)$ but the R-module structure precomposed with the involution.
The *dual R-module* of M is $M^* := \overline{\mathrm{Hom}_R}(M, R)$.
An *involution* on an R-module N is a group homomorphism $N \to N,\ x \mapsto \overline{x}$ such that for $\lambda \in R, x \in N$

$$\overline{\lambda x} = \overline{\lambda} \cdot \overline{x}$$

Let $\epsilon \in \{\pm 1\}$. A map $\Phi \colon M \times M \to N$ is a *ϵ-Hermitian form* if for all $x, y, z \in M$, $\lambda \in R$:

- $\Phi(\lambda x + y, z) = \overline{\lambda} \cdot \Phi(x, z) + \Phi(y, z)$
- $\Phi(x, \lambda y + z) = \lambda \cdot \Phi(x, y) + \Phi(x, z)$
- $\Phi(x, y) = \epsilon \cdot \overline{\Phi(x, y)}$

It is *non-singular* if the *adjoint map*

$$\begin{aligned} M &\to \overline{\mathrm{Hom}}(M, N) \\ x &\mapsto \Phi(x, -) \end{aligned}$$

is an isomorphism.

For now, we will be interested in this definition in the following setting:

Definition. Consider the *conjugation* ring homomorphism $\mathbb{Q}(t) \to \mathbb{Q}(t),\ f \mapsto \overline{f}$ determined by $\overline{t} = t^{-1}$ and note that it restricts to $\mathbb{Z}[t^{\pm 1}]$. We also extend it linearly to free $\mathbb{Z}[t^{\pm 1}]$-modules and to quotients of free $\mathbb{Z}[t^{\pm 1}]$-modules by submodules that are invariant under conjugation.

[28] In this section we follow [Tro73].

Definition. Let $q \geq 1$ and $K \subseteq S^{2q+1}$ be an oriented $(2q-1)$-dimensional knot. Let A be an $(n \times n)$ Seifert matrix for K. Set

$$\mathrm{Al}_K := \mathbb{Z}[t^{\pm 1}]^n / (tA + (-1)^q A^T)\mathbb{Z}[t^{\pm 1}]^n$$

The *Blanchfield form* of K is

$$\begin{aligned} \mathrm{Bl}_K \colon \mathrm{Al}_K \times \mathrm{Al}_K &\to \mathbb{Q}(t)/\mathbb{Z}[t^{\pm 1}] \\ ([v],[w]) &\mapsto (t-1) \cdot \overline{v}^T \cdot (tA + (-1)^q A^T)^{-1} \cdot w \end{aligned}$$

We note some algebraic properties of this form:

Proposition 5.11. *Let $K \subseteq S^{2q+1}$ be an oriented $(2q-1)$-dimensional knot.*

(1) (a) *The module Al_K is a finitely generated $\mathbb{Z}[t^{\pm 1}]$-module of type K.*
  (b) *The Blanchfield form $\mathrm{Bl}_K \colon \mathrm{Al}_K \times \mathrm{Al}_K \to \mathbb{Q}(t)/\mathbb{Z}[t^{\pm 1}]$ is a well-defined non-singular $(-1)^{q-1}$-Hermitian form.*

(2) *The isomorphism class of the $\mathbb{Z}[t^{\pm 1}]$-module Al_K and the isometry class of the Blanchfield form $\mathrm{Bl}_K \colon \mathrm{Al}_K \times \mathrm{Al}_K \to \mathbb{Q}(t)/\mathbb{Z}[t^{\pm 1}]$ do not depend on the choice of Seifert matrix for K.*

Proof.

(1) (a) Let A be an $(n \times n)$ Seifert matrix for K. The $\mathbb{Z}[t^{\pm 1}]$-module

$$\mathrm{Al}_K = \mathbb{Z}[t^{\pm 1}]^n / (tA + (-1)^q A^T)\mathbb{Z}[t^{\pm 1}]^n$$

is finitely generated by definition. To show that multiplication by $t-1$ defines an isomorphism $\mathrm{Al}_K \to \mathrm{Al}_K$, set $d(t) := \det(tA + (-1)^q A^T)$. Then $d(t) - d(1) \in \mathbb{Z}[t^{\pm 1}]$ vanishes at 1, so there exists $\varphi(t) \in \mathbb{Z}[t^{\pm 1}]$ with $(t-1)\varphi(t) = d(t) - d(1)$. Multiplication by $-d(1)\varphi(t)$ defines an inverse to multiplication by $t-1$:
For $x \in \mathbb{Z}[t^{\pm 1}]^n$

$$-d(1)\varphi(t) \cdot (t-1) \cdot x = -d(1)(d(t) \cdot x - d(1) \cdot x) = -d(1)d(t) \cdot x + d(1)^2 \cdot x$$

Now observe that by Cramer's rule (see [Lan02: Proposition XIII.4.16])

$$d(t) \cdot x = \det(tA + (-1)^q A^T) \cdot x \in (tA + (-1)^q A^T)\mathbb{Z}[t^{\pm 1}]^n$$

and $d(1) = \det(A + (-1)^q A^T) = \pm 1$ by Corollary 5.6.

(b) Let A be a $(n \times n)$ Seifert matrix for K. By Corollary 5.6, $\det(A + (-1)^q A^T) = \pm 1$. Hence, $\det(tA + (-1)^q A^T) \neq 0$ and $tA + (-1)^q A^T$ is invertible over $\mathbb{Q}(t)$.
Let $v, w \in \mathbb{Z}[t^{\pm 1}]^n$. If $w = (tA + (-1)^q A^T) \cdot x$ for $x \in \mathbb{Z}[t^{\pm 1}]^n$,

$$\overline{v}^T \cdot (tA + (-1)^q A^T)^{-1} \cdot w = \overline{v}^T \cdot x \in \mathbb{Z}[t^{\pm 1}]$$

If $v = (tA + (-1)^q A^T) \cdot x$ for $x \in \mathbb{Z}[t^{\pm 1}]^n$,

$$\begin{aligned} \overline{v}^T \cdot (tA + (-1)^q A^T)^{-1} \cdot w &= \overline{x}^T \cdot (t^{-1}A^T + (-1)^q A) \cdot (tA + (-1)^q A^T)^{-1} \cdot w \\ &= (-1)^q t^{-1} \cdot \overline{x}^T \cdot ((-1)^q A^T + tA) \cdot (tA + (-1)^q A^T)^{-1} \cdot w \\ &= (-1)^q t^{-1} \cdot \overline{x}^T \cdot w \in \mathbb{Z}[t^{\pm 1}] \end{aligned}$$

This shows that the Blanchfield form is well-defined. It is ϵ-Hermitian since

$$\begin{aligned} \Big((t-1) \cdot (tA + (-1)^q A^T)^{-1}\Big)^T &= (t-1) \cdot (tA^T + (-1)^q A)^{-1} \\ &= (t-1) \cdot \Big((-1)^q t \cdot (t^{-1}A + (-1)^q A^T)\Big)^{-1} \\ &= (t-1) \cdot (-1)^q t^{-1} \cdot (t^{-1}A + (-1)^q A^T)^{-1} \\ &= (-1)^{q-1} \cdot (t^{-1} - 1) \cdot (t^{-1}A + (-1)^q A^T)^{-1} \\ &= (-1)^{q-1} \cdot \overline{(t-1) \cdot (tA + (-1)^q A^T)^{-1}} \end{aligned}$$

The Blanchfield form is non-singular by standard linear algebra since multiplication by $t-1$ defines an automorphism of Al_K by (1).

(2) By Proposition 5.3 we need to show that the isomorphism class of the $\mathbb{Z}[t^{\pm 1}]$-module Al_K and the isometry class of the Blanchfield form $\mathrm{Bl}_K\colon \mathrm{Al}_K \times \mathrm{Al}_K \to \mathbb{Q}(t)/\mathbb{Z}[t^{\pm 1}]$ are invariant under S-equivalence of the Seifert matrix defining them. It is relatively clear that they are invariant under congruence of the Seifert matrix, so it remains to consider elementary enlargements:
Let A be a Seifert matrix for K and

$$A' := \begin{pmatrix} A & * & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}$$

The Blanchfield form defined by A' is represented by

$$(t-1)\begin{pmatrix} A & t\cdot * & 0 \\ (-1)^q \cdot * & 0 & t \\ 0 & (-1)^q & 0 \end{pmatrix}^{-1}$$

Over $\mathbb{Q}(t)$ this is isometric to the form represented by

$$(t-1)\begin{pmatrix} A^{-1} & 0 \\ 0 & U \end{pmatrix}^{-1} \qquad \text{where} \qquad U = \begin{pmatrix} 0 & t \\ (-1)^q & 0 \end{pmatrix}$$

Since U is invertible over $\mathbb{Z}[t^{\pm 1}]$, this form is in turn isometric to the Blanchfield form determined by A.
The other elementary enlargement of A can be dealt with similarly. ∎

With this we have defined a bilinear form whose isometry class is a knot invariant. In the process we have lost no information compared to the Seifert from we started with:

Theorem 5.12. *Let $q \geq 1$. Two oriented $(2q-1)$-dimensional knots have S-equivalent Seifert forms if and only if their Blanchfield forms are isometric.*

Proof. The 'if' part is Proposition 5.11 (2), the converse is the content of [Tro73]. ∎

In Theorem 5.9 we have classified simple knots in terms of their Seifert form. In light of the last theorem, this must also be possible in terms of the Blanchfield form by translating the necessary and sufficient properties of the Seifert form to Blanchfield forms. Most of this is a relatively direct exercise in linear algebra. The only real trouble is the signature restriction arising from Rokhlin's Theorem when $q = 2$. Here we need to turn the Blanchfield form back into a form over an ordered field to give it a signature:

Definition. Let $\chi\colon \mathbb{Q}(t)/\mathbb{Q}[t^{\pm 1}, (1-t)^{-1}] \to \mathbb{Q}$ be the unique $\mathbb{Q}$-linear map such that for $f, g \in \mathbb{Q}[t]$ with $\deg(f) < \deg(g)$ and g coprime to t and $1-t$

$$\chi\left(\left[\frac{f}{g}\right]\right) = \left(\frac{f}{g}\right)'(1)$$

Let $\epsilon \in \{\pm 1\}$ and $\Phi\colon M \times M \to \mathbb{Q}(t)/\mathbb{Z}[t^{\pm 1}]$ be a ϵ-Hermitian form on a $\mathbb{Z}[t^{\pm 1}]$ module M. The *scalar form of* Φ is $\chi \circ (\mathrm{id}_{\mathbb{Q}} \otimes \Phi)\colon (\mathbb{Q} \otimes M) \times (\mathbb{Q} \otimes M) \to \mathbb{Q}$.

Lemma 5.13.
(1) *The $\mathbb{Q}$-linear map $\chi\colon \mathbb{Q}(t)/\mathbb{Q}[t^{\pm 1}, (1-t)^{-1}] \to \mathbb{Q}$ is well-defined and unique.*
(2) *For $\alpha \in \mathbb{Q}(t)/\mathbb{Q}[t^{\pm 1}, (1-t)^{-1}]$*

$$\chi(\overline{\alpha}) = -\chi(\alpha) \qquad \textit{and} \qquad \chi((t-1)\alpha) = \alpha(1)$$

(3) *Let $\epsilon \in \{\pm 1\}$ and $\Phi\colon M \times M \to \mathbb{Q}(t)/\mathbb{Z}[t^{\pm 1}]$ be an ϵ-Hermitian form on a $\mathbb{Z}[t^{\pm 1}]$ module M. The scalar form of Φ is a $(-\epsilon)$-symmetric bilinear form $(\mathbb{Q} \otimes M) \times (\mathbb{Q} \otimes M) \to \mathbb{Q}$.*

Proof.

(1) As a $\mathbb{Q}$-vector space

$$\mathbb{Q}(t) = \mathbb{Q}[t^{\pm 1}, (1-t)^{-1}] \oplus \overbrace{\left\{ \frac{f}{g} \;\middle|\; \begin{matrix} f, g \in \mathbb{Q}[t], \deg(f) < \deg(g) \\ g \text{ coprime to } t, 1-t \end{matrix} \right\}}^{=:P}$$

This follows from the following observations:
- If $f, g, h \in \mathbb{Q}[t]$ with g, h coprime there exists $a, b \in \mathbb{Q}[t]$ such that $ag + bh = 1$ and thereby
$$\frac{f}{gh} = \frac{fa}{h} + \frac{fb}{g}$$
- Let $f, g \in \mathbb{Q}[t]$ with $\deg(f) < \deg(g)$ and g coprime to $t, 1-t$. Wlog. f and g are also coprime. Then $\frac{f}{g} \in \mathbb{Q}[t^{\pm 1}, (1-t)^{-1}]$ implies that the only prime factors of g are t and $1-t$. Thereby, $g \in \mathbb{Q}^*$ and $\deg(f) < \deg(g) = 0$, so $f = 0$.

For $f, g \in \mathbb{Q}[t]$ with g coprime to $t, 1-t$ the derivative $\left(\frac{f}{g}\right)'(1)$ is defined since $g(1) = 0$ implies that $1-t$ divides g. Hence, the claim follows from the linearity of the derivative.

(2) We begin with an intermediate claim:

Claim. *Let $f, g \in \mathbb{Q}[t]$ with g coprime to $t, 1-t$ and $\deg(f) \leq \deg(g)$. Then*
$$\chi\left(\left[\frac{f}{g}\right]\right) = \left(\frac{f}{g}\right)'(1)$$

Proof. Suppose g has degree $n \geq 0$ with leading coefficient $q \neq 0$. Let p be the coefficient of t^n in f. Then
$$\frac{f}{g} = \frac{p}{q} + \left(\frac{f}{g} - \frac{p}{q}\right) = \frac{p}{q} + \frac{qf - pg}{qg}$$
and $\frac{p}{q} \in \mathbb{Q}[t^{\pm 1}, (1-t)^{-1}]$, $\deg(qf - pg) < \deg(qg)$. So
$$\chi\left(\left[\frac{f}{g}\right]\right) = \chi\left(\left[\frac{f}{g} - \frac{p}{q}\right]\right) = \left(\frac{f}{g} - \frac{p}{q}\right)'(1) = \left(\frac{f}{g}\right)'(1)$$
□

To prove the original claim, we may wlog. assume that $\alpha \in P$. Then $\chi(\alpha) = \alpha'(1)$. In general, $\overline{\alpha} \notin P$, but it still satisfies the hypotheses of the claim. Similarly, the claim is applicable to $(t-1)\alpha$. Hence, (2) follows since by the chain and product rule
$$\overline{\alpha}'(t) = (\alpha(t^{-1}))' = -t^{-2} \cdot \alpha'(t^{-1}) \qquad \text{and} \qquad ((t-1)\alpha(t))' = \alpha(t) + (t-1)\alpha'(t)$$

(3) The map $\mathrm{id}_{\mathbb{Q}} \otimes \Phi \colon (\mathbb{Q} \otimes M) \times (\mathbb{Q} \otimes M) \to \mathbb{Q}(t)/\mathbb{Q}[t^{\pm 1}, (1-t)^{-1}]$ is $\mathbb{Q}$-bilinear. Hence, the scalar form $\chi \circ (\mathrm{id}_{\mathbb{Q}} \otimes \Phi)$ is also $\mathbb{Q}$-bilinear by (1). To see that it is $(-\epsilon)$-symmetric observe that by (2) for $x, y \in M$
$$\begin{aligned}(\chi \circ (\mathrm{id}_{\mathbb{Q}} \otimes \Phi))(1 \otimes x, 1 \otimes y) &= \chi(1 \otimes \Phi(x, y)) = \chi(1 \otimes \epsilon \cdot \overline{\Phi(y, x)}) = \epsilon \cdot \chi(\overline{1 \otimes \Phi(x, y)}) \\ &= -\epsilon \cdot (\chi \circ (\mathrm{id}_{\mathbb{Q}} \otimes \Phi))(1 \otimes y, 1 \otimes x)\end{aligned}$$
■

With this form established we can restate the algebraic classification of simple knots in terms of Blanchfield forms:

Theorem 5.14. *Let $q \geq 1$.*

(1) *Let $K \subseteq S^{2q+1}$ be a simple oriented $(2q-1)$-dimensional knot.*
   (a) *The module Al_K is a finitely generated $\mathbb{Z}[t^{\pm 1}]$-module of type K.*
   (b) *The Blanchfield form $\mathrm{Bl}_K \colon \mathrm{Al}_K \times \mathrm{Al}_K \to \mathbb{Q}(t)/\mathbb{Z}[t^{\pm 1}]$ is a non-singular $(-1)^{q-1}$-Hermitian form.*
   (c) *If $q = 2$, the signature of the scalar form of Bl_K is divisible by 16.*

(2) *Conversely, any $\mathbb{Z}[t^{\pm 1}]$-module with a form satisfying* (a)-(c) *arises from a simple oriented $(2q-1)$-dimensional knot.*

(3) *For $q \neq 1$, two simple oriented $(2q-1)$-dimensional knots are isotopic if and only if they have isometric Blanchfield forms.*

Proof.

(1) Statements (a) and (b) are proven in Proposition 5.11 (1). For (c) observe that the scalar form of Bl_K is represented by the matrix

$$\chi\Big(((t-1)(tA+(-1)^q A^T)^{-1}\Big) \overset{\text{Lemma 5.13 (2)}}{\doteq} (A+(-1)^q A^T)^{-1}$$

which has signature divisible by 16 by Theorem 5.9.

(2) see [Lev77: Theorem 16.1]

(3) This follows from Theorem 5.12 and Theorem 5.9. ∎

To use this classification in our quest to study factorization in $\mathcal{K}_{2q-1}$, we again need to prove it is given by a module homomorphism:

Definition. For $\epsilon \in \{\pm 1\}$ let $\mathcal{H}^\epsilon$ be the monoid of isometry classes of ϵ-Hermitian forms over $\mathbb{Z}[t^{\pm 1}]$ with values in $\mathbb{Q}(t)/\mathbb{Z}[t^{\pm 1}]$ under direct sum.

Proposition 5.15. *Let $q \geq 1$. The Blanchfield form defines a homomorphism $\mathcal{K}_{2q-1} \to \mathcal{H}^{(-1)^q}$.*

Proof. By Proposition 5.4 it suffices to show that for $\epsilon \in \{\pm 1\}$

$$\begin{aligned}
&\mathcal{S} \to \mathcal{H}^\epsilon \\
&A \mapsto \left(\begin{array}{rcl}\Phi_A\colon \mathbb{Z}[t^{\pm 1}]^n/A\mathbb{Z}[t^{\pm 1}]^n) \times (\mathbb{Z}[t^{\pm 1}]^n/A\mathbb{Z}[t^{\pm 1}]^n) & \to & \mathbb{Q}(t)/\mathbb{Z}[t^{\pm 1}] \\ (v,w) & \mapsto & (t-1)\cdot \overline{v}^T \cdot (tA+\epsilon A^T)^{-1}\cdot w\end{array}\right)
\end{aligned}$$

is a monoid homomorphism which follows from elementary linear algebra. ∎

To close out this section, we note that the module Al_K is not quite as mysterious as it seems at first glance – as its name and it being of type K might already have suggested, Al_K arises as an Alexander module of K:

Proposition 5.16. *Let $q \geq 1$ and $K \subseteq S^{2q+1}$ be a simple oriented $(2q-1)$-dimensional knot. Then*

$$\mathrm{Al}_K \cong \mathrm{H}_q(\widetilde{(S^{2q+1}\setminus K)}_{\mathrm{ab}})$$

as $\mathbb{Z}[t^{\pm 1}]$-modules.

Sketch of a proof. By Theorem 5.7 there exists a Seifert manifold $\Sigma \subseteq S^{2q+1}$ for K with $\pi_i(\Sigma) = 0$ for $i \in \{1, \dots, q-1\}$. In Construction 4.7 we found a pushout of subspaces

$$\begin{tikzcd}
(N\setminus\Sigma)\times\mathbb{Z} \arrow[r] \arrow[d] & (S^{2q+1}\setminus\Sigma)\times\mathbb{Z} \arrow[d] \\
N\times\mathbb{Z} \arrow[r] & \widetilde{(S^{2q+1}\setminus K)}_{\mathrm{ab}}
\end{tikzcd}$$

Observe that for a topological space X we can identify $\mathrm{H}_i(X\times\mathbb{Z}) \cong \mathrm{H}_i(X)\otimes\mathbb{Z}[t^{\pm 1}]$. With this observation the Mayer–Vietoris Theorem leads to a long exact sequence

$$\mathrm{H}_q(N\setminus\Sigma)\otimes\mathbb{Z}[t^{\pm 1}] \to \begin{matrix}\mathrm{H}_q(N)\otimes\mathbb{Z}[t^{\pm 1}]\oplus \\ \mathrm{H}_q(S^{2q+1}\setminus\Sigma)\otimes\mathbb{Z}[t^{\pm 1}]\end{matrix} \to \mathrm{H}_q(\widetilde{(S^{2q+1}\setminus K)}_{\mathrm{ab}}) \to \mathrm{H}_{q-1}(N\setminus\Sigma)\otimes\mathbb{Z}[t^{\pm 1}]$$

The last entry in this sequence is 0 by the Hurewicz Theorem since $N\setminus\Sigma \simeq \Sigma\sqcup\Sigma$. By Poincaré duality and the Universal Coefficient Theorem $\mathrm{H}_{q+1}(\Sigma) = 0$. Hence, the Universal Coefficient Theorem implies that $\mathrm{H}_q(\Sigma)$ is free (it is finitely generated and torsion would need to appear in the homology one dimension higher). The remainder of the argument is essentially the same as in classical knot theory, see [Lic97: Theorem 6.5]. ∎

Given that the Blanchfield pairing is defined on Al_K, the last proposition raises the question whether it may be defined directly on the Alexander module $\mathrm{H}_q((\widetilde{S^{2q+1} \setminus K})_{\mathrm{ab}})$. This is indeed the case:
The Alexander modules of a knot $K \subseteq S^{2q+1}$ are the homology of $X := S^{2q+1} \setminus \nu K$ with coefficients twisted by the abelianization $\pi_1(X) \to \langle t \rangle$. Hence, Poincaré duality gives an isomorphism

$$\mathrm{PD}\colon\ \mathrm{H}_q(X; \mathbb{Z}[t^{\pm 1}]) \to \mathrm{H}^{q+1}(X, \partial X; \mathbb{Z}[t^{\pm 1}])$$

The Bockstein sequence of $0 \to \mathbb{Z}[t^{\pm 1}] \to \mathbb{Q}(t) \to \mathbb{Q}(t)/\mathbb{Z}[t^{\pm 1}] \to 0$ gives rise to an isomorphism

$$\mathrm{BS}\colon\ \mathrm{H}^q(X; \mathbb{Z}[t^{\pm 1}]) \to \mathrm{H}^{q+1}(X; \mathbb{Q}(t)/\mathbb{Z}[t^{\pm 1}])$$

as a $\mathbb{Z}[t^{\pm 1}]$-module of type K is $\mathbb{Z}[t^{\pm 1}]$-torsion, i.e. vanishes when tensored with $\mathbb{Q}(t)$. Combining these with standard isomorphisms from homological algebra gives an isomorphism

$$\begin{tikzcd}
\mathrm{H}_q(X;\mathbb{Z}[t^{\pm1}]) \arrow[rr, "\Phi"] \arrow[d, "\mathrm{PD}"'] & & \overline{\mathrm{Hom}}_{\mathbb{Z}[t^{\pm1}]}(\mathrm{H}_q(X;\mathbb{Z}[t^{\pm1}]), \mathbb{Q}(t)/\mathbb{Z}[t^{\pm1}]) \\
\mathrm{H}^{q+1}(X,\partial X;\mathbb{Z}[t^{\pm1}]) \arrow[r] & \mathrm{H}^{q+1}(X;\mathbb{Z}[t^{\pm1}]) \arrow[r, "\mathrm{BS}^{-1}"] & \mathrm{H}^q(X, \mathbb{Q}(t)/\mathbb{Z}[t^{\pm1}]) \arrow[u]
\end{tikzcd}$$

which is the adjoint of the Blanchfield form, i.e. Bl_K is isometric to

$$\begin{aligned} \mathrm{H}_q(X;\mathbb{Z}[t^{\pm1}]) \times \mathrm{H}_q(X;\mathbb{Z}[t^{\pm1}]) &\to \mathbb{Q}(t)/\mathbb{Z}[t^{\pm1}] \\ (x,y) &\mapsto \Phi(y)(x) \end{aligned}$$

We did not take this approach since – as pointed out in [Lev77: after Theorem 14.1] – the signature restriction on the Blanchfield form for $q = 2$ does not naturally arise in this setting: There are no 4-manifolds in sight to which Rokhlin's Theorem could be applied.
Also note that Poincaré duality applied to Theorem 5.7 in the above manner shows that $\mathrm{H}_q(X;\mathbb{Z}[t^{\pm1}])$ is in fact the only non-trivial Alexander module of a simple $(2q-1)$-dimensional knot $K \subseteq S^{2q+1}$.

5.3. Forms and number fields

In this section we show that $\mathcal{H}^\epsilon$ does not allow cancellation. With the classification of simple knots from the last section, this will imply that $\mathcal{K}_n$ does not allow cancellation for $n \geq 3$ odd. To do so, we need to find forms $M \times M \to \mathbb{Q}(t)/\mathbb{Z}[t^{\pm1}]$. We begin by sketching how ideas from number theory can be used to obtain them. However, the account that follows is entirely self-contained and does not rely on number theoretic arguments in the proofs, as the author wishes to keep it accessible to readers without a background in number theory. The trade-off is that, without these preliminary remarks, it would be difficult to see how one might arrive at the examples in the first place.[29]
The $\mathbb{Z}[t^{\pm1}]$-module $\mathbb{Q}(t)/\mathbb{Z}[t^{\pm1}]$ is hard to grasp. But since we are considering forms on finitely generated modules, we only need to consider the finitely generated submodules of $\mathbb{Q}(t)/\mathbb{Z}[t^{\pm1}]$. Many of those are contained in cyclic submodules which are isomorphic to quotients of $\mathbb{Z}[t^{\pm1}]$ – similar to the finitely generated subgroups of $\mathbb{Q}/\mathbb{Z}$ being isomorphic to $\mathbb{Z}/n\mathbb{Z}$. One might therefore consider forms over quotient rings $\mathbb{Z}[t^{\pm1}]/(f)$ for $f \in \mathbb{Z}[t]$.
If $f \in \mathbb{Z}[t]$ is irreducible, this ring arises in number theory since $\mathbb{Z}[t^{\pm1}]/(f)$ embeds into $\mathbb{Q}(\alpha)$ for $\alpha \in \mathbb{C}$ with $f(\alpha) = 0$. More precisely, $\mathbb{Z}[t^{\pm1}]/(f)$ is a finite index subring of a localization of the ring of integers of $\mathbb{Q}(\alpha)$ at finitely many primes (note that everything but the two finiteness conditions in this sentence is vacuous). In number theoretic terms, it is an order of the number field $\mathbb{Q}(\alpha)$.

[29] This section is based on [Bay85], which also contains some more number theoretic background.

However, orders of arbitrary number fields are still fairly complicated – if the degree of the extensions is large. If the degree of the extension – i.e. $\deg(f)$ – is 2, we understand their structure and their ideals fairly well. It will turn out that indeed $\deg(f) = 2$ will suffice to find our examples.
The last (or first) piece is now a way to turn a module over such a ring – in particular, an ideal of the ring – into a form:

Definition. Let R be a commutative ring with an involution. An R-module M is *reflexive* if the *evaluation homomorphism*

$$\begin{aligned} M &\mapsto M^{**} \\ m &\mapsto (f \mapsto \overline{f(m)}) \end{aligned}$$

is an isomorphism. Define for $\epsilon \in \{\pm 1\}$

$$\begin{aligned} \Phi^\epsilon(M)\colon (M \oplus M^*) \times (M \oplus M^*) &\to R \\ ((x, f), (y, g)) &\mapsto \overline{f(y)} + \epsilon g(x) \end{aligned}$$

Lemma 5.17. *Let R be a commutative ring with an involution, M, N be reflexive R-modules and $\epsilon \in \{\pm 1\}$.*
(1) *The map $\Phi^\epsilon(M)\colon (M \oplus M^*) \times (M \oplus M^*) \to R$ is a non-singular ϵ-Hermitian form.*
(2) *If M and N are isomorphic, $\Phi^\epsilon(M)$ and $\Phi^\epsilon(N)$ are isometric.*
(3) *The forms $\Phi^\epsilon(M \oplus N)$ and $\Phi^\epsilon(M) \oplus \Phi^\epsilon(N)$ are isometric.*

Proof. All claims follow from elementary linear algebra. Observe in particular that for R-modules A, B

$$\begin{aligned} A^* \oplus B^* &\to (A \oplus B)^* \\ (f, g) &\mapsto ((a, b) \mapsto f(a) + g(b)) \end{aligned}$$

is a natural isomorphism. ■

We can now provide the desired examples:

Lemma 5.18. *Let $f := 43t^2 - 85t + 43 \in \mathbb{Z}[t^{\pm 1}]$ and consider the ring $R := \mathbb{Z}[t^{\pm 1}]/(f)$ and the ideal $I := (5, t-2) \subseteq R$. Set furthermore $a := \frac{43t-41}{3}$ and consider the ring $\tilde{R} := R[a]$.*
Let $\epsilon \in \{\pm 1\}$. Then $\Phi^\epsilon(I) \oplus \Phi^\epsilon(\tilde{R}) \cong \Phi^\epsilon(R) \oplus \Phi^\epsilon(\tilde{R})$, but $I \oplus I^* \not\cong R \oplus R^*$, i.e. $\Phi^\epsilon(I) \not\cong \Phi^\epsilon(R)$.

Proof. The involution $R \to R$ given by $\overline{r}(t) := r(t^{-1})$ is well-defined, since $t^{-1}f$ is symmetric in t and t^{-1} which implies that $\overline{(f)} = (f)$.
The polynomial $f \in \mathbb{Z}[t]$ is irreducible since it is irreducible modulo 2. Hence, the ideal generated by f in $\mathbb{Z}[t]$ is prime. As it does not contain powers of t, its localization $(f) \subseteq \mathbb{Z}[t^{\pm 1}]$ is also prime (see [Lan02: Exercise II.5]). Hence, R and $\tilde{R}$ are integral domains.
To show that $\Phi^\epsilon(I) \oplus \Phi^\epsilon(\tilde{R}) \cong \Phi^\epsilon(R) \oplus \Phi^\epsilon(\tilde{R})$ it suffices to show that $I \oplus \tilde{R} \cong R \oplus \tilde{R}$ by Lemma 5.17. To this end, consider the R-linear map

$$\begin{aligned} \varphi\colon R \oplus \tilde{R} &\to I \oplus \tilde{R} \\ \begin{pmatrix} \lambda \\ \mu \end{pmatrix} &\mapsto \begin{pmatrix} 10\lambda + 9a\mu \\ \lambda + a\mu \end{pmatrix} \end{aligned}$$

This is well-defined since for $\lambda \in R$ clearly $10\lambda \in I$ and for $\mu \in \tilde{R}$

$$9a\mu = 3(43t - 41)\mu = 5(26t - 25) - (t - 2) \in I$$

To see that it is injective, let $(\lambda, \mu) \in \ker(\varphi)$. Then $\lambda = 0$ and $a\mu = 0$. As $\tilde{R}$ is an integral domain, $\mu = 0$, too.

For surjectivity, one can solve the relevant system of equations to obtain that

$$\varphi\begin{pmatrix}5\\a-1\end{pmatrix}=\begin{pmatrix}5\\0\end{pmatrix}\qquad\varphi\begin{pmatrix}t-2\\\frac{3}{43}(3a-2)\end{pmatrix}=\begin{pmatrix}t-2\\0\end{pmatrix}$$
$$\varphi\begin{pmatrix}-9\\-2a+2\end{pmatrix}=\begin{pmatrix}0\\1\end{pmatrix}\qquad\varphi\begin{pmatrix}-9a\\10\end{pmatrix}=\begin{pmatrix}0\\a\end{pmatrix}$$

where we use that 43 is invertible in R since

$$43\cdot(2-(t+t^{-1}))=-t^{-1}\cdot f+1$$

It remains to show that $I\oplus I^*$ and $R\oplus R^*$ are not isomorphic. Observe that for any invertible ideal $J\subseteq R$ [30]

$$\begin{aligned}\overline{J}^{-1}=\{x\in \mathrm{Q}(R)\mid x\overline{J}\subseteq R\}&\cong\overline{\mathrm{Hom}_R(J,R)}=J^*\\ x&\mapsto(j\mapsto x\overline{j})\end{aligned}$$

Hence, $R\oplus R^*\cong R^2$ and $I\oplus I^*\cong I\oplus\overline{I}^{-1}$. These modules are distinguished by their determinant[31]. The determinant of R^2 is R, and the determinant of $I\oplus\overline{I}^{-1}$ is $I\overline{I}^{-1}$ which are not isomorphic by direct calculation:
The ideal I is the kernel of the unique ring homomorphism $R\to\mathbb{F}_5$ with $t\mapsto 2$. Hence, $\overline{I}$ is the kernel of the unique ring homomorphism $R\to\mathbb{F}_5$ with $t\mapsto 3$, i.e. $\overline{I}=(5,t-3)$. Note that $f\equiv 3(t-2)(t-3)\mod 5$. Then

$$\overline{I}^{-1}=\left(1,\frac{43t+44}{5}\right)$$

Hence,

$$I\overline{I}^{-1}=\left(5,t-2,43t+44,(t-2)\frac{43t+44}{5}\right)$$

This is not isomorphic to R since 5 and $t-2$ have no common divisor in R as $I=(5,t-2)$ is maximal. ■

Number-theoretically, the ring R is the localization of $\mathbb{Z}[3\alpha]$ at 43, where

$$\alpha:=\frac{1-i\sqrt{19}}{2}$$

and $\tilde{R}$ is the localization of $\mathbb{Z}[\alpha]$ at 43. The ring $\mathbb{Z}[\alpha]$ is the ring of integers of $\mathbb{Q}(i\sqrt{19})$ and a familiar source of counterexamples: For example, it is – by some measures the easiest example of – a principal ideal domain that does not allow a Euclidean function[32].

[30] A *fractional ideal* over an integral domain R is an R-submodule I of the quotient field $\mathrm{Q}(R)$ of R such that there exists an $r\in R$ with $rI\subseteq R$. Fractional ideals form a monoid under the usual multiplication of ideals with the unit given by R. compare [Lan02: Example p.88].

[31] The *determinant* of a projective module M over an integral domain R is the r-th exterior power of M where $r:=\dim_{\mathrm{Q}(R)}(M\otimes_R \mathrm{Q}(R))$ is the rank of M. It is always a projective R-module of rank 1. The determinant of a direct sum is the tensor product of the determinants. As fractional ideals have rank 1, the determinant of a fractional ideal is itself. compare [Lan02: p.735].

[32] The ring $\mathbb{Z}[\alpha]$ can be seen to be a principal ideal domain by calculating its class group, see [Neu99: p.37]. That it does not allow a Euclidean function is completely elementary and hence not in the standard textbooks – which is why we will quickly sketch it here:
By considering the multiplicative norm $\mathbb{Z}[\alpha]\to\mathbb{N}_0$, $\alpha\mapsto|\alpha|^2$, one can see that $\mathbb{Z}[\alpha]^*=\{\pm1\}$. Suppose $d\colon\mathbb{Z}[\alpha]\to\mathbb{N}_0\cup\{-\infty\}$ is a Euclidean function. Choose $m\in\mathbb{Z}[\alpha]\setminus\{0,\pm1\}$ with minimal $d(m)$. There exist $q,r\in\mathbb{Z}[\alpha]$ with $2=mq+r$ and $d(r)<d(m)$. By the minimality, $r\in\{0,\pm1\}$ and $mq\in\{1,2,3\}$. As 2, 3 are irreducible in $\mathbb{Z}[\alpha]=\mathbb{Z}\oplus\mathbb{Z}[\alpha]$ (their norm is 4, 9 and there are no elements of norm 2, 3), we have $m\in\{\pm2,\pm3\}$. Analogously, there exists $q'\in\mathbb{Z}[\alpha]$ with $mq'\in\{\alpha,\alpha\pm1\}$. Hence, an element of this set must be divisible by 2 or 3 – which they are not. Contradiction!

We now turn these examples into forms with codomain $\mathbb{Q}(t)/\mathbb{Z}[t^{\pm 1}]$, giving us the desired knot-theoretic examples:

Proposition 5.19. *Let $f \in \mathbb{Z}[t^{\pm 1}]$.*

(1) *The map*

$$\begin{aligned} \iota\colon \mathbb{Z}[t^{\pm 1}]/(f) &\to \left\langle \tfrac{1}{f} \right\rangle \subseteq \mathbb{Q}(t)/\mathbb{Z}[t^{\pm 1}] \\ [x] &\mapsto \left[\frac{x}{f}\right] \end{aligned}$$

is an isomorphism of $\mathbb{Z}[t^{\pm 1}]$-modules.

(2) *Let $\epsilon \in \{\pm 1\}$ and $\Phi\colon M \times M \to \mathbb{Z}[t^{\pm 1}]/(f)$ be a non-singular ϵ-Hermitian form over $\mathbb{Z}[t^{\pm 1}]/(f)$. The projection $\mathbb{Z}[t^{\pm 1}] \twoheadrightarrow \mathbb{Z}[t^{\pm 1}]/(f)$ induces a $\mathbb{Z}[t^{\pm 1}]$-module structure on M. Then Φ and $\iota \circ \Phi$ are non-singular ϵ-Hermitian forms over $\mathbb{Z}[t^{\pm 1}]$.*

Proof.

(1) This follows since for $x \in \mathbb{Z}[t^{\pm 1}]$

$$\frac{x}{f} \in \mathbb{Z}[t^{\pm 1}] \quad \Leftrightarrow \quad x \in (f)$$

(2) By construction of the restriction of scalars, $\Phi\colon M \times M \to \mathbb{Z}[t^{\pm 1}]/(f)$ also defines a non-singular ϵ-Hermitian form over $\mathbb{Z}[t^{\pm 1}]$, and $\iota \circ \Phi\colon M \times M \to \mathbb{Q}(t)/\mathbb{Z}[t^{\pm 1}]$ is an ϵ-Hermitian form. It is non-singular, since if $\varphi\colon M \to \mathbb{Q}(t)/\mathbb{Z}[t^{\pm 1}]$ is a monomorphism, we have for all $x \in M$

$$f \cdot \varphi(x) = \varphi(f \cdot x) = \varphi(0) = 0$$

Hence, $\operatorname{im}(\varphi) \subseteq \langle \frac{1}{f} \rangle$. ■

Lemma 5.20. *Let $q \geq 2$. There exist simple oriented $(2q-1)$-dimensional knots $A, B, C \subseteq S^{2q+1}$ such that*

- *$A\#C \cong B\#C$*
- *$\mathrm{Al}_A \ncong \mathrm{Al}_B$, in particular, $A \ncong B$.*

Proof. As in Lemma 5.18 let $f := 43t^2 - 85t + 43 \in \mathbb{Z}[t^{\pm 1}]$, $R := \mathbb{Z}[t^{\pm 1}]/(f)$ $I := (5, t-2) \subseteq R$, $a := \frac{43t-41}{3}$ and $\tilde{R} := R[a]$. We have seen there that for $\epsilon \in \{\pm 1\}$

$$\Phi^\epsilon(I) \oplus \Phi^\epsilon(\tilde{R}) \cong \Phi^\epsilon(R) \oplus \Phi^\epsilon(\tilde{R})$$

and $I \oplus I^* \ncong R \oplus R^*$.
From Proposition 5.19 we obtain a monomorphism

$$\begin{aligned} \iota\colon \mathbb{Z}[t^{\pm 1}]/(f) &\to \mathbb{Q}(t)/\mathbb{Z}[t^{\pm 1}] \\ [x] &\mapsto \left[\frac{x}{f}\right] \end{aligned}$$

and non-singular ϵ-Hermitian forms

$$\Phi_A^\epsilon := \iota \circ \Phi^\epsilon(I) \quad \Phi_B^\epsilon := \iota \circ \Phi^\epsilon(R) \quad \Phi_C^\epsilon := \iota \circ \Phi^\epsilon(\tilde{R})$$

If $\epsilon = -1$, in particular if $q = 2$, their scalar forms have signature 0 as the bilinear form

$$\begin{aligned} (V \oplus V^*) \times (V \oplus V^*) &\to \mathbb{Q} \\ ((x, f), (y, g)) &\mapsto f(y) - g(x) \end{aligned}$$

has signature 0 for any finite dimensional $\mathbb{Q}$-vector space V.
Furthermore, $t - 1$ is a unit in R since

$$(t-1)(-43t + 42) = -f + 1$$

This implies that the $\mathbb{Z}[t^{\pm 1}]$-module structure induced by $\mathbb{Z}[t^{\pm 1}] \twoheadrightarrow R$ on a finitely generated R-module is of type K.
The claim now follows from Theorem 5.14 and Proposition 5.15. ■

Theorem 5.21. *Let $q \geq 2$. The monoid $\mathcal{K}_{2q-1}$ is not a unique factorization monoid.*

Proof. In Lemma 5.20 we found an element $K \in \mathcal{K}_{2q-1}$ that is not cancellable. By Theorem 4.20 (2) the neutral element is the only unit of $\mathcal{K}_{2q-1}$, so K is also not weakly cancellable. Hence, the claim follows from Proposition 2.5 (1). ∎

For $q = 1$, i.e. in the classical dimension, $\mathcal{K}_1$ is a unique factorization monoid by a result of H. Schubert:

Theorem 5.22. *The monoid $\mathcal{K}_1$ is a unique factorization monoid.*

Proof. see [Sch49: Satz 7] or [Lic97: Theorem 2.12] for an English textbook account ∎

5.4. String knots

The examples we obtained in the last section were exclusively in odd dimensions. To find examples in even dimensions, we 'spin' them to increase their dimension by 1. To make this construction precise, we first need to introduce knotted strings in balls as an intermediate step:[33]

Definition. Let $n \geq 1$. An *n-dimensional string knot* is a smooth submanifold $K \subseteq \overline{B}^{n+2}$ diffeomorphic to $\overline{B}^n$ that is *trivial near the boundary*, i.e.

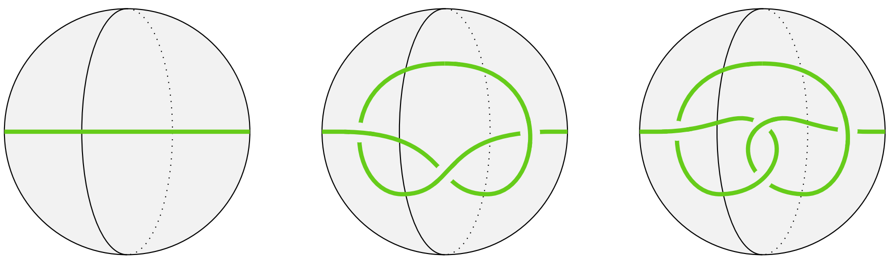
$$K \cap (\overline{B}^{n+2} \setminus B^{n+2}_{\frac{8}{9}}) = (\overline{B}^n \times \{0\}) \cap (\overline{B}^{n+2} \setminus B^{n+2}_{\frac{8}{9}})$$

The *trivial n-dimensional string knot* is $\overline{B}^n \times \{0\} \subseteq \overline{B}^{n+2}$.
Two n-dimensional string knots $K, L \subseteq \overline{B}^{n+3}$ are *smoothly isotopic* if there is a *smooth isotopy* between them, i.e. a smooth map $F\colon K \times [0,1] \to \overline{B}^{n+2}$ such that

- $F_0\colon K \times \{0\} \to K$ and $F_1\colon K \times \{1\} \to L$ are diffeomorphisms.
- for each $t \in [0,1]$, $F_t(K) \subseteq \overline{B}^{n+2}$ is a string knot.

Let $\mathcal{S}_n$ be the set of isotopy classes of n-dimensional string knots.

Example 5.23. A trivial and two non-trivial string knots:



We can glue a trivial string knot to the boundary of a string knot to obtain a knot:

Definition. Consider the smooth embeddings[34]

$$\iota_\pm\colon \overline{B}^{n+2} \to S^{n+2}$$
$$x = (x_1, \ldots, x_{n+2}) \mapsto \frac{1}{\|x\|^2 + 1}\Big((-1)^{n+1} \cdot 2x_1, \ldots, 2x_{n+2}, \pm(1 - \|x\|^2)\Big)$$

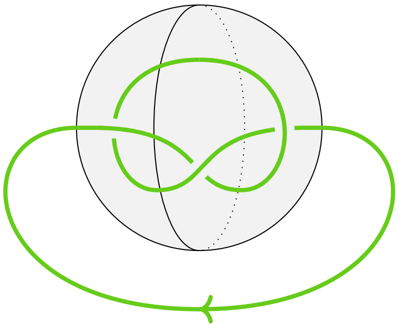
The *closure* of an n-dimensional string knot $K \subseteq \overline{B}^{n+2}$ is the n-dimensional knot

$$\mathrm{cl}(K) \coloneqq \iota_+(K) \cup \iota_-(\overline{B}^n \times \{0\}) \subseteq S^{n+2} \subseteq S^{n+2}$$

oriented such that $\iota_-\colon (\overline{B}^{n+2}, \overline{B}^n \times \{0\}) \to (S^{n+2}, \mathrm{cl}(K))$ is orientation-preserving.

[33]We loosely follow [Fri25: Section 96.4].
[34]compare [Fri25: Example p.297]

Example 5.24. A string knot whose closure is a trefoil:



It might seem counterintuitive that we can obtain an *oriented* knot from a string knot despite the latter carrying no orientation. Notice, however, that since all string knots share the same boundary one can compatibly choose an orientation on all of them by fixing the boundary orientation.

Theorem 5.25. *The closure operation gives a well-defined bijection* $\mathrm{cl}\colon \mathcal{S}_n \to \mathcal{K}_n$ *for all* $n \geq 1$*.*

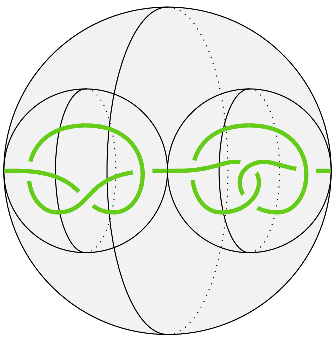
Proof. see [Fri25: Proposition 96.11], or use the following sketch to construct an inverse: Consider the smooth embeddings $\iota_\pm\colon \overline{B}^{n+2} \to S^{n+2}$ from the definition of the closure operation. Let $K \subseteq S^{n+2}$ be an oriented n-dimensional knot and $\varphi\colon (\overline{B}^{n+2}, \overline{B}^n) \to (S^{n+2}, K)$ be an orientation-preserving smooth embedding. By Theorem 1.7 we may assume after an isotopy of K that $\varphi = \iota_-$. Then $\iota_+^{-1}(K) \subseteq \overline{B}^{n+2}$ is a string knot. The Isotopy Extension Theorem for pairs from the proof of Theorem 1.7 together with an argument similar to Theorem 1.14 shows that its smooth isotopy class is independent from the choice of smooth isotopy in the above and from the choice of representative for the smooth isotopy class of K. Hence, this gives a well-defined map $\mathcal{K}_n \to \mathcal{S}_n$. It is easy to see that this is the desired inverse. ∎

Under this bijection, the connected sum operation corresponds to the following operation akin to boundary connected sum:

Definition. Let $n \geq 1$. Consider the diffeomorphisms $\Phi_\pm\colon \overline{B}^{n+2} \to \overline{B}^{n+2}_{\frac{1}{2}} \pm (\frac{1}{2}, 0, \ldots, 0)$ given by translation and radial shrinking. The *boundary connected sum* of two n-dimensional string knots $K, L \subseteq \overline{B}^{n+3}$ is

$$K\#L \coloneqq (\Phi_+(K) \cup \Phi_-(L)) \cup \left(\overline{B}^n \times \{0\} \cap (\overline{B}^{n+3} \setminus (\mathrm{im}(\Psi_+) \cup \mathrm{im}(\Psi_-)))\right)$$

Example 5.26. The boundary connected sum of two string knots.



Proposition 5.27. *Let* $n \geq 1$*.*

(1) *The boundary connected sum defines an abelian monoid structure on* $\mathcal{S}_n$ *with the neutral element given by the trivial string knot.*

(2) *The map* $\mathrm{cl}\colon \mathcal{S}_n \to \mathcal{K}_n$ *is a monoid isomorphism.*

Proof. see [Fri25: Proposition 96.13, Remark p.1951] ∎

We will later also need the observation that the closure operation does not change the complement away from the boundary:

Proposition 5.28. *Let $n \geq 1$ and $K \subseteq \overline{B}^{n+2}$ be an n-dimensional string knot. Then*

$$B^{n+2} \setminus K \cong S^{n+2} \setminus \mathrm{cl}(K)$$

Proof. The key observation is that in $S^{n+2} \setminus \mathrm{cl}(K)$ the south pole S has been removed. A diffeomorphism $S^{n+2} \setminus S \cong B^{n+2}$ can then be adapted such that it pushes the lower hemisphere into the trivial near the boundary part of $K \subseteq B^{n+2}$. We are sure that the reader will not find it enlightening if we write down a concrete map. ∎

5.5. Spinning

The spinning operation turns a string knot into a knot one dimension higher by spinning it around a circle and closing off the end. With the closure bijection from last section, we can transfer it to an operation on knots:[35]

Definition. Let $n \geq 1$. Consider the smooth map

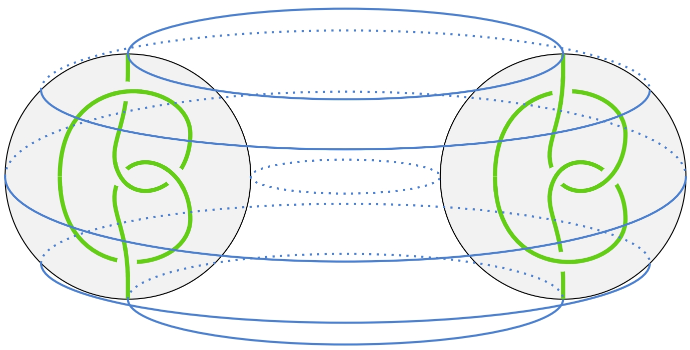
$$\begin{aligned} \Phi \colon S^1 \times \overline{B}^{n+2} &\to S^{n+3} \\ (x, y) &\mapsto \left(\sqrt{1 - \|y\|^2} \cdot x, y\right) \end{aligned}$$

The *spin* of an n-dimensional string knot $K \subseteq \overline{B}^{n+2}$ is the $(n+1)$-dimensional knot

$$\sigma(K) \coloneqq \Phi(K \times S^1)$$

We orient $\sigma(K)$ such the orientations of $S^{n-1} \times \{0\}, \sigma(K) \subseteq S^{n+1}$ agree on their intersection. The *spin* of the n-dimensional knot $\mathrm{cl}(K)$ also is $\sigma(\mathrm{cl}(K)) \coloneqq \sigma(K)$.

Example 5.29. When spinning a string knot the blue S^1's each get mapped to a single point.



Example 5.30. Let $U \coloneqq \overline{B}^n \times \{0\} \subseteq \overline{B}^{n+2}$ be the trivial string knot. Its spin is the unknot

$$\Phi((\overline{B}^n \times \{0\}^2) \times K) = S^{n-1} \times \{0\} \subseteq S^{n+3}$$

We naturally begin by showing the spinning operation is well-defined:

Lemma 5.31. *Let $n \geq 1$.*

(1) *The spin of an n-dimensional string knot $K \subseteq \overline{B}^{n+2}$ is an $(n+1)$-dimensional oriented knot whose smooth isotopy class only depends on the smooth isotopy class of K.*
(2) *The spin of an n-dimensional knot $K \subseteq \overline{S}^{n+2}$ is an $(n+1)$-dimensional oriented knot whose smooth isotopy class only depends on the smooth isotopy class of K.*

[35]This construction was developed by E. Artin [Art25] as the first way to construct non-trivial higher dimensional knots.

Proof.
(1) Let $\Phi\colon S^1 \times \overline{B}^{n+2} \to S^{n+3}$ be the map from the definition. The spin $\Phi(S^1 \times K) \subseteq S^{n+3}$ is a smooth submanifold by the following observations:
   - Since $\Phi\colon S^1 \times B^{n+2} \to S^{n+3}$ is a smooth embedding, $\Phi(S^1 \times K) \subseteq S^{n+3}$ is a smooth submanifold.
   - As $K \subseteq \overline{B}^{n+2}$ is trivial near the boundary, $\Phi(S^1 \times K)$ is equal to the spin of a trivial string knot in a neighbourhood of $\Phi(S^1 \times S^{n+1}) = \{0\} \times S^{n-1}$. This is a smooth submanifold by Example 5.30.

   Observe $\Phi\colon S^1 \times \overline{B}^{n+2} \to S^{n+3}$ descends to a bijection
$$(S^1 \times \overline{B}^{n+2})/{\sim} \to S^{n+3}$$
   where $(x, y) \sim (x, y')$ for $x \in S^{n+1}$, $y, y' \in S^1$, and similarly for $S^1 \times \overline{B}^{n+2} \to S^{n+1}$. Since $S^1 \times K \cong S^1 \times \overline{B}^n$, this can be used to construct a diffeomorphism $\Phi(S^1 \times K) \cong S^{n+1}$, which shows that $\sigma(K) \subseteq S^{n+3}$ is an $(n+1)$-dimensional knot.
   If $F\colon K \times [0,1] \to \overline{B}^{n+2}$ is a smooth isotopy, triviality near the boundary analogously implies that $\Phi \circ F$ also is a smooth isotopy.
(2) This follows immediately from (1) and Theorem 5.25. ∎

To show that this operation is interesting, we need to understand how it interacts with knot invariants:

Proposition 5.32. *Let $n \geq 1$ and $K \subseteq \overline{S}^{n+2}$ be an oriented n-dimensional knot.*
(1) $\pi_1(S^{n+2} \setminus K) \cong \pi_1(S^{n+3} \setminus \sigma(K))$
(2) *for $i \geq 1$*
$$\mathrm{H}_i((\widetilde{S^{n+3} \setminus \sigma K})_{\mathrm{ab}}) \cong \begin{cases} \mathrm{H}_i((\widetilde{S^{n+2} \setminus K})_{\mathrm{ab}}) \oplus \mathrm{H}_{i-1}((\widetilde{S^{n+2} \setminus K})_{\mathrm{ab}}), & \text{if } i \geq 2 \\ \mathrm{H}_1((\widetilde{S^{n+2} \setminus K})_{\mathrm{ab}}), & \text{if } i = 1 \end{cases}$$
   as $\mathbb{Z}[t^{\pm 1}]$-modules

Proof. By Theorem 5.25 there exists n-dimensional string knot $\tilde{K} \subseteq \overline{B}^{n+2}$ with closure K. Then $B^{n+2} \setminus \tilde{K} \cong S^{n+2} \setminus K$ by Proposition 5.28. Since $\tilde{K}$ is trivial near the boundary, the inclusion $X := \overline{B}^{n+2}_{\frac{9}{10}} \setminus \tilde{K} \to B^{n+2} \setminus \tilde{K}$ is a homotopy equivalence. Let $\Phi\colon S^1 \times \overline{B}^{n+2} \to S^{n+3}$ be the smooth map from the definition of spinning.
We decompose $S^{n+3} \setminus \sigma(K) = A \cup B$ where
$$A := \Phi(S^1 \times X)$$
$$B := S^{n+3} \setminus \left(\mathring{A} \cup (S^{n-1} \times \{0\})\right)$$
$$A \cap B = \Phi(S^1 \times (S^{n+1}_{\frac{9}{10}} \setminus \partial\tilde{K})) = \Phi(S^1 \times (S^{n+1}_{\frac{9}{10}} \setminus (S^{n-1}_{\frac{9}{10}} \times \{0\})))$$
The smooth submanifold
$$\mu := \{0\} \times S^1 \subseteq S^{n+1} \setminus (S^{n-1} \times \{0\})$$
is a 'meridian' of all string knots. We have a commutative diagram
$$\begin{tikzcd} S^1 \times X \arrow[d, "\Phi"', "\cong"] & S^1 \times \mu \arrow[l, hook'] \arrow[r, "\mathrm{pr}"] \arrow[d, "\Phi"', "\simeq"] & \mu \arrow[d, "\Phi"', "\simeq"] \\ A & A \cap B \arrow[l, hook'] \arrow[r, hook] & B \end{tikzcd}$$
where pr is the projection. To see that $\Phi\colon \mu \to B$ is a homotopy equivalence, observe first that $S^{n+3} \setminus \mathring{A} \cong S^{n+1} \times \overline{B}^2 \simeq S^{n+1}$ (compare [Fri25: Lemma 98.6]) and then note that $\Phi(\mu)$ is an unknot in this S^{n+1}. The middle map is a homotopy equivalence for similar reasons.
Set $\pi := \pi_1(\overline{B}^{n+2}_{\frac{9}{10}} \setminus \tilde{K}) \cong \pi_1(S^{n+2} \setminus K)$.

(1) Apply the Seifert–van Kampen Theorem to the decomposition $S^{n+3} \setminus \sigma(K) = A \cup B$. The above diagram then yields a pushout

$$\begin{tikzcd} \mathbb{Z} \times \mathbb{Z} \arrow[r] \arrow[d] & \mathbb{Z} \arrow[d] \\ \mathbb{Z} \times \pi \arrow[r] & \pi_1(S^{n+3} \setminus \sigma(K)) \end{tikzcd}$$

with the top map given by the projection onto the second factor. The claim now follows since π also has the universal property of this pushout.

(2) Let $p\colon \widetilde{(S^{2q+2} \setminus \sigma K)}_{\mathrm{ab}} \to S^{n+3} \setminus \sigma(K)$ be the universal abelian covering. By elementary set theory

$$\widetilde{(S^{n+3} \setminus \sigma K)}_{\mathrm{ab}} = p^{-1}(A) \cup p^{-1}(B)$$

and $p^{-1}(A) \cap p^{-1}(B) = p^{-1}(A \cap B)$. To understand these preimages we make repeated use of [Fri25: Proposition 115.7] and the classification of coverings:
Observe that restricting $\pi_1(S^{n+3} \setminus \sigma(K)) \to \pi_1(S^{n+3} \setminus \sigma(K))_{\mathrm{ab}}$ to the fundamental groups of A, B and $A \cap B$ still gives an epimorphism by the calculation in (1). Hence, $p^{-1}(A)$, $p^{-1}(B)$ and $p^{-1}(C)$ are connected and the inclusions induce isomorphisms

$$(*) \qquad \mathrm{Deck}(p^{-1}(A) \to A) \cong \mathrm{Deck}(p^{-1}(A \cap B) \to A \cap B) \cong \mathrm{Deck}(p^{-1}(B) \to B) \cong \mathbb{Z}$$

We thereby can consider the homology of $p^{-1}(A)$, $p^{-1}(B)$ and $p^{-1}(C)$ as $\mathbb{Z}[t^{\pm 1}]$-modules.

- The covering $p^{-1}(A) \to A$ corresponds to the subgroup

$$\Phi_*(\pi_1(S^1) \times [\pi : \pi])$$

Hence, $p^{-1}(A)$ is homeomorphic to $S^1 \times \widetilde{X}_{\mathrm{ab}}$.

- The covering $p^{-1}(B) \to B$ corresponds to the subgroup

$$\Phi_*(\pi_1(\mu)) = \pi_1(B)$$

Since $B \simeq S^1$, it follows that $p^{-1}(B)$ is contractible.

- The covering $p^{-1}(A \cap B) \to A \cap B$ corresponds to the subgroup

$$\Phi_*(\pi_1(S^1) \times \{0\})$$

Hence, $p^{-1}(A \cap B)$ is homotopy equivalent to $S^1 \times \mathbb{R} \simeq S^1$.

More precisely, we have a commutative diagram

$$\begin{tikzcd} S^1 \times \widetilde{X}_{\mathrm{ab}} \arrow[d, "\cong"] & S^1 \arrow[l, "i"'] \arrow[r] \arrow[d, "\simeq"] & \{*\} \arrow[d, "\simeq"] \\ p^{-1}(A) & p^{-1}(A \cap B) \arrow[l] \arrow[r] & p^{-1}(B) \end{tikzcd}$$

where i is given by $S^1 \times \{*\} \to S^1 \times \widetilde{X}_{\mathrm{ab}}$ for some $* \in \widetilde{X}_{\mathrm{ab}}$. Now consider the Mayer–Vietoris sequence of this decomposition. All morphisms in this sequence are homomorphisms of $\mathbb{Z}[t^{\pm 1}]$-modules by $(*)$.[36] The map $i\colon S^1 \to S^1 \times \widetilde{X}_{\mathrm{ab}}$ induces a monomorphism on all homology groups since its composition with the projection is the identity. Hence, the boundary maps vanish. For $i = 1$, the Künneth Theorem shows that the sequence is given by

$$0 \to \mathbb{Z} \xrightarrow{f} \mathbb{Z} \oplus \mathrm{H}_1(\widetilde{X}_{\mathrm{ab}}) \to \mathrm{H}_1(\widetilde{(S^{n+3} \setminus \sigma K)}_{\mathrm{ab}}) \to 0$$

where the first coordinate of f is an isomorphism, implying the claim.

[36] To see that the boundary maps are homomorphisms of $\mathbb{Z}[t^{\pm 1}]$-modules, remember that the $\mathbb{Z}[t^{\pm 1}]$-module structure is already defined on chain level. Hence, the short exact sequence of chain complexes arising in the proof of Mayer–Vietoris is a sequence of $\mathbb{Z}[t^{\pm 1}]$-module complexes.
Taking the point of view from p. 54, this is a special case of the Mayer–Vietoris sequence with twisted coefficients.

For $i \geq 2$, the Mayer–Vietoris sequence yields

$$\mathrm{H}_i((\widetilde{S^{n+3} \setminus \sigma K})_{\mathrm{ab}}) \cong \mathrm{H}_i(S^1 \times \widetilde{X}_{\mathrm{ab}})$$

Now calculate $\mathrm{H}_i(S^1 \times \widetilde{X}_{\mathrm{ab}})$ not using Künneth, but by decomposing the S^1 into the upper and lower hemisphere and applying the Mayer–Vietoris sequence as in [Fri25: Proposition 139.7]. Then one can as above consider the $\mathbb{Z}$-actions to get the remaining isomorphism of $\mathbb{Z}[t^{\pm 1}]$-modules. ∎

We want to spin the example of non-cancellation from last section to an example one dimension higher. To do so, we need that spinning is compatible with the connected sum operation:

Proposition 5.33. *Let $n \geq 1$. Spinning defines a monoid homomorphism $\sigma\colon \mathcal{K}_n \to \mathcal{K}_{n+1}$.*

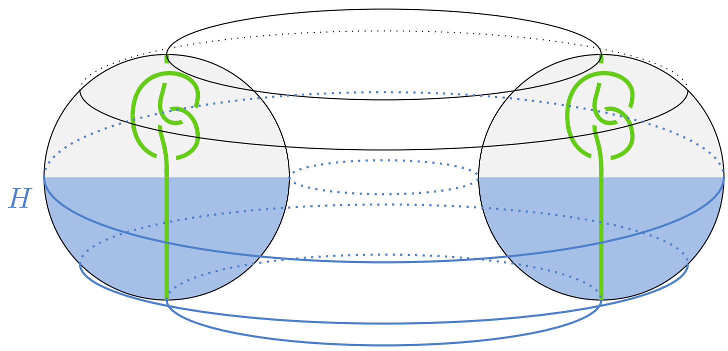
Proof. By Proposition 5.27 (2) it suffices to show that $\sigma\colon \mathcal{S}_n \to \mathcal{K}_{n+1}$ is a monoid homomorphism. In Example 5.30 we have already shown that the spin of the trivial string knot $U \subseteq \overline{B}^{n+2}$ is trivial.
Let $\Phi\colon S^1 \times \overline{B}^{n+2} \to S^{n+3}$ be the smooth map from the definition of spinning. Let $K, L \subseteq \overline{B}^{n+2}$ be n-dimensional string knots. By Proposition 5.27 (1) it suffices to show that

$$(*) \qquad \sigma\big((U\#K)\#(L\#U)\big) \cong \sigma(K)\#\sigma(L)$$

Observe that

$$H \coloneqq \Phi\Big(S^1 \times \{(x_1, \dots, x_{n+2}) \in \overline{B}^{n+2} \mid x_1 \leq 0\}\Big) = \Big\{(x_0, \dots, x_{n+3}) \in S^{n+3} \;\Big|\; x_2 \leq 0\Big\}$$

is a 'lower hemisphere' with $H \cap \sigma(U\#K) = H \cap \sigma(U)$. Therefore, Example 5.30 implies that $(H, H \cap \sigma(U\#K))$ is diffeomorphic to $(\overline{B}^{n+2}, \overline{B}^n)$.



Similarly, one can proceed for the corresponding upper hemisphere and $L\#U$. Using these pairs $(\overline{B}^{n+2}, \overline{B}^n)$ to construct the connected sum $\sigma(K)\#\sigma(L)$ directly shows $(*)$. ∎

It is now straightforward to spin the example from last section and obtain a counterexample for cancellation in even dimensions. Here comes into play that we carefully constructed our example knots to be distinguished by their Alexander module – not only by the Blanchfield form. Since the former behaves well under spinning, we can still use it to distinguish the spun knots:[37]

Theorem 5.34. *Let $q \geq 2$. The monoid $\mathcal{K}_{2q}$ is not a unique factorization monoid.*

Proof. By Lemma 5.20 there exist simple oriented $(2q-1)$-dimensional knots $A, B, C \subseteq S^{2q+1}$ such that $A\#C \cong B\#C$ and $\mathrm{Al}_A \not\cong \mathrm{Al}_B$. By Proposition 5.33, $\sigma(A)\#\sigma(C) \cong \sigma(B)\#\sigma(C)$. By Proposition 5.16 and Proposition 5.32 (2), $\sigma(A) \not\cong \sigma(B)$. Hence, $\sigma(C)$ is not cancellable in $\mathcal{K}_{2q}$. By Theorem 4.20 (2) the neutral element is the only unit of $\mathcal{K}_{2q}$, so $\sigma(C)$ is also not weakly cancellable. The claim hence follows from Proposition 2.5 (1). ∎

[37] The idea is of course again from [Bay85].

Summing up the results of this chapter, we have shown that $\mathcal{K}_n$ is a unique factorization monoid for $n = 1$ and not a unique factorization monoid for $n \geq 3$. The case $n = 2$ is missing. Our method is not applicable as there are no 1-dimensional examples which we could spin into 2-dimensional ones. In fact, it appears that the question whether $\mathcal{K}_2$ is a unique factorization monoid remains open.
As an aside we want to conclude this chapter by constructing interesting examples of spinning not being injective. Incidentally, this gives a monoid homomorphism with trivial kernel which is not injective:[38]

Example 5.35. We only consider the classical case $\sigma\colon \mathcal{K}_1 \to \mathcal{K}_2$.

- If $\sigma(K) \subseteq S^4$ is unknotted, Proposition 5.32 and Theorem 4.10 imply that $\pi_1(S^3 \setminus K) \cong \mathbb{Z}$ and for $i \geq 1$
$$\mathrm{H}_i(\widetilde{S^3 \setminus K}_{\mathrm{ab}}) = 0$$
and thereby further that $K \subseteq S^3$ is trivial.
- Let $\rho\colon \overline{B}^3 \to \overline{B}^3,\ (x, y, z) \mapsto (-x, y, z)$ be an orientation-reversing diffeomorphism and consider the two tangles below:

T $\qquad$ $-T := \rho(T)$

Their closures are a trefoil T and its inverse[39] $-T$ which are not smoothly isotopic by [Fri25: Proposition 102.15]. The diffeomorphism
$$\begin{aligned} f\colon S^1 \times \overline{B}^3 &\to S^1 \times \overline{B}^3 \\ (x, y) &\mapsto (-x, \rho(y)) \end{aligned}$$
is orientation-preserving and compatible with the spinning map $\Phi\colon S^1 \times \overline{B}^3 \to S^4$ in the sense that there exists an orientation-preserving diffeomorphism $g\colon S^4 \to S^4$ with $\Phi \circ f = g \circ \Phi$. Since the restriction $f\colon S^1 \times T \to S^1 \times \rho(T)$ is also orientation-preserving, $g\colon (S^4, \sigma(T)) \to (S^4, \sigma(-T))$ is an orientation-preserving diffeomorphism of pairs. Hence, the oriented knots $\sigma(T)$ and $\sigma(-T)$ are smoothly isotopic by Theorem 1.14.
- One can go even further: The (string) knots $T\#T$ and $T\# -T$ are also not smoothly isotopic – not even after mirroring (see [Fri25: Exercise 102.8]). But Proposition 5.33 implies that their spins are still smoothly isotopic.

[38] see [Gor76]

[39] A priori, one could think that $\rho(T)$ closes to the mirror of T, but going carefully through the orientation conventions will reveal that it really closes to the inverse. Observe in this regard that mirroring the y or z coordinate might produce a different string than mirroring x – this string will close to the mirror.

A. Bass–Serre theory and accessibility

A.1. Groups acting on trees

In this appendix we are motivated by the question raised before Lemma 4.19: Given a group G, is there a bound on the number of factors in a decomposition of G into an amalgamated free product? There is a generalization in which this can be nicely resolved, namely graphs of groups. Before we discuss those, let us recall some basic notions of graph theory:[40] [41] [42]

Definition. A *graph* consists of a set of *vertices* Γ_V, a set of *edges* Γ_E with two maps *origin*, *terminus* $o, t\colon \Gamma_E \to \Gamma_V$ and an *inversion* $\Gamma_E \to \Gamma_E,\ e \mapsto \overline{e}$ such that for all $e \in \Gamma_E$

$$e \neq \overline{e} \qquad \overline{\overline{e}} = e \qquad o(e) = t(\overline{e})$$

A graph Γ is *finite* (resp. *non-empty*) if Γ_V and Γ_E are finite (resp. non-empty).
A *morphism* of graphs $f\colon \Gamma \to \Gamma'$ consists of maps $f_V\colon \Gamma_V \to \Gamma'_V$, $f_E\colon \Gamma_E \to \Gamma'_E$ such that for all $e \in \Gamma_E$

$$f_V(o(e)) = o(f_E(e)) \quad f_V(t(e)) = t(f_E(e)) \quad f_E(\overline{e}) = \overline{f_E(e)}$$

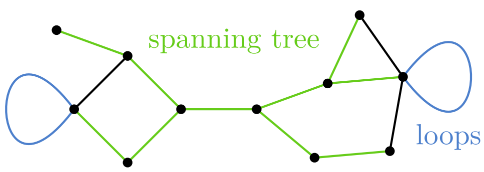
A *subgraph* $\Gamma' \subseteq \Gamma$ is a graph Γ' such that $\Gamma'_V \subseteq \Gamma_V$, $\Gamma'_E \subseteq \Gamma_E$ and the origin, terminus and inversion maps of Γ' are the restrictions of the respective maps of Γ.
For an edge $e \in \Gamma_E$ we consider the subgraph Γ^e with $\Gamma^e_V = \Gamma_V$ and $\Gamma^e_E = \Gamma_E \setminus \{e, \overline{e}\}$.

The graphs we consider are unoriented, but every intuitive edge exists twice – once in either direction. When drawing graphs we will therefore only draw one edge for every pair $e, \overline{e}$.
A type of graph of particular interest is a tree:

Definition. For $n \geq 0$ the graph I^n has $I^n_V := \{0, \dots, n\}$, $I^n_E := \{1, \overline{1}, \dots, n, \overline{n}\}$ and for $i \in \{1, \dots, n\}$

$$\begin{aligned} o(i) &= i-1 \qquad & t(i) &= i \\ o(\overline{i}) &= i \qquad & t(\overline{i}) &= -i \end{aligned}$$

Let Γ be a graph. For $n \geq 0$, a morphism $\gamma\colon I^n \to \Gamma$ is a *path of length n from $\gamma_V(0)$ to $\gamma_V(n)$* in Γ. The path is a *geodesic* if γ_V is injective.
The graph Γ is *connected* if for all vertices $v, w \in \Gamma_V$ there exists a path from v to w. Otherwise, Γ is *disconnected.*
An edge $e \in \Gamma_E$ is a *loop* if $t(e) = o(e)$.
A graph Γ is a *tree* if Γ_E does not contain any loops and for $v, w \in \Gamma_V$ there exists a unique geodesic from v to w.
A *subtree* of a graph is a subgraph that is a tree. A *spanning tree* of a graph Γ is a subtree T of Γ such that $T_V = \Gamma_V$.



[40]This appendix is primarily based on the author's notes of a lecture by H. Wilton at the University of Cambridge. A literature reference is [SW79: Section 4].
[41]As this is merely an appendix, we allow ourselves to be a bit briefer in some arguments than we would otherwise be.
[42]Our definition of graphs is taken from [Ser80: Section 2.1].

The following proposition gives a criterion for checking whether a graph is a tree. With it we can show that every finite graph has a spanning tree:[43]

Proposition A.1. *Let Γ be a graph.*

(1) *Let $v, w \in \Gamma_V$. There exists a path from v to w if and only if there exists a geodesic from v to w.*
(2) *The graph Γ is a tree if and only if it is connected and for every $e \in \Gamma_E$ the subgraph Γ^e is disconnected.*
(3) *If Γ is finite and connected, it admits a spanning tree.*

Proof.

(1) The 'only if'-direction is obvious. For the 'if'-direction let $n \in \mathbb{N}$ be minimal such that there exists a path $\gamma\colon I^n \to \Gamma$ of length n from v to w. It is not hard to see that this path is a geodesic.
(2) Assume Γ is a tree and let $e \in \Gamma_E$. Since e is not a loop $o(e) \neq t(e)$. Suppose the graph Γ^e is connected. By (1) there exists a geodesic from v to w in Γ^e. This is then also a geodesic in Γ. But the unique geodesic in Γ from v to w consists only of the edge e. Contradiction!
For the reverse direction, let Γ be an arbitrary graph and observe:
   - If $e \in \Gamma_E$ is a loop, Γ^e is connected.
   - Let $v, w \in \Gamma_V$ and $\gamma\colon I^n \to \Gamma$, $\gamma'\colon I^{n'} \to \Gamma$ geodesics from v to w. If they are distinct, there exists $e \in \operatorname{im}(\gamma_E)$ with $e \notin \operatorname{im}(\gamma'_E)$. Then Γ^e is connected.
(3) If Γ is finite, we proceed inductively on $\frac{1}{2}\#\Gamma_E$. If $\frac{1}{2}\#\Gamma_E = 0$, $\Gamma_V = \{v\}$ since Γ is connected. Then Γ is a tree.
For the inductive step observe that if Γ^e is disconnected for all $e \in \Gamma_E$, Γ itself is a tree by (2). Otherwise, there exists an edge $e \in \Gamma_E$ such that Γ^e is connected. By induction, a spanning tree for Γ^e exists. It is also a spanning tree for Γ. ■

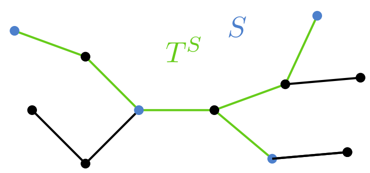
The defining property of a tree is that any two vertices are joined by a unique geodesic. This extends to arbitrary collections of vertices which determine unique subtrees:

Proposition A.2. *Let T be a tree and $S \subseteq T_V$ a subset.*

(1) *There exists a unique subgraph T^S of T such that*
   (i) *$S \subseteq T^S_V$ and T^S is connected.*
   (ii) *If T' is a connected subgraph of T with $S \subseteq T'_V$, T^S is a subgraph of T'.*
(2) *If S is finite, T^S is finite.*

Proof.

(1) Let T^S_E be the set of edges $e \in T_E$ that lie on the (unique) geodesic from v to w for some $v, w \in S$ and set
$$T^S_V := o(T^S_E) \cup t(T^S_E) \cup S$$
Then T^S is a subgraph of T. Since T is connected, T^S is also connected and Proposition A.1 (1) implies that every pair of vertices in T^S is connected by a geodesic. Since T is a tree and geodesics of T^S are geodesics in T, it follows that T^S is a tree.
The minimality condition (ii) follows from the uniqueness of the geodesics in the tree T and again the observation that geodesics in T' remain geodesics in T.
(2) If S is finite, T^S_E is finite as paths have finite lengths. Then T^S_V is also finite. ■

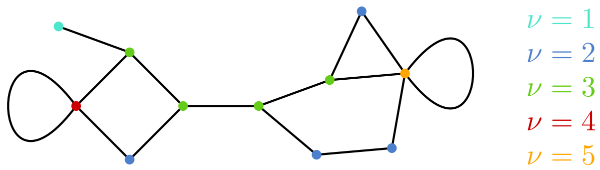


[43] This finiteness assumption is purely technical. It only enables us to give nice inductive proofs instead of having to argue with exhaustions. The same holds for most of the finiteness assumptions in the following.
We will see in Footnote 46 that we do not lose much by this restriction.

For later reference we note the following:

Definition. Let Γ be a graph. The *valence* of a vertex $v \in \Gamma_V$ is

$$\nu(v) := \# o^{-1}(v)$$



Lemma A.3. *If T is a finite tree and $T_E \neq \varnothing$, there exist at least two vertices of valence 1.*

Proof. We proceed by induction on $\frac{1}{2}\# T_E \geq 1$. If $T_E = \{e, \overline{e}\}$ it follows that $T_V = \{t(e), o(e)\}$ and the claim is true. For the inductive step, choose $e \in T_E$ and consider the subgraph T^e. It has precisely two components T^+, T^-. The claim follows by observing that $T^\pm$ both contain at least one vertex having valence 1 in T:

- If $T^\pm_E \neq \varnothing$, it inductively contains at least two vertices with valence 1 in $T^\pm$. At most one of them can be origin or terminus of e. The other must still have valence 1 in T.
- If $T^\pm_E = \varnothing$, it contains a single vertex which is the origin or terminus of e. This vertex thereby has valence 1 in T. ∎

To set us on our path back to group theory, we introduce actions of a group on a graph:

Definition. An *action of a group G on a graph* Γ consists of G-actions on Γ_V and Γ_E such that $o, t\colon \Gamma_E \to \Gamma_V$ and the inversion $\Gamma_E \to \Gamma_E$ are G-equivariant.
For $v \in \Gamma_V$ (resp. $e \in \Gamma_E$) we consider the *vertex stabilizer* (resp. *edge stabilizer*)

$$G_v := \{g \in G \mid gv = v\} \qquad \text{resp.} \qquad G_e := \{g \in G \mid ge = e\}$$

The action is *without edge inversions* if $ge \neq \overline{e}$ for all $e \in \Gamma_E$, $g \in G$.
If the action is without edge inversions, the *quotient graph* Γ/G has

$$(\Gamma/G)_V := (\Gamma_V)/G \qquad \text{and} \qquad (\Gamma/G)_E := (\Gamma_E)/G$$

with the induced origin, terminus and inversion maps.
The action is *cofinite* if the quotient graph Γ/G is finite.

Lemma A.4. *Let G be a group acting on a graph Γ without edge inversions.*
(1) *The quotient graph Γ/G is well-defined.*
(2) *We have $G_{\overline{e}} = G_e \subseteq G_{o(e)} \cap G_{t(e)}$ all $e \in \Gamma_E$.*
(3) *We have $G_e = \{g \in G \mid g\{e, \overline{e}\} = \{e, \overline{e}\}\}$ for all $e \in \Gamma_E$.*

Proof.
(1) The only non-trivial aspect is to show that $\overline{Ge} \neq Ge$ for all $Ge \in (\Gamma/G)_E$ which follows since G acts without edge inversions.
(2) This follows since the origin, terminus and inversion maps are G-equivariant.
(3) The inclusion '$\subseteq$' follows from (2). The other inclusion follows since the action is without edge inversions. ∎

We note some basic properties of group actions on trees for later reference:

Lemma A.5. *Let G be a group acting on a tree T without edge inversions.*
(1) *Let $x, y \in T_V$. If $e \in T_E$ lies on the unique geodesic from x to y, then $G_x \cap G_y \subseteq G_e$.*
(2) *If T is finite and non-empty, there exists $v \in T_V$ with $G_v = G$.*
(3) *Let $g \in G$. Suppose there exist $m \geq 1$ and $x \in T_V$ such that $g^m \in G_x$. Then $g \in G_y$ for some $y \in T_V$.*

Proof.
(1) Let $\gamma\colon I^n \to T$ be the unique geodesic from x to y. Then $g \cdot \gamma\colon I^n \to T$ is a geodesic from $g \cdot x = x$ to $g \cdot y = y$, i.e. $g \cdot \gamma = \gamma$ since T is a tree. Hence, $g \in G_e$ for all $e \in \operatorname{im}(\gamma_E)$.
(2) We proceed inductively on $\frac{1}{2}\#T_E$. If $\frac{1}{2}\#T_E = 0$, we have $T_V = \{v\}$ and the claim is clear. For the induction step assume $\frac{1}{2}\#T_E \geq 1$ and consider $L \coloneqq \{v \in T_V \mid \nu(v) = 1\}$. Then L is non-empty by Lemma A.3.
If $T_V = L$, T consists of a single edge connecting two vertices and the claim follows since G acts without edge inversions. Hence, we may assume $T_V \neq L$.
Note that $G \cdot L = L$. Hence, the action of G restricts to an action on the subgraph with vertices $T_V \setminus L$ and edges $T_E \setminus \{e, \overline{e} \mid e \in o^{-1}(L)\}$ and the claim follows by induction.
(3) The finite set $S \coloneqq \{g^i \cdot x \mid 0 \leq i \leq m-1\}$ is permuted by g. By Proposition A.2 there exists a unique minimal subgraph T^S of T with $S \subseteq T^S_V$ and T^S is finite. Since gT^S is a subgraph of T with $S = g \cdot S \subseteq gT^S_V$, T^S is a subgraph of gT^S. By finiteness $T^S = gT^S$ and $\langle g \rangle \subseteq G$ acts on T^S by restricting the G-action. Then (2) implies the claim. ■

As topologists we of course want consider graphs topologically. There is an obvious way to realize a graph as a topological space:

Definition. Let Γ be a graph. The *realization* of Γ is the 1-dimensional CW-complex $|\Gamma|$ with $|\Gamma|^{(0)} \coloneqq \Gamma_V$ and
$$|\Gamma|^{(1)} \coloneqq (\Gamma_V \sqcup \Gamma_E \times [-1,1]) /{\sim}$$
where for $e \in \Gamma_E$, $h \in [-1,1]$
$$(e,1) \sim t(e) \quad (e,-1) \sim o(e) \quad (e,t) \sim (\overline{e}, -t)$$

Perhaps one should now draw an example of the realization of a graph. But whenever we have drawn a graph, we have in fact drawn its realization.
Next we observe that a graph being a tree is a topological condition:

Proposition A.6. *Let Γ be a finite connected graph. Then Γ is a tree if and only if $|\Gamma|$ is simply connected.*

Proof. Suppose Γ is not a tree. By Proposition A.1 (2) there exists $e \in \Gamma_E$ with Γ^e connected. Then $\pi_1(|\Gamma|)$ is an HNN-extension of $\pi_1(|\Gamma^e|)$ (see [Fri25: Theorem 125.27]) and therefore non-trivial (see [Fri25: Lemma 93.5]). This shows that Γ is a tree if $|\Gamma|$ is simply connected.
For the reverse implication we proceed by induction on $\frac{1}{2}\#\Gamma_E \geq 0$: If $\#\Gamma_E = 0$, connectedness of Γ implies that $\Gamma_V = \{v\}$. Hence, Γ is a tree. For the inductive step assume $\frac{1}{2}\#\Gamma_E \geq 1$ and let $e \in \Gamma_E$. The subgraph $\Gamma^e \subseteq \Gamma$ consists of two components $\Gamma^{\pm}$ by Proposition A.1 (2). Both are trees with less edges than Γ. Hence $|\Gamma^{\pm}|$ are simply connected by induction. Then $|\Gamma|$ also is simply connected by the Seifert–van Kampen Theorem. ■

A.2. Graphs of groups

In the last proposition we have seen that the topology of graphs naturally leads to certain HNN-extensions and free products. In this section we develop this further to the notion of a graph of groups which generalizes and unifies HNN-extensions and amalgamated free products:

Definition. A *graph of groups* $\mathcal{G}$ consists of
- a finite connected graph $\mathcal{G}_\Gamma \eqqcolon \Gamma$, the *underlying graph*
- a group $\mathcal{G}_v$ for every $v \in \Gamma_V$, the *vertex group at v*
- a group $\mathcal{G}_e$ for every $e \in \Gamma_E$, the *edge group at e*
- a monomorphism $\delta^t_e\colon \mathcal{G}_e \hookrightarrow \mathcal{G}_{t(e)}$ for every $e \in \Gamma_E$

such that $\mathcal{G}_e = \mathcal{G}_{\overline{e}}$ for all $e \in \Gamma_E$. We set
$$\delta^o_e \coloneqq \delta^t_{\overline{e}}\colon \mathcal{G}_e = \mathcal{G}_{\overline{e}} \hookrightarrow \mathcal{G}_{t(\overline{e})} = \mathcal{G}_{o(e)}$$

We again want to realize this as a topological space. To do so, we introduce graphs of spaces as an intermediate step:

Definition. A *graph of spaces* $\mathcal{X}$ consists of

- a finite connected graph $\mathcal{X}_\Gamma =: \Gamma$, the *underlying graph*
- a pointed CW-complex $(\mathcal{X}_v, p_v)$ for every $v \in \Gamma_V$, the *vertex space at* v
- a pointed CW-complex $(\mathcal{X}_e, p_e)$ for every $e \in \Gamma_E$, the *edge space at* e
- a map $\delta_e^t\colon (\mathcal{X}_e, p_e) \to (\mathcal{X}_{t(e)}, p_{t(e)})$ inducing a monomorphism on fundamental groups for every $e \in \Gamma_E$

such that $(\mathcal{X}_e, p_e) = (\mathcal{X}_{\overline{e}}, p_{\overline{e}})$ for all $e \in \Gamma_E$. We set

$$\delta_e^o := \delta_{\overline{e}}^t\colon (\mathcal{X}_e, p_e) = (X_{\overline{e}}, p_{\overline{e}}) \to (\mathcal{X}_{t(\overline{e})}, p_{t(\overline{e})}) = (\mathcal{X}_{o(e)}, p_{o(e)})$$

The *graph of groups associated to* $\mathcal{X}$ has underlying graph Γ, vertex groups $\pi_1(\mathcal{X}_v, p_v)$ for $v \in \Gamma_V$, edge groups $\pi_1(\mathcal{X}_e, p_e)$ for $e \in \Gamma_E$ and the monomorphisms $(\delta_e^t)_*\colon \pi_1(\mathcal{X}_e, p_e) \hookrightarrow \pi_1(\mathcal{X}_{t(e)}, p_{t(e)})$.
The *realization of* $\mathcal{X}$ is the CW-complex

$$|\mathcal{X}| := \Big(\bigsqcup_{v \in \Gamma_V} \mathcal{X}_v \sqcup \bigsqcup_{e \in \Gamma_E} \mathcal{X}_e \times [-1, 1] \Big) / \sim$$

where for $e \in \Gamma_E$, $x \in \mathcal{X}_e$, $t \in [-1, 1]$

$$(x, 1) \sim \delta_e^t(x) \qquad (x, -1) \sim \delta_e^o(x) \qquad (x, t) \sim (\overline{x}, -t)$$

The *realization of a path* $\gamma\colon I_n \to \Gamma$ is the path

$$\begin{aligned} |\gamma|\colon [0, n] &\to |\mathcal{X}| \\ t &\mapsto \Big\{ (p_{\gamma_E(i)}, 2(t-i)-1), \quad \text{if } t \in [i-1, i] \text{ where } i \in \{1, \ldots, n\} \end{aligned}$$

We view $\mathcal{X}_v$ as a subspace of $|\mathcal{X}|$ is the obvious way for $v \in \Gamma_V$ and $\mathcal{X}_e$ as the subspace $\mathcal{X}_e \times \{0\}$ for $e \in \Gamma_E$.

Example A.7.

(1) Let $\mathcal{X}$ be a graph of spaces where every vertex and edge space is just a single point. Then the realization of $\mathcal{X}$ is just the realization of its underlying graph.

(2) Let $g \geq 0$ and consider the following graph of spaces $\mathcal{X}$:
   - The underlying graph Γ has $\Gamma_V = \{v\}$ and $\Gamma_E = \{1, \overline{1}, \ldots, g, \overline{g}\}$.
   - The vertex space $\mathcal{X}_v$ is an oriented S^2 with the interiors of $2g$ open discs removed.
   - The edge spaces are all oriented S^1's.
   - For $e \in \{1, \ldots, g\}$ the inclusion $\mathcal{X}_e \to \mathcal{X}_v$ is an orientation-preserving diffeomorphism to a boundary component and the inclusion $\mathcal{X}_{\overline{e}} \to \mathcal{X}_v$ is an orientation-reversion diffeomorphism to a boundary component such that every boundary component of $\mathcal{X}_v$ is identified with precisely one edge space.

   Then $|\mathcal{X}|$ is the orientable surface of genus g.

Using the realization we can associate a group to a graph of spaces:

Definition. Let $\mathcal{X}$ be a graph of spaces on a graph Γ. Let $v_0 \in \Gamma_V$. The *fundamental group of* $\mathcal{X}$ *at* v_0 is

$$\pi_1(\mathcal{X}, v_0) := \pi_1(|\mathcal{X}|, p_{v_0})$$

Lemma A.8.

(1) *Let $\mathcal{X}$ be a graph of spaces on a graph Γ and $\gamma\colon I^n \to \Gamma$ a path. Then*

$$\begin{aligned} \pi_1(\mathcal{X}, \gamma_V(0)) = \pi_1(|\mathcal{X}|, p_{\gamma_V(0)}) &\to \pi_1(|\mathcal{X}|, p_{\gamma_V(n)}) = \pi_1(\mathcal{X}, \gamma_V(n)) \\ \alpha &\mapsto |\gamma| * \alpha * |\overline{\gamma}| \end{aligned}$$

is an isomorphism.

(2) *Let $\mathcal{G}$ be a graph of groups on a graph Γ and $v_0 \in \Gamma_V$. If $\mathcal{X}$ and $\mathcal{X}'$ are graphs of spaces associated to $\mathcal{G}$, $\pi_1(\mathcal{X}, v_0) \cong \pi_1(\mathcal{X}', v_0)$.*

Proof.

(1) This follows from the analogous statement in topology, see [Fri25: Proposition 76.14].

(2) We proceed inductively on $\#\frac{1}{2}\Gamma_E \geq 0$:
If $\#\frac{1}{2}\Gamma_E = 0$, connectedness of Γ implies that $\Gamma_V = \{v_0\}$. Then

$$\pi_1(\mathcal{X}, v_0) \cong G_{v_0} \cong \pi_1(\mathcal{X}', v_0)$$

For the inductive step, assume $\#\frac{1}{2}\Gamma_E \geq 1$ and let $e \in \Gamma_E$, We consider two cases:

case 1: Γ^e is disconnected
Then Γ^e has precisely two path components $\Gamma^\pm$ which both have less edges than Γ. Wlog. $t(e) \in \Gamma_V^+$ and $o(e) \in \Gamma_V^-$. Restricting the graph of group (resp. spaces) structure defines graph of groups (resp. spaces) on $\Gamma^\pm$. Then

$$\begin{aligned}\pi_1(\mathcal{X}, v_0) = \pi_1(|\mathcal{X}|, p_{v_0}) &\overset{\substack{\text{topological change of basepoint}\\ \downarrow}}{\cong} \pi_1(|\mathcal{X}|, p_e) \overset{\substack{\text{Sefiert-van Kampen Theorem}\\ \downarrow}}{\cong} \pi_1(|\mathcal{X}^-|, p_{o(e)}) *_{\pi_1(\mathcal{X}_e, p_e)} \pi_1(|\mathcal{X}^+|, p_{t(e)})\\ &\underset{\substack{\uparrow\\ \text{induction}}}{\cong} \pi_1(|\mathcal{X}'^-|, p_{i(e)}) *_{\pi_1(\mathcal{X}'_e, p_e)} \pi_1(|\mathcal{X}'^+|, p_{t(e)}) \cong \pi_1(|\mathcal{X}'|, v_0)\end{aligned}$$

case 2: Γ^e is connected
In this case we can proceed similarly by applying the HNN-Seifert–van Kampen Theorem (see [Fri25: Theorem 125.27]). ■

In the last lemma we saw that the fundamental group of a graph of spaces only depends on the associated graph of groups. If every graph of groups is associated to a graph of spaces, we can thereby define a fundamental group of a graph of groups:

Construction A.9. Let $\mathcal{G}$ be a graph of groups on a graph Γ. For $v \in \Gamma_V$ (resp. $e \in \Gamma_E$) let $(\mathcal{X}_v, p_v)$ (resp. $(\mathcal{X}_e, p_e)$) be a the canonical pointed Eilenberg-Maclane space of type $K(\mathcal{G}_v, 1)$ (resp. $K(\mathcal{G}_e, 1)$) as defined in [Fri25: Definition 281.11]. Identify $\pi_1(\mathcal{X}_v, p_v) \cong \mathcal{G}_v$ (resp. $\pi_1(\mathcal{X}_e, p_e) \cong \mathcal{G}_e$) along the canonical isomorphism defined there and identify these groups in the following.
Since maps between such Eilenberg-Maclane Spaces are classified by the fundamental group (see [Fri25: Proposition 281.18]), there exists a natural map $(\mathcal{X}_e, p_e) \to (\mathcal{X}_{t(e)}, p_{t(e)})$ inducing δ_e^t on fundamental groups. These define a graph of spaces associated to Γ.

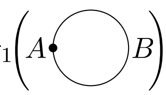
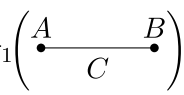
We now define the fundamental group of a graph of spaces as the fundamental group of the precise associated graph of spaces constructed above. Up to isomorphism we can of course use any associated graph of spaces to compute it:

Definition. Let $\mathcal{G}$ be a graph of groups on a graph Γ. Let $v_0 \in \Gamma_V$. Let $\mathcal{X}$ be the canonical graph of spaces with associated graph of groups $\mathcal{G}$ constructed in Construction A.9. The *fundamental group of $\mathcal{G}$ at v_0* is $\pi_1(\mathcal{G}, v_0) := \pi_1(\mathcal{X}, v_0)$.

With the fundamental group defined, we can see how graphs of groups generalize HNN-extensions and amalgamated free products:

Example A.10.

(1) It follows from the two versions of the Seifert–van Kampen Theorem that

π_1  $\cong A *_B$ and π_1  $\cong A *_C B$

(2) By inductively applying the Seifert–van Kampen Theorem it follows that the fundamental group of a graph of groups on a tree T is the colimit of the diagram defined by the vertex and edge groups with the given monomorphisms.

(3) The fundamental group of a graph of groups with trivial vertex and edge groups is the fundamental group of the underlying graph.

The vertex and edge groups of a graph of groups can be viewed as subgroups of its fundamental group:

Lemma A.11. *Let $\mathcal{G}$ be a graph of groups on a graph Γ. Let $\gamma\colon I^n \to \Gamma$ be a path. The maps*

$$\begin{array}{rcl} \mathcal{G}_{\gamma_V(n)} & \hookrightarrow & \pi_1(\mathcal{G}, \gamma_V(0)) \\ \alpha & \mapsto & \overline{\gamma} * \alpha * \gamma \end{array} \qquad \textit{and} \qquad \begin{array}{rcl} \mathcal{G}_{\gamma_E(n)} & \hookrightarrow & \pi_1(\mathcal{G}, \gamma_V(0)) \\ \alpha & \mapsto & \overline{\gamma} * \delta^t_{\gamma_E(n)}(\alpha) * \gamma \end{array}$$

are monomorphisms of groups where we identify $\mathcal{G}_{\gamma_V(n)}$ and $\mathcal{G}_{\gamma_E(n)}$ with the fundamental group of the corresponding Eilenberg-Maclane space of type $K(-,1)$.

Proof. The given maps are well-defined as the Eilenberg-Maclane spaces are subspaces of the canonical realization of $\mathcal{G}$ which is used to define the fundamental group. By Lemma A.8 (1) we only need to prove the claim for $n = 0$. Here we can use the argument of Lemma A.8 (2) to reduce to the case of HNN-extensions or amalgamated free products where the claim is classical (see [Fri25: Propositions 85.25, 93.6]). For the second map observe furthermore that $\delta^t_{\gamma_E(n)}\colon \mathcal{G}_{\gamma_E(n)} \hookrightarrow \mathcal{G}_{\gamma_V(n)}$ is a monomorphism by assumption. ■

Of course, these inclusions from the vertex and edge groups into the fundamental group depend on the chosen path – at least up to conjugation. We can choose sufficiently many paths at once using a spanning tree. This allows us to identify the vertex and edge groups with subgroups of the fundamental group:

Definition. Let $\mathcal{G}$ be a graph of groups on a graph Γ and $v_0 \in \Gamma_V$. Given a spanning tree T of Γ we make the following identifications:

- For $v \in \Gamma_V$ there exists a unique geodesic $\gamma\colon I^n \to T \subseteq \Gamma$ from v_0 to v. Using Lemma A.11 we identify G_v with a subgroup of $\pi_1(\mathcal{G}, v_0)$ along γ.
- For $e \in \Gamma_E$ there exists a unique geodesic $\gamma\colon I^n \to T \subseteq \Gamma$ from v_0 to $t(e)$. Using Lemma A.11 we identify G_e with a subgroup of $\pi_1(\mathcal{G}, v_0)$ along γ.

A.3. The Bass–Serre tree

Using the language of graphs of groups our goal can now be reformulated (and generalized) as follows: Given a group G find a bound on the number of vertices in a graph of groups with fundamental group G. But we have made no apparent progress towards actually achieving this. In this section we will discuss the main tool we will use: We turn a graph of groups with fundamental group G into an action by G on a tree T. Actions of groups on trees can then be studied geometrically.

Definition. Let $\mathcal{G}$ be a graph of groups on a graph Γ. Let $v_0 \in \Gamma_V$ and set $G := \pi_1(\mathcal{G}, v_0)$. A *Bass–Serre tree* of $\mathcal{G}$ consists of

- a tree $\tilde{\mathcal{G}}$ with a G-action
- a morphism of graphs $p\colon \tilde{\mathcal{G}} \to \Gamma$ inducing an isomorphism $\tilde{\mathcal{G}}/G \to \Gamma$

such that for every spanning tree $T \subseteq \Gamma$ used to identify the vertex and edge groups of $\mathcal{G}$ with subgroups of G there exist *lifting maps*

$$\begin{array}{rcl} \Gamma_V & \to & \tilde{\mathcal{G}}_V \\ v & \mapsto & \tilde{v} \end{array} \qquad \text{and} \qquad \begin{array}{rcl} \Gamma_E & \to & \tilde{\mathcal{G}}_E \\ e & \mapsto & \tilde{e} \end{array}$$

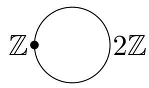
with

- $p_V(\tilde{v}) = v$ and $G_{\tilde{v}} = \mathcal{G}_v$ for all $v \in \Gamma_V$
- $p_E(\tilde{e}) = e$ and $G_{\tilde{e}} = \mathcal{G}_e$ for all $e \in \Gamma_E$
- $t(\tilde{e}) = \widetilde{t(e)}$ for all $e \in \Gamma_E$

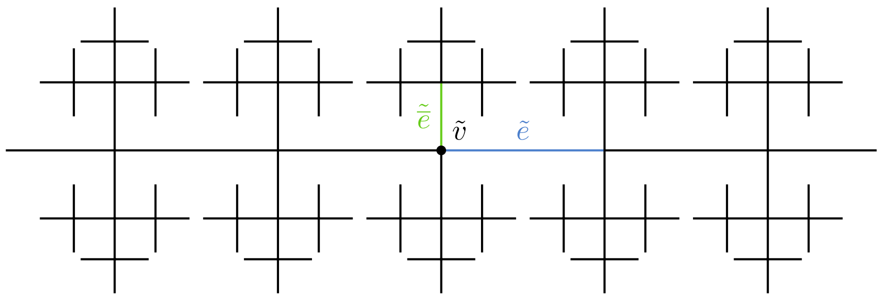
Example A.12. Let $\mathcal{G}$ be a graph of groups on a graph Γ.

- If all vertex and edge groups trivial, a Bass–Serre tree of $\mathcal{G}$ is just the universal covering of Γ with the action by $\pi_1(\mathcal{G}) \cong \pi_1(\Gamma)$ via deck transformations.

- For a more interesting example consider the following graph of groups:



Its fundamental group is $\langle a, t \mid ta^2t^{-1} = a^2 \rangle$. A Bass–Serre tree is the infinite tree where every vertex has valence 4



with a acting by 'mirroring along the diagonal with slope -1' through $\tilde{v}$ and t acting by a shift by 1 to the right.

In the next proposition we give a general description of a Bass–Serre tree. With it, one can in principle find the Bass–Serre tree of any given graph of groups:

Proposition A.13. *Let $\mathcal{G}$ be a graph of groups on a graph Γ. Let $v_0 \in \Gamma_V$ and set $G := \pi_1(\mathcal{G}, v_0)$. Let $p\colon \tilde{\mathcal{G}} \to \Gamma$ be a Bass–Serre tree. Fix a spanning tree $T \subseteq \Gamma$. Use T to identify the vertex and edge groups of $\mathcal{G}$ with subgroups of G, and consider the lifting maps corresponding to T.*

(1) *Let $e \in \Gamma_E$. For all $\hat{e} \in p_E^{-1}(e)$ there exist a $g \in G$ with $\hat{e} = g\tilde{e}$. Then $t(\hat{e}) = g\widetilde{t(e)}$ and we have a commutative diagram*

$$\begin{tikzcd} G_{\hat{e}} \arrow[r, hook] \arrow[d, "\cong"'] & G_{t(\hat{e})} \arrow[d, "\cong"] \\ \mathcal{G}_e \arrow[r, hook] & \mathcal{G}_{t(e)} \end{tikzcd}$$

with the vertical isomorphisms given by conjugation with g^{-1}.

(2) (a) *For $v \in \Gamma_V$ and $e \in \Gamma_E$ the maps*

$$\begin{array}{rcl} G/\mathcal{G}_v & \to & p_V^{-1}(v) \\ g\mathcal{G}_v & \mapsto & g\tilde{v} \end{array} \qquad \text{and} \qquad \begin{array}{rcl} G/\mathcal{G}_e & \to & p_E^{-1}(e) \\ g\mathcal{G}_e & \mapsto & g\tilde{e} \end{array}$$

are a G-equivariant bijections with G acting by left-translation on the domains.

(b) *For $e \in \Gamma_E$ and $\hat{e} \in p_E^{-1}(e)$ the map*

$$\begin{array}{rcl} \mathcal{G}_{t(e)}/\mathcal{G}_e & \to & \{e' \in E(\tilde{\mathcal{G}}) \mid t(\hat{e}) = t(e')\} \\ g\mathcal{G}_e & \mapsto & ge' \end{array}$$

is a $\mathcal{G}_{t(e)}$-equivariant bijection with $\mathcal{G}_{t(e)}$ acting by left-translation on $\mathcal{G}_{t(e)}/\mathcal{G}_e$.

Proof.

(1) As the morphism $p\colon \tilde{\mathcal{G}} \to \Gamma$ induces an isomorphism $\tilde{\mathcal{G}}/G \to \Gamma$, there exists $g \in G$ with $\overline{e} = g\tilde{e}$. Then

$$t(\overline{e}) = t(g\tilde{e}) = gt(\tilde{e}) = g\widetilde{t(e)}$$

For commutativity of the diagram observe that $\overline{e} = g\tilde{e}$ implies purely by group theory that on stabilizers $G_{\overline{e}} = gG_{\tilde{e}}g^{-1}$, and similarly for $t(\overline{e})$.

(2) All of the claims are essentially instances of the Orbit-Stabilizer Theorem. ■

From this proposition one can relatively easily conclude that there is at most one Bass–Serre tree for a given graph of groups. Much less trivial is the existence of the Bass–Serre tree.
We again prove this topologically: The universal covering of a graph of spaces can again be decomposed as a graph of spaces. The underlying graph of this decomposition is the desired tree:

Theorem A.14. *Let $\mathcal{G}$ be a graph of groups on a graph Γ. Let $v_0 \in \Gamma_V$ and set $G := \pi_1(\mathcal{G}, v_0)$. A Bass–Serre tree $p\colon \tilde{\mathcal{G}} \to \Gamma$ for $\mathcal{G}$ exists and if $p'\colon \tilde{\mathcal{G}}' \to \Gamma$ is also a Bass–Serre tree for $\mathcal{G}$, there exists an isomorphism $\tilde{\mathcal{G}} \to \tilde{\mathcal{G}}'$ of graphs with G-actions such that*

$$\begin{tikzcd} \tilde{\mathcal{G}} \arrow[rr, "\cong"] \arrow[dr, "p"'] & & \tilde{\mathcal{G}}' \arrow[dl, "p'"] \\ & \Gamma & \end{tikzcd}$$

commutes.

Proof. By Construction A.9 there exists a graph of spaces $\mathcal{X}$ associated to $\mathcal{G}$. Let $p\colon \widetilde{|\mathcal{X}|} \to |\mathcal{X}|$ be the universal covering and consider the following CW-subcomplexes of $|\mathcal{X}|$

$$\mathfrak{V} := \bigcup_{v\in V} \mathcal{X}_v \qquad \mathfrak{E} := \bigcup_{e\in E} \mathcal{X}_e$$

Define $\tilde{\mathcal{G}}_V := \pi_0(p^{-1}(\mathfrak{V}))$. These will be the vertices of $\tilde{\mathcal{G}}$. We similarly would want to define the edges as $\pi_0(p^{-1}(\mathfrak{E}))$, but in our definition of a graph we need every edge twice – once for each choice of 'orientation'. To make this choice, we need to proceed carefully: First note that if $c \in \pi_0(p^{-1}(\mathfrak{E}))$ is a path component, $p(c) = \mathcal{X}_e = \mathcal{X}_{\overline{e}}$ for some $e \in \Gamma_E$. Hence, it is natural to define $\tilde{\mathcal{G}}_E$ as the set of pairs (c, f) where $c \in \pi_0(p^{-1}(\mathfrak{E}))$ and $f\colon c \times [-1,1] \to \widetilde{|\mathcal{X}|}$ is a continuous map for which there exists $e \in \Gamma_E$ such that the diagram

$$(*) \qquad \begin{tikzcd} c \arrow[r, "x\mapsto(x,0)"'] \arrow[rr, bend left, "\text{incl.}"] & c \times [-1,1] \arrow[r, "f"'] \arrow[d, "p|_c\times\mathrm{id}"'] & \widetilde{|\mathcal{X}|} \arrow[d, "p"] \\ & \mathcal{X}_e \times [-1,1] \arrow[r, "\text{incl.}"] & \mathcal{X} \end{tikzcd}$$

commutes. Observe the following:

- Given $c \in \pi_0(p^{-1}(\mathfrak{E}))$ and $e \in \Gamma_E$ with $p(c) = \mathcal{X}_e$ there exists a unique continuous map $f\colon c \times [-1,1] \to \widetilde{|\mathcal{X}|}$ such that the diagram commutes:
  By covering space theory this is true if and only if the composition $c \times [-1,1] \to X$ induces the trivial map on fundamental groups – but this follows since the curved map exists and the inclusion $c \to c \times [-1,1]$ is a homotopy equivalence.
- Given $c \in \pi_0(p^{-1}(\mathfrak{E}))$, there exists precisely two $e \in \Gamma_E$ with $p(c) = \mathcal{X}_e$ and they are inversions of each other. Hence, the $e \in \Gamma_E$ such that the diagram commutes is uniquely determined by (c, f). It therefore makes sense to define $p_E(c, f) := e$ giving a map $p_E\colon \tilde{\mathcal{G}}_E \to \Gamma_E$.
- The last point also implies that we have doubled up the elements $\pi_0(p^{-1}(\mathfrak{E}))$, i.e. for every $c \in \pi_0(p^{-1}(\mathfrak{E}))$ there exist precisely two maps $f, f'\colon c\times[-1,1] \to \widetilde{|\mathcal{X}|}$ with $(c,f), (c,f') \in \tilde{\mathcal{G}}_E$.

On vertices we define a map $p_V\colon \tilde{\mathcal{G}}_V \to \Gamma_V$ by letting $p_V(c)$ be the unique $v \in \Gamma_V$ with $p(c) = \mathcal{X}_v$.
To define the graph $\tilde{\mathcal{G}}$ it remains to give its origin, terminus and inversion maps. We let the terminus of an edge $(c, f) \in \tilde{\mathcal{G}}_E$ be the unique path component of $p^{-1}(\mathfrak{V})$ that contains $f(c\times\{1\})$. Analogously, the origin is the path component containing $f(c \times \{-1\})$. The inversion of (c, f) is (c, f') where

$$f'\colon c \times [-1,1] \to \widetilde{|\mathcal{X}|}$$
$$(x,t) \mapsto f(x,-t)$$

The above bullet points now imply that this yields a well-defined graph $\tilde{\mathcal{G}}$ and a morphism of graphs $p\colon \tilde{\mathcal{G}} \to \Gamma$.

This graph is a tree by Proposition A.1 (2) since for $c \in \pi_0(p^{-1}(\mathfrak{E}))$, $\widetilde{|\mathcal{X}|} \setminus c$ is disconnected: Otherwise, $\pi_1(\widetilde{|\mathcal{X}|})$ would be an HNN-extension over $\pi_1(c)$ (see [Fri25: Theorem 125.27]) and therefore non-trivial (see [Fri25: Lemma 93.5]).
Next we define the group action by G on $\tilde{\mathcal{G}}$:
Pick a lift $\tilde{p}_{v_0} \in p^{-1}(p_{v_0})$. The group $G = \pi_1(\mathcal{G}, v_0) = \pi_1(|\mathcal{X}|, p_{v_0})$ acts on $\widetilde{|\mathcal{X}|}$ by deck transformations, i.e. $g \in G$ acts as the unique deck transformation $f \colon \widetilde{|\mathcal{X}|} \to \widetilde{|\mathcal{X}|}$ where $f(\tilde{p}_{v_0})$ is the endpoint of the lift of the loop $g \in \pi_1(|\mathcal{X}|, p_{v_0})$ to $p_{\tilde{v}_0}$ (compare [Fri25: Proposition 114.9]). Since $p \colon \widetilde{|\mathcal{X}|} \to |\mathcal{X}|$ induces an isomorphism $\widetilde{|\mathcal{X}|}/G \to |\mathcal{X}|$, this action induces an action on $\tilde{\mathcal{G}}_V = \pi_0(p^{-1}(\mathfrak{V}))$ and $\pi_0(p^{-1}(\mathfrak{E}))$. We may hence define an action on $\tilde{\mathcal{G}}_E$ by $g(c, f) = (gc, gf)$. By definition of deck transformations commutativity of the diagram $(*)$ implies that the diagram

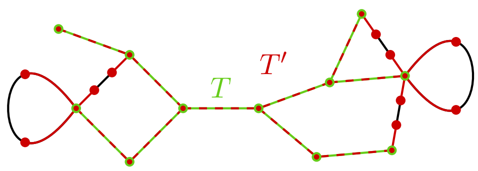
$$\begin{tikzcd}
gc \arrow[r, "x \mapsto (x,0)"'] \arrow[rr, bend left, "\text{incl.}"] & gc \times [-1,1] \arrow[r, "f"'] \arrow[d, "p|_{gc} \times \mathrm{id}"'] & \widetilde{|\mathcal{X}|} \arrow[d, "p"] \\
& \mathcal{X}_e \times [-1,1] \arrow[r, "\text{incl.}"] & |\mathcal{X}|
\end{tikzcd}$$

also commutes. Hence, the G-actions give a G-action without edge inversions on $\tilde{\mathcal{G}}$ and $p \colon \tilde{\mathcal{G}} \to \Gamma$ induces an isomorphism $\tilde{\mathcal{G}}/G \to \Gamma$.
The last objects we need to construct are the lifting maps. For these we need a spanning tree $T \subseteq \Gamma$. To allow us to find lifts for the edges missing in T, we need to extend T:
Consider the graph T' with $T'_V = T_V \sqcup (\Gamma_E \setminus T_E)$,

$$T'_E \coloneqq T_E \sqcup (\{\pm 1\} \times (\Gamma_E \setminus T_E))$$

where the origin, terminus and inversion maps for $e \in T_E \subseteq T'_E$ are the ones from T' and for $(s, e) \in \Gamma_E \setminus T_E \subseteq T'_E$

$$\overline{(s,e)} \coloneqq (-s, e) \qquad \text{and} \qquad o(\overline{s,e}) \coloneqq t(s,e) \coloneqq \begin{cases} t(e) \in T_V \subseteq T'_V, & \text{if } s = +1 \\ e \in \Gamma_E \setminus T_E \subseteq T'_V, & \text{if } s = -1 \end{cases}$$



Note that T' is still a tree: If one attaches a new vertex to a tree along a single edge to a vertex in the tree, the result is again a tree.
Consider the map

$$\begin{aligned} f \colon |T'| &\to |T| \\ P &\mapsto \begin{cases} v \in |T|^{(0)}, & \text{if } P = v \in T_V \subseteq T'_V = |T'|^{(0)} \\ (e, 0) \in e \times [-1,1], & \text{if } P = e \in \Gamma_E \setminus T_E \subseteq T'_V = |T'|^{(0)} \\ (e, t) \in e \times [-1,1], & \text{if } P = (e,t) \in e \times [-1,1] \subseteq |T'| \text{ for } e \in T_E \\ (e, \tfrac{1}{2}(st+1)) \in e \times [-1,1], & \text{if } \begin{array}{l} P = ((s,e),t) \in (s,e) \times [-1,1] \subseteq |T'| \\ \quad \text{for } e \in \Gamma_E \setminus T_E, s \in \{\pm 1\} \end{array} \end{cases} \end{aligned}$$

This map is surjective but every edge is 'cut open' in the middle of the edges not belonging to T. Composing with the obvious inclusion

$$\begin{aligned} |T| &\to |\mathcal{X}| \\ P &\mapsto \begin{cases} p_v, & \text{if } P = v \in T_V = |T|^{(0)} \\ (p_e, t), & \text{if } P = (e,t) \in T_E \times [-1,1] \subseteq |T| \end{cases} \end{aligned}$$

gives a continuous map $f \colon T' \to |\mathcal{X}|$.

Since T' is a tree, $|T'|$ is simply connected by Proposition A.6. Hence, there exists a unique lift $\tilde{f}\colon |T'| \to \tilde{X}$ such that

$$\begin{tikzcd} & (\widetilde{|\mathcal{X}|}, \tilde{p}_{v_0}) \arrow[d, "p"] \\ (|T'|, v_0) \arrow[ur, "\tilde{f}"] \arrow[r, "f"] & (|\mathcal{X}|, p_{v_0}) \end{tikzcd}$$

commutes. For $v \in \Gamma_V$ there exists a unique point in $p^{-1}(p_v) \cap \tilde{f}(|T'|)$. Let $\tilde{v}$ be its path component in $\tilde{\mathcal{G}}_V = \pi_0(p^{-1}(\mathfrak{V}))$. Proceed similarly for $e \in E_T \subseteq \Gamma_E$.
For $e \in \Gamma_E \setminus T_E$, there exist two points in $p^{-1}(p_e) \cap \tilde{f}(|T'|)$ – precisely one of which is connected by an edge of T' to $\widetilde{t(e)}$. Let $\hat{e} \in \pi_0(p^{-1}(\mathfrak{E}))$ be the path component of this point. There is a unique $f\colon e \times [-1, 1] \to \widetilde{|\mathcal{X}|}$ such that $p_E(\hat{e}, f) = e$. Set $\tilde{e} := (\hat{e}, f)$. Note that this implies that e and $\overline{e}$ have different lifts. A generous reading of [Fri25: Proposition 115.7] now implies that the stabilizers are as desired.
For the uniqueness statement, pick a single spanning tree $T \subseteq \Gamma$ and use it to define the lifting maps to $\tilde{\mathcal{G}}$ and $\tilde{\mathcal{G}}'$. The desired isomorphism can then by constructed from the bijections provided in Proposition A.13 (2). ∎

Having established how to turn a graph of groups into an action on a tree, we turn to the reverse direction – defining a graph of groups from an action on a tree:

Definition. Let G be a group acting cofinitely without edge inversions on a tree T. Consider the quotient graph $p\colon T \to T/G$ and let $S \subseteq T/G$ be a spanning tree. Choose $O \subseteq (T/G)_E$ such that $(T/G)_E = O \sqcup \overline{O}$[44]. Choose a morphism of graphs $l\colon S \to T$ such that $p \circ l = \mathrm{id}$. Define the lift of $v \in (T/G)_V$ to be $\tilde{v} := l_V(v)$. For $e \in E(S)$ similarly define $\tilde{e} := l_E(e)$. For $e \in ((T/G)_E \setminus S_E) \cap O$ there exists a unique $\tilde{e} \in T_E$ such that $t(\tilde{e}) = l_E(t(e))$. Set $\tilde{\overline{e}} := \overline{\tilde{e}}$. This defines lifts for all edges of T/G.
A *quotient graph of groups* of G acting on T is now obtained in the following way:
- The vertex and edge groups are the stabilisers of the respective lifts.
- For $e \in O \cup S_E$, the monomorphism $G_e \hookrightarrow G_{t(e)}$ is the inclusion. For $e \notin O \cup S_E$

  $$p(t(\tilde{e})) = t(e) = p(\widetilde{t(e)})$$

  Hence, we can choose $g \in G$ with $t(\tilde{e}) = g \cdot \widetilde{t(e)}$. The monomorphism $G_e \hookrightarrow G_{t(e)}$ is then given by

  $$\begin{aligned} G_e = G_{\tilde{e}} \subseteq G_{t(\tilde{e})} = gG_{\widetilde{t(e)}}g^{-1} &\to G_{\widetilde{t(e)}} = G_{t(e)} \\ h &\mapsto g^{-1}hg \end{aligned}$$

We will only need this construction in passing and hence do not establish its properties. In particular, we do not state a uniqueness result since this would require introducing the category of graphs of groups which we otherwise have no need for.
Hence, we only refer to the literature for showing that the Bass–Serre tree and quotient constructions are inverses in a suitable sense:

Theorem A.15.
(1) *Let G be a group acting cofinitely on a tree T without edge inversions. Let $\mathcal{G}$ be the quotient graph of groups.*
  (a) *Let $v_0 \in (G/T)_V$. Then $G \cong \pi_1(\mathcal{G}, v_0)$.*
  (b) *The Bass–Serre tree of the quotient graph $\mathcal{G}$ is given by the G-action on T.*
(2) *Let $\mathcal{G}$ be a graph of groups. Then $\mathcal{G}$ is a quotient graph of groups of the Bass–Serre tree $\tilde{\mathcal{G}}$.*

Proof. see [SW79: Theorem 4.3] ∎

[44]This O stands for *orientation*.

We have approached Bass–Serre theory from a topological point of view. Alternatively, one can achieve the same goals combinatorially as done in [Ser80]. This does have advantages, for example our way of defining the fundamental group of a graph of groups does via Eilenberg-Maclane spaces – which leads to huge intermediate steps. It is intuitive that the fundamental group can be defined directly from the graph of groups: Its elements should be loops in the graph decorated by suitably compatible elements of the vertex and edge groups. Of course, when one takes this approach one has to reprove many things manually that were obvious to us since we knew the topological analogues. For example, when defining the Bass–Serre tree the combinatorial approach can easily write down an explicit model – but it is somewhat complicated to show that it is a tree. On the contrary, we struggled to construct the Bass–Serre tree – but it was immediately obvious that it is a tree.

A.4. Accessibility

The goal we set out to achieve in this appendix was to bound the number of vertices in a graph of groups with a fixed fundamental group. In the last section we learned how to translate this into a geometric problem about groups acting on trees. In this section we will use the resolution of this geometric problem to find our desired bounds.
We first observe that one cannot naively have such a bound as the number of vertices can become arbitrarily large:

Example A.16.

- On any finite graph Γ we can define a graph of groups with trivial fundamental group by letting all vertex and edge groups be trivial. More generally, we can extend any graph of groups without changing the fundamental group by adding an edge from an existing to a new vertex and labelling this edge and vertex trivially.
- For any $n \geq 0$

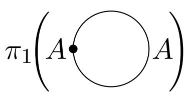
$$\begin{array}{ccccccccc} \mathbb{Z}x & & \mathbb{Z}x^2 & & \mathbb{Z}x^4 & & \mathbb{Z}x^{2^n} & & \mathbb{Z}y \\ \bullet & \text{---} & \bullet & \text{---} & \bullet & \cdots & \bullet & \text{---} & \bullet \\ & \mathbb{Z}x^2 & & \mathbb{Z}x^4 & & & & \{e\} & \end{array}$$

  is a graph of groups with $n+1$ vertices whose fundamental group is $\langle x, y\rangle$.

These examples arise when the monomorphism from an edge group to a vertex group is an isomorphism. Hence, we need to exclude this possibility:

Definition. Let $\mathcal{G}$ be a graph of groups on a graph Γ. A vertex $v \in \Gamma_V$ is *inessential* if there exists an edge $e \in \Gamma_E$ with $t(e) = v \neq o(e)$ such that the monomorphism $d_e^t \colon \mathcal{G}_e \hookrightarrow \mathcal{G}_v$ is an isomorphism. The graph of groups $\mathcal{G}$ is *reduced* if all vertices are essential.
Let T be a tree with a G-action without edge inversions. A vertex $v \in T_V$ is *inessential* if there exists an edge $e \in T_E$ with $t(e) = v$, $o(e) \notin Gv$ such that $G_e = G_v$. The G-action on T is *reduced* if all vertices are essential.

The condition that the edge is not a loop arises since an HNN-extension of a group over the entire group is not isomorphic to the original group.



$$\cong A*_A \not\cong A$$

Unsurprisingly, the two notions of reducedness correspond to each other:

Lemma A.17.

(1) *The group action on the Bass–Serre tree of a reduced graph of groups is reduced.*
(2) *The quotient graph of groups of a reduced group action on a tree is reduced.*

Proof.

(1) This follows from Proposition A.13 (1).
(2) This follows directly from the definition. ■

Given an inessential vertex we can collapse it along the relevant edge to remove it. Doing so for sufficiently many inessential vertices leads to a reduced action on a tree:

Proposition A.18. *Let T be a tree with a G-action without edge inversions.*[45] *There exists a tree T' with a reduced G-action and*

(i) *an injective G-equivariant map $f\colon T'_E \hookrightarrow T_E$ such that $G_{e'} = G_{f(e')}$ for $e' \in T'_E$*

(ii) *a surjective G-equivariant map $g\colon T_V \twoheadrightarrow T'_V$ such that for $v' \in T'_V$*

$$G_{v'} = \bigcup_{v \in g^{-1}(v')} G_v$$

(iii) *for all $e' \in T'_E$*

$$\overline{f(e')} = f(\overline{e}') \qquad (g \circ t \circ f)(e') = t(e') \qquad (g \circ o \circ f)(e') = o(e')$$

Proof. By collapsing a subset $I \subseteq T_E$ we obtain a tree T^I in the following sense: Let $T^I_V := T_V/{\sim}$ where $t(e) \sim o(e)$ for $e \in I$ and $T^I_E := T_E \setminus I$. Let $\iota_I\colon T^I_E = T_E \setminus I \hookrightarrow T_E$ be the inclusion and $\pi_I\colon T_V \twoheadrightarrow T^I_V = T_V/{\sim}$ the projection. The origin, terminus and inversion maps of T^I are induced by those of T, implying that the analogue of (iii) holds. If additionally $G \cdot I = I$, the G-action on T induces a G-action on T^I. The analogues to (i), (ii) then also hold as they follow purely from considerations about quotienting sets with group actions and restricting group actions to subsets. It remains to identify a suitable subset $I \subseteq T_E$ for which the action on T^I is reduced. Here we proceed in two steps: We certainly can collapse all edges whose origin and terminus are inessential as T is a tree: Set

$$A := \{e \in T_E \mid G_e = G_{o(e)} = G_{t(e)},\ o(e) \notin Gt(e)\}$$

and consider T^A. In the second step, choose for every orbit $Gv \subseteq T^A$ of inessential vertices in T^A, an orbit $Ge_v \subseteq T^A_E$ of edges such that $t(e_v) = v$, $o(e_v) \notin Gt(e_v)$ and $G_{e_v} = G_v$. Let $B \subseteq T^A_E \hookrightarrow T_E$ be the union of these orbits. Set $I := A \cup B$. Note that there is a canonical isomorphism $T^I \cong (T^A)^B$ and use it to identify these trees.
Suppose $v \in T^I_V$ is an inessential vertex, i.e. there exists $e \in T^I_E$ with $t(e) = v$, $o(e) \notin Gv$ and $G_e = G_v$. Then $\iota_B(e) \in T^A_E$ and $t(\iota_B(e)) \in \pi_B^{-1}(v)$, so

$$G_v = G_e = G_{\iota_B(e)} \subseteq G_{t(\iota_B(e))} \subseteq G_v$$

i.e. these inclusions are equalities and $v' := t(\iota_B(e)) \in T^A_V$ is an inessential vertex.
Consider the edge $e_{v'} \in B$. Then $o(e_{v'}) \in \pi_B^{-1}(v)$ and

$$G_{v'} = G_{e_{v'}} \subsetneq G_{o(e_{v'})} \subseteq G_v = G_{v'}$$

Contradiction! The G-action on $T' := T^I$ is therefore reduced. ∎

As we have already pointed out, a reduced action by a finitely generated group will always be cofinite:[46]

Lemma A.19. *A reduced action without edge inversions by a finitely generated group on a tree is cofinite.*

Proof. Let G be a group with a finite generating set S and consider a G-action without edge inversions on a tree T that is not cofinite. Wlog. $S = S^{-1}$ and S contains the neutral element. Pick $v_0 \in T_V$. Let $T^S \subseteq T$ be the unique minimal subtree of T with $Sv_0 \subseteq T^S_V$ from Proposition A.2. Note that T^S is finite, hence $G \cdot T^S \neq T$. Then there exists $e \in T_E$ such that $o(e) \in (G \cdot T_0)_V$, $v := t(e) \notin (G \cdot T_0)_V$.

[45] Note that here we for once do not assume that the action is cofinite. The reason for this is that the G-actions we obtain later need not be cofinite. However, as we previously always assumed (co)finiteness, we cannot apply Bass–Serre Theory unless this is the case. We wiggle our way out of this issue by using this proposition to assume all actions are reduced and then applying the next lemma to see that reduced actions are cofinite.

[46] If we had developed Bass–Serre theory on possibly infinite graphs, this lemma would of course also have a graph of groups versions – namely that a reduced graph of groups with finitely generated fundamental groups is finite. Hence, we see that we did not lose much by restricting to finite graphs.

Observe that for $s \in S$, $v_0 \in sT_V^S$ since $S = S^{-1}$. Hence, more generally gT^S contains v_0 for all $g \in G$ and $G \cdot T^S$ is connected. As T is a tree this implies that e is the unique edge in T with $o(e) \in (G \cdot T_0)_V$ and $t(e) = v$. For $g \in G_v$, ge also fulfils this, hence $ge = e$ and $G_v = G_e$. Then v is an inessential vertex. ∎

We can now use the trick established in the footnotes to prove statements about trees with not necessarily cofinite actions by proving them for graphs of groups. We apply this to the following technical lemma which we will need later:

Lemma A.20. *Let G be a finitely generated group and T be a tree with a G-action without edge inversions, such that all vertex stabilizers are abelian. If there exists $v \in T_V$ such that G is the normal closure of G_v, $G = G_v$.*

Proof. Let T', $f\colon T'_E \to T_E$ and $g\colon T_V \to T'_V$ be the reduction of T as in Proposition A.18. Then G is still the normal closure of $G_{g(v)}$. Infact, since all vertex groups are abelian, $G_{g(v)} = G_v$. By Lemma A.19, Theorem A.15 and Lemma A.17 it therefore suffices to prove the analogous claim for graphs of groups:

Claim. *Suppose $\mathcal{G}$ is a reduced graph of groups with fundamental group G on a graph Γ such that all vertex groups are abelian. If there exists $v \in \Gamma_V$ such that G is the normal closure of $\mathcal{G}_v$, $G = \mathcal{G}_v$*[47]*.*

Suppose Γ is not a tree. By Proposition A.1 (2) there exists $e \in \Gamma_E$ such that Γ^e is connected. As in Lemma A.8 (2)

$$G \cong \pi_1(\Gamma^e) *_{\mathcal{G}_e}$$

Then $\mathcal{G}_v \subseteq \pi_1(\Gamma^e)$ is contained in a proper normal subgroup of G (see [Fri25: Lemma 93.5]) and the normal closure of $\mathcal{G}_v$ is not G. Contradiction! Hence, Γ is a tree.
If $T_E = \varnothing$, we have $T_V = \{v\}$ and the claim holds. Suppose $T_E \neq \varnothing$. By Lemma A.3 there exists $w \in \Gamma_V \setminus \{v\}$ of valence 1, i.e. there exists a unique edge $e \in T_E$ with $t(e) = w$. Consider the subtree $T' \subseteq T$ with $T'_V = T_V \setminus \{w\}$ and $T'_E = T_E \setminus \{e\}$. By the inductive argument from Lemma A.8 (2)

$$G \cong \pi_1(T') *_{\mathcal{G}_e} \mathcal{G}_w$$

Since $\mathcal{G}_w$ is abelian, there exists a commutative diagram

$$\begin{tikzcd}
\mathcal{G}_e \arrow[r] \arrow[d] & \mathcal{G}_w \arrow[d, two heads] \\
\pi_1(T') \arrow[r, "0"] & \mathcal{G}_w/\mathcal{G}_e
\end{tikzcd}$$

The universal property of $G \cong \pi_1(T') *_{\mathcal{G}_e} \mathcal{G}_w$ defines an epimorphism $G \twoheadrightarrow \mathcal{G}_w/\mathcal{G}_e$ whose kernel contains $\pi_1(T')$. Since $\mathcal{G}$ is reduced, $\mathcal{G}_w/\mathcal{G}_e$ is non-trivial and $\pi_1(T') \subseteq G$ is contained in a proper normal subgroup – contradicting that G is the normal closure of $\mathcal{G}_v$ as $\mathcal{G}_v \subseteq \pi_1(T')$. ∎

Having established some theory of reduced actions and graphs of groups, we can finally come to our main goal. The theorems in the following all share a common form: They have two parts – one for trees and one for graphs of groups – corresponding to each other under Bass–Serre theory. We start by sketching a special case that easily follows from the Grushko–Neumann Theorem 2.22:

Proposition A.21. *For a finitely presented group G there exists $n \geq 1$ such that the following hold:*

(1) *If T is a reduced tree with a G-action without edge inversions with trivial edge stabilizers, $\#(T/G)_V \leq n$.*
(2) *If G is the fundamental group of a reduced graph of groups $\mathcal{G}$ on a graph Γ with trivial edge groups, $\#\Gamma_V \leq n$.*

[47] Note that by Lemma A.11 this being an equality does not depend on the choice of spanning tree used to identify $\mathcal{G}_v$ with a subgroup of G.

Sketch of a proof. By Theorem A.15, Lemma A.19 and Lemma A.17 it suffices to prove (2): The inductive argument from Lemma A.8 (2) implies that

$$\pi_1(\mathcal{G}) \cong \pi_1(\Gamma) * \mathop{\ast}_{v \in \Gamma_V} \mathcal{G}_v$$

Hence, by Theorem 2.22

$$d(G) = d(\pi_1(\Gamma)) + \sum_{v \in \Gamma_V} d(\mathcal{G}_v) \geq \#\{v \in \Gamma_V \mid \mathcal{G}_v \neq \{e\}\}$$

Let $v \in \Gamma_V$ with $\mathcal{G}_v = \{e\}$. Since $\mathcal{G}$ is reduced, every $e \in \Gamma_E$ with $t(e) = v$ must also have $o(e) = v$. Since Γ is connected, there can be at most one $v \in \Gamma_V$ with $G_v = \{e\}$, and in this case $\Gamma_V = \{v\}$. We may therefore choose $n := \max\{d(G), 1\}$. ∎

The first interesting case of this phenomenon was observed by M. Dunwoody under the condition that the edge groups are finite:

Theorem A.22. *For a finitely presented group G there exists $n \geq 1$ such that the following hold:*
(1) *If T is a reduced tree with a G-action without edge inversions with finite edge stabilizers, $\#(T/G)_V \leq n$.*
(2) *If G is the fundamental group of a reduced graph of groups $\mathcal{G}$ on a graph Γ with finite edge groups, $\#\Gamma_V \leq n$.*

Proof. see [Dun85] ∎

In our motivating example from knot theory, the edge groups are isomorphic to $\mathbb{Z}$ generated by meridians – hence we cannot directly apply this theorem. There is however a slight generalization which deals with this case. Its geometric input is encapsulated by the following theorem which would also have formed the basis of the omitted proof for the last theorem:

Theorem A.23 (Dunwoody Resolution). *Let G be a finitely presented group. There exists $n \geq 1$ such that if T is a tree with a G-action there exists a map $\varphi \colon T' \to T$ of trees with G-actions from a tree T' with at most n G-orbits of essential vertices.*

Proof. see [DF87: Theorem 1.6], or [DD89: Theorem VI.4.4] for a textbook account ∎

We can now prove the theorem we need in the proof of Lemma 4.19. Its hypotheses are precisely such that they are satisfied by knot groups amalgamated along meridians:[48]

Theorem A.24. *For a finitely presented group G there exists $n \geq 1$ such that the following hold:*
(1) *If T is a reduced tree with a G-action without edge inversions such that*
  (i) *for all $e \in T_E$, $G_e \cong \mathbb{Z}$*
  (ii) *for all $e, f \in T_E$, the subgroups G_e and G_f are conjugate in G*
  (iii) *for all $v \in T_V$, the group G_v is finitely generated*
  (iv) *for all $e \in T_E$, the normal closure of G_e in $G_{t(e)}$ is $G_{t(e)}$*
  then $\#(T/G)_V \leq n$.
(2) *If G is the fundamental group of a reduced graph of groups $\mathcal{G}$ on a graph Γ such that*
  (i) *for all $e \in \Gamma_E$, $\mathcal{G}_e \cong \mathbb{Z}$*
  (ii) *for all $e, f \in \Gamma_E$, the edge groups $\mathcal{G}_e$ and $\mathcal{G}_f$ are conjugate in G*[49]
  (iii) *for all $v \in \Gamma_V$, the vertex group $\mathcal{G}_v$ is finitely generated*
  (iv) *for all $e \in \Gamma_E$, the normal closure of $\mathcal{G}_e$ in $\mathcal{G}_{t(e)}$ is $\mathcal{G}_{t(e)}$*
  then $\#\Gamma_V \leq n$.

Proof. By Theorem A.14, Proposition A.13 (1) and Lemma A.17, (1) implies (2). Hence, we only need to prove (1):

[48] The proof presented here is from [DF87].
[49] Note that by Lemma A.11 this does not depend on the choice of spanning tree for Γ used to identify $\mathcal{G}_e$ and $\mathcal{G}_f$ as subgroups of G.

By Theorem A.23 there exists $n \geq 1$ such that there exists a map $\varphi\colon T' \to T$ of trees with G-actions from a tree T' with at most n G-orbits of essential vertices. It suffices to show that for $v \in T_V$ there exists an essential vertex $v' \in T'_V$ with $\varphi_V(v') = v$.
Suppose there exists $v \in T_V$ such that $\varphi_V^{-1}(v)$ contains no essential vertices. Consider the subgroup action of $K := G_v \subseteq G$ on T'. We want to apply Lemma A.20 to this action.
First observe the following claim:

Claim. *For all $v' \in T'_V$ the stabilizer $K_{v'}$ is trivial or isomorphic to $\mathbb{Z}$.*

Proof. Assume $\varphi_V(v') \neq v$. Then there exists an edge $e \in T_E$ lying on the unique geodesic from v to $\varphi_V(v')$. By Lemma A.5 (1) $G_v \cap G_{\varphi_V(v')} \subseteq G_e$. Hence, the groups

$$K_{v'} \subseteq G_v \cap G_{\varphi_V(v')} \subseteq G_e$$

are all trivial or isomorphic to $\mathbb{Z}$ by (i).
Now assume $\varphi_V(v') = v$. Then $v' \in \varphi_V^{-1}(v)$ is inessential, i.e. there exists an edge $e' \in T'_E$ with $t(e') = v'$ and $G_{e'} = G_{v'}$. Then the groups

$$K_{v'} \subseteq G_{v'} = G_{e'} \subseteq G_{\varphi_E(e')}$$

are all trivial or isomorphic to $\mathbb{Z}$ by (i). □

Now let $e' \in T'_E$ be an edge. By (i) $G_{\varphi_E(e')}$ is generated by some $g \in G_{\varphi_E(e')}$. By (ii) and (iv), $K = G_v$ is normally generated by hgh^{-1} for some $h \in G$. The subgroup

$$K_{e'} \subseteq G_v \cap G_{\varphi_E(e')} \subseteq G_{\varphi_E(e')}$$

is generated by g^m for some $m \geq 1$. In particular, $g^m \in K_{t(e')}$. By Lemma A.5 (3) there exists $v' \in T'_V$ such that $g \in K_{v'}$. Hence, $hgh^{-1} \in K_{hv'}$ and K is the normal closure of $K_{hv'}$. Then Lemma A.20 implies that $K = K_{hv'}$ and by the claim this is trivial or isomorphic to $\mathbb{Z}$.
If $K = G_v$ is trivial, reducedness and connectedness of T imply that $T_V = Gv$. Then $\#(T/G)_V = 1$ and the theorem holds. Therefore we may now assume $T_V \neq Gv$ and $K \cong \mathbb{Z}$. By connectedness there exists $e \in T_E$ with $t(e) = v$, $o(e) \notin Gv$. By (iv) it follows that $G_e = G_{t(e)}$. But this contradicts T being reduced. ■

For completeness we point out that there now is a much more general version of these theorems with only small restrictions on the edge groups:

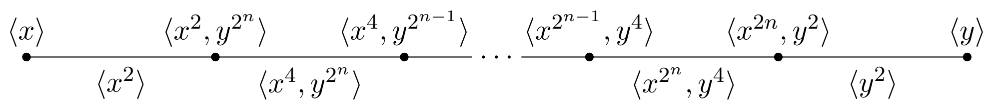
Theorem A.25. *For a finitely presented group G there exists $n \geq 1$ such that the following hold:*

1. *If T is a reduced tree with a G-action without edge inversions with 'small'[50] edge stabilizers, $\#(T/G)_V \leq n$.*
2. *If G is the fundamental group of a reduced graph of groups $\mathcal{G}$ on a graph Γ with 'small' edge groups, $\#\Gamma_V \leq n$.*

Proof. see [BF91] ■

Note however that there always needs to be some restriction on the edge groups:

Example A.26. For every $n \geq 0$



is a graph of groups with $n+2$ vertices whose fundamental group is $\langle x, y \rangle$. Of course, most of the edge groups of this graph of groups are free on two generators, i.e. not small.

[50] We will not provide a precise definition of 'small' here. Suffice it to say that a group is small if it does not allow a monomorphism from the free group on two generators. In particular, this is a generalisation of the previous theorems.

Bibliography

[Ale24] J.W. Alexander. 'On the Subdivision of 3-Space by a Polyhedron'. In: *Proc. Natl. Acad. Sci. U.S.A.* vol. 10, no. 1, pp. 6–8, 1924.

[Art25] E. Artin. 'Zur Isotopie zweidimensionaler Flächen im R_4'. In: *Abh. Math. Sem. Univ. Hamburg* vol. 4, no. 1, pp. 174–177, 1925.

[Bay85] E. Bayer-Fluckiger. 'Cancellation of hyperbolic ϵ-Hermitian forms and of simple knots'. In: *Math. Proc. Cambridge Philos. Soc.* vol. 98, no. 1, pp. 111–115, 1985.

[BCF21] I. Bokor, D. Crowley, S. Friedl, F. Hebestreit, D. Kasprowski, M. Land, J. Nicholson. 'Connected sum decompositions of high-dimensional manifolds'. In: *2019–20 MATRIX annals*. Vol. 4. MATRIX Book Ser. Springer, Cham, 2021, pp. 5–30.

[BF91] M. Bestvina, M. Feighn. 'Bounding the complexity of simplicial group actions on trees'. In: *Invent. Math.* vol. 103, no. 3, pp. 449–469, 1991.

[Bra42] A. Brauer. 'On a problem of partitions'. In: *Amer. J. Math.* vol. 64, pp. 299–312, 1942.

[Bro82] K.S. Brown. *Cohomology of groups*. Vol. 87. Graduate Texts in Mathematics. Springer, New York-Berlin, 1982.

[Cap76] S.E. Cappell. 'A splitting theorem for manifolds'. In: *Invent. Math.* vol. 33, no. 2, pp. 69–170, 1976.

[CD21] T. Ceccherini-Silberstein, M. D'Adderio. 'The Burnside Problem'. In: *Topics in Groups and Geometry*. Springer Monographs in Mathematics. Springer, Cham, 2021, pp. 87–98.

[CKS04] S. Carter, S. Kamada, M. Saito. *Surfaces in 4-space*. Vol. 142. Encyclopaedia of Mathematical Sciences. Springer, Berlin, 2004.

[DD89] W. Dicks, M.J. Dunwoody. *Groups acting on graphs*. Vol. 17. Cambridge Studies in Advanced Mathematics. Cambridge University Press, Cambridge, 1989.

[DF87] M.J. Dunwoody, R.A. Fenn. 'On the finiteness of higher knot sums'. In: *Topology* vol. 26, no. 3, pp. 337–343, 1987.

[Dun85] M.J. Dunwoody. 'The accessibility of finitely presented groups'. In: *Invent. Math.* vol. 81, no. 3, pp. 449–457, 1985.

[FQ90] M.H. Freedman, F. Quinn. *Topology of 4-manifolds*. Vol. 39. Princeton Mathematical Series. Princeton University Press, Princeton, NJ, 1990.

[Fri25] S. Friedl. *Topology*. AMS Open Math Notes 202307.111368. 2025.
URL: `www.ams.org/open-math-notes/omn-view-listing?listingId=111368`.

[Gor76] C.M. Gordon. 'A note on spun knots'. In: *Proc. Amer. Math. Soc.* vol. 58, pp. 361–362, 1976.

[Gra74] A. Grams. 'Atomic rings and the ascending chain condition for principal ideals'. In: *Proc. Cambridge Philos. Soc.* vol. 75, pp. 321–329, 1974.

[Hae61] A. Haefliger. 'Plongements différentiables de variétés dans variétés'. In: *Comment. Math. Helv.* vol. 36, pp. 47–82, 1961.

[Hae62] A. Haefliger. 'Knotted $(4k-1)$-spheres in $6k$-space'. In: *Ann. of Math. (2)* vol. 75, pp. 452–466, 1962.

[Hem76] J. Hempel. 3-*Manifolds*. Annals of Mathematics Studies, No. 86. Princeton University Press, Princeton, NJ, 1976.

[Kne29] H. Kneser. 'Geschlossene Flächen in dreidimensionalen Mannigfaltigkeiten'. In: *Jahresber. Dtsch. Math. Ver.* vol. 38, pp. 248–259, 1929.

[Kos93] A.A. Kosinski. *Differential manifolds*. Vol. 138. Pure and Applied Mathematics. Academic Press, Inc., Boston, MA, 1993.

[Kur34] A. Kurosch. 'Die Untergruppen der freien Produkte von beliebigen Gruppen'. In: *Math. Ann.* vol. 109, no. 1, pp. 647–660, 1934.

[Lan02] S. Lang. *Algebra*. third. Vol. 211. Graduate Texts in Mathematics. Springer, New York, 2002.

[Lee97] J. Lee. *Riemannian Manifolds*. Graduate Texts in Mathematics. Springer, New York, 1997.

[Lev65] J. Levine. 'Unknotting spheres in codimension two'. In: *Topology* vol. 4, pp. 9–16, 1965.

[Lev66] J. Levine. 'Polynomial invariants of knots of codimension two'. In: *Ann. of Math. (2)* vol. 84, pp. 537–554, 1966.

[Lev70] J. Levine. 'An algebraic classification of some knots of codimension two'. In: *Comment. Math. Helv.* vol. 45, pp. 185–198, 1970.

[Lev77] J. Levine. 'Knot modules. I'. In: *Trans. Amer. Math. Soc.* vol. 229, pp. 1–50, 1977.

[Lic97] W.B.R. Lickorish. *An introduction to knot theory*. Vol. 175. Graduate Texts in Mathematics. Springer, New York, 1997.

[Liv24] C. Livingston. 'Connected sums of codimension two locally flat submanifolds'. In: *Proc. Roy. Soc. Edinburgh Sect. A* vol. 154, no. 6, pp. 1937–1944, 2024.

[Löh17] C. Löh. *Geometric group theory*. Universitext. Springer, Cham, 2017.

[Mae77] T. Maeda. 'On a composition of knot groups. II. Algebraic bridge index'. In: *Math. Sem. Notes Kobe Univ.* vol. 5, no. 3, pp. 457–464, 1977.

[Mil62] J. Milnor. 'A Unique Decomposition Theorem for 3-Manifolds'. In: *Amer. J. Math.* vol. 84, no. 1, pp. 1–7, 1962.

[MT07] J. Morgan, G. Tian. *Ricci flow and the Poincaré conjecture*. Vol. 3. Clay Mathematics Monographs. American Mathematical Society, Providence, RI; Clay Mathematics Institute, Cambridge, MA, 2007.

[Neu99] J. Neukirch. *Algebraic number theory*. Vol. 322. Grundlehren der mathematischen Wissenschaften. Translated from German. Springer, Berlin, 1999.

[Pap57] C.D. Papakyriakopoulos. 'On Dehn's lemma and the asphericity of knots'. In: *Ann. of Math.* vol. 66, no. 1, pp. 1–26, 1957.

[Rub97] D. Ruberman. 'Null-homotopic embedded spheres of codimension one'. In: *Tight and taut submanifolds*. Vol. 32. Math. Sci. Res. Inst. Publ. Cambridge Univ. Press, Cambridge, 1997, pp. 229–232.

[Sch49] H. Schubert. 'Die eindeutige Zerlegbarkeit eines Knotens in Primknoten'. In: *S.-B. Heidelberger Akad. Wiss. Math.-Nat. Kl.* vol. 1949, no. 3, pp. 57–104, 1949.

[Ser80] J.-P. Serre. *Trees*. Translated from the French. Springer, Berlin-New York, 1980.

[Sha68] J.L. Shaneson. 'Embeddings with codimension two of spheres in spheres and H-cobordisms of $S^1 \times S^3$'. In: *Bull. Amer. Math. Soc.* vol. 74, pp. 972–974, 1968.

[Sma61] S. Smale. 'Generalized Poincaré's conjecture in dimensions greater than four'. In: *Ann. of Math. (2)* vol. 74, pp. 391–406, 1961.

[Sma62] S. Smale. 'On the structure of manifolds'. In: *Amer. J. Math.* vol. 84, pp. 387–399, 1962.

[Sta61] J. Stallings. 'On fibering certain 3-manifolds'. In: *Topology of 3-manifolds and related topics*. Proc. The Univ. of Georgia Institute. Prentice-Hall, Inc., Englewood Cliffs, NJ, 1961, pp. 95–100.

[SW79] P. Scott, C.T.C. Wall. 'Topological methods in group theory'. In: *Homological group theory*. Vol. 36. London Math. Soc. Lecture Note Ser. Cambridge Univ. Press, Cambridge, 1979, pp. 137–203.

[Tra87] B. Trace. 'Two comments concerning the uniqueness of prime factorizations for 3-manifolds'. In: *Bull. London Math. Soc.* vol. 19, no. 1, pp. 75–77, 1987.

[Tro73] H.F. Trotter. 'On S-equivalence of Seifert matrices'. In: *Invent. Math.* vol. 20, pp. 173–207, 1973.

[Wal65] C.T.C. Wall. 'Unknotting tori in codimension one and spheres in codimension two'. In: *Proc. Cambridge Philos. Soc.* vol. 61, pp. 659–664, 1965.

[Wal99] C.T.C. Wall. *Surgery on compact manifolds*. Second. Vol. 69. Mathematical Surveys and Monographs. Edited by A. Ranicki. American Mathematical Society, Providence, RI, 1999.