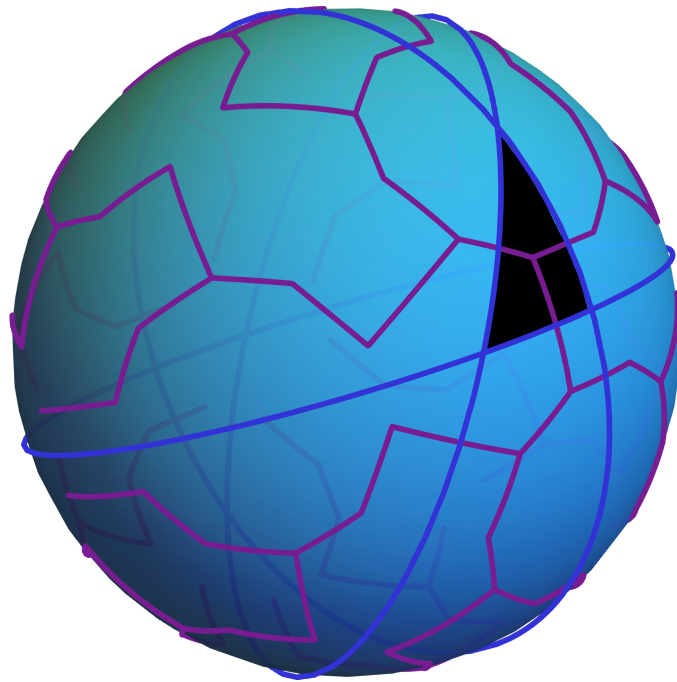UNIVERSITY COLLEGE LONDON

DEPARTMENT OF MATHEMATICS

Nielsen equivalence in Coxeter groups; and a geometric approach to group equivariant machine learning

THESIS SUBMITTED FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

BY

DAVID SHEARD

Supervised by
Associate Professor Larsen LOUDER
Second supervisor
Professor Konstanze RIETSCH

August 2022

Declaration

I, David Sheard, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Abstract

In this thesis we study two main threads. In Part I, we initiate the study of Nielsen equivalence in Coxeter groups—the classification of finite generating sets up to a natural action of the automorphism group of a free group. We explore different Nielsen equivalence invariants and adapt the method of Lustig and Moriah [79] to the Coxeter case. We also adapt the completion sequences of Dani and Levcovitz [31] to give a method of testing when generating sets of right-angled Coxeter groups are Nielsen equivalent.

Coxeter systems have a distinguished set of elements, called the reflections, from which generating sets can be drawn. We study generating sets of reflections separately. In this case, the natural notion of equivalence is generated by partial conjugations of one generator by another. This arises naturally for Weyl groups in the context of cluster algebras via quiver mutations [6]. We study this *mutation equivalence* for Weyl groups, and reflection equivalence for arbitrary Coxeter systems. In the latter case we leverage hyperplane arrangements in the Davis complex associated to a Coxeter system to give geometric criteria from when a set of reflections generates and a test for when generating sets of reflections are reflection equivalent.

In Part II, we discuss the other main topic of the thesis is group equivariant machine learning, based on joint work with Aslan and Platt [3]. We propose a novel approach to defining machine learning algorithms for problems which are equivariant with respect to some discrete group action. Our approach involves pre-processing the input data from a learning algorithm by projecting it onto a fundamental domain for the group action. We give explicit and efficient algorithms for computing this projection. We test our approach on three example learning problems, and demonstrate improvements in accuracy over other methods in the literature.

Impact statement

The primary impacts of the work presented in Part I are to mathematics academia. Coxeter groups are an well-studied class of groups which connections with many areas of mathematics. Nielsen equivalence, meanwhile, has a long history of interest, with applications to generating random group elements, for example. This work contributes to, and connects, these two separate areas.

In Chapter 3 we build on the work of Michael Barot and Bethany Marsh [6], trying to understand the connection between cluster algebras, quiver mutations, and presentations of Weyl groups. There the authors related these presentations to companion bases of the associated root system, while in [53], analogous presentations of braid groups are given a categorical interpretation. We work towards a new interpretation in terms of reflection equivalence of reflection generating sets of Weyl groups which may help understanding the connection further. Another impact of this work is cultural, as the proof of Proposition 3.18 was the inspiration for a collaborative art project with Melissa Rodd [98].

One of the most natural applications of studying Nielsen equivalence in Coxeter groups is to studying Nielsen equivalence in Artin groups. These form another very important class of closely related groups. This is because any Artin group (with a given Artin presentation) surjects onto a Coxeter group with a corresponding Coxeter presentation. Because of this, their theories parallel each other. More specifically, understanding Nielsen equivalence in Coxeter groups gives a method of finding Nielsen equivalence invariants for Artin groups via the surjection just mentioned, see Section 2.2.

A potential future application of the work in Chapter 4 is to the study of reflection quotients of Coxeter groups which have only received a little attention, see for example [84].

These impacts will be brought about by the distribution of this work in future publications.

The work in Part II is much more obviously applicable outside of its academic field, as well as inside. Our theoretical contributions include a unified framework

for intrinsic approaches to group equivariant machine learning based on lifting maps between quotient spaces. Separately we contribute to the well-developed field of algorithmic approaches to permutation groups. In particular we give the first general algorithm to construct fundamental domains for arbitrary groups acting on $\mathbb{R}^n$ by permuting coordinates, as well as for finding a projection map onto that fundamental domain.

Finally, of course, the machine learning algorithms we propose can be applied to a very wide range of real-world applications to improve efficiency and accuracy over existing algorithms. There are ethical questions, because our theoretical method, applied to some image recognition task, could in principle be used to improve detection of cancer just as well as it could be used by a nefarious organisation or government to improve facial recognition. These impacts will be brought about through publication of the work, which is already available as a preprint, see [3].

Acknowledgements

No man is an Iland, intire of itselfe

JOHN DONNE (1572–1631)

from 'Devotions Upon Emergent Occasions'

I have had the privilege to encounter and work with many amazing people throughout the course of my PhD, in whose absence this whole process would have been a great deal harder and less enjoyable. I wish to acknowledge the support, wisdom, friendship, love, advice, and companionship—mathematical and personal—for which I am extremely grateful. As is always the case, it is not possible to rank or compare the different contributions various people have made; however, our writing system demands I set forth my thanks in a linear order.

My thanks go to my supervisor, Lars Louder, for the time and effort he has put into guiding my research. For the many things I have learned from him; in particular, for my introduction to the wonderful world of geometric group theory, and the beautiful way of understanding groups through pictures. I must also thank him for introducing me to the problem of Nielsen equivalence.

To my secondary supervisor, Konni Rietsch, for her support and supervision of a project in my first year on Bruhat order. She is a pleasure to work with as an individual and taught me the skill of reading difficult papers.

To the incomparable centre manager for the LSGNT, Nicky Townsend, whose support has been stalwart in times of trouble and times of calm. Her many qualities and contributions are too numerous to list here, but she is indefatigably hard working, always a friendly face there to listen, and a dispenser of excellent advice.

To the other professional service staff in the Mathematics Department of UCL, my sincere thanks for all the work they have done to support students like me, which often goes unnoticed and unrecognised. Helen Higgins, Soheni Datta, Kate Fraser, Sukh Thiara, Harry Donnelly, Sam Hopkins, Ibrahim Bulime, Adrianna Mickiewicz, Richard Hoyle, Kay Musonda, and Abdal Asif. I would also like to

8

thank Helen Wilson for being a superb head of the UCL Mathematics Department.

I would like to thank my examiners Alessandro Sisto and Mark Hagen for their careful reading of my thesis; for their conscientious questions and comments on my work; and for making my viva a fun, relaxed, and very interesting occasion.

It has been my great pleasure to know Luciano Rila both as a colleague but more importantly a friend. He has been in so many ways my guide: for teaching; outreach; equity, diversity, and inclusion; being queer in a mathematical space; careers,; and so much more. Always a great friend and source of advice and inspiration.

Niki Kalaydzhieva is one of the most amazing people I know and she has been a superlative friend to me—a buttress to my Leaning Tower of Pisa in hard times. She is also an inspiration for me as a teacher and in her care for others and her work. Overall a wonderful and fun person to be around.

I have many fellow PhD students to thank. Ben Aslan and Daniel Platt, my collaborators for the work featured in Part II, as well as my friends in the LSGNT. Thanks to my PhD siblings Joe MacColl and Ben Millard for learning geometric group theory with me. Angela Wu, a great friend, supporter, companion, and a person with whom I have spent many happy hours. Jakob Stein who was a great friend throughout our undergraduates, and continues to be a great friend during our PhDs.

I want to thank Antonio d'Alfonso del Sordo for his friendship and love. He has made an indelible mark on the final year of my PhD and is someone I admire as a teacher and as a mathematician, and most importantly as a friend and companion. Every little thing he does amazes me.

I am also grateful to my other PhD colleagues at UCL and in the LSGNT. In particular, thanks to Petru Constantinescu, Bradley Doyle, Johannes Girsch, Andrew Graham, Matthew Habermann, Ashwin Iyengar, Lorenzo La Porta, Sarah Nowell, Pol van Hoften (for all the music), Hanneke Wiersema, Jenny Swinson, Galen Voysey, Albert Wood, Kristina Kubiliūtė, Sam Porritt, Federico Barbacovi, Federico Bongiorno, Dominik Bullach, Roberto Ladu, Jaime Mendizabal, Miriam Norris, Corvin Paul, Qaasim Shafi, Tom Sharpe, Teresa Ludenbach, Hannah Tillmann-

UCL research paper declaration form

1. **For a research manuscript that has already been published:**

   Not yet published.

   (a) **Where was this work published?**

   (b) **Who published the work?**

   (c) **When was the work published?**

   (d) **Was the work peer reviewed?**

   (e) **Have you retained the copyright?**

2. **For a research manuscript prepared for publication but that has not yet been published:**

   (a) **Has the manuscript been uploaded to a preprint server? If so, please specify which server:**

   Yes, arXiv, see [3].

   (b) **Where is the work intended to be published?**

   The Thirty-Sixth Annual Conference on Neural Information Processing Systems (NeurIPS 2022). It will be published as part of the conference proceedings.

   (c) **List the manuscript's authors in the intended authorship order:**

   Benjamin Aslan, Daniel Platt, and David Sheard.

   (d) **Stage of publication:**

   Accepted.

3. **For multi-authored work, please give a statement of contribution covering all authors:**

   The text in Chapters 6–9 of this thesis is adapted from [3]. The first draft of that paper was mostly written by David Sheard, although the literature review and experimental results sections (which became Sections 6.2 and 7.4) were drafted by Daniel Platt, and the sections on the Universal Approximation Theorem and Dirichlet fundamental domains (which became Sections 7.2.1 and 9.1) were drafted by Benjamin

Aslan. All authors subsequently contributed to and edited all sections of the paper. Moreover, significant changes and several additions were made when compiling this thesis.

In terms of the content, the initial idea to use some form of geometric projection map was Daniel Platt's. Benjamin Aslan implemented the projections we used for running experiments, and computed the polynomial invariants in Example 9.6. These experiments were then run by Daniel Platt. Benjamin Aslan also proposed a significant simplification to the original proof of Proposition 8.30.

While Benjamin Aslan and Daniel Platt also made valuable contributions in discussion to the other mathematical content, this was the work of David Sheard. In particular, the Universal Approximation Theorems in Section 7.2.1, the unified approach to equivariant machine learning in Section 7.3; the combinatorial projection maps and related algorithms discussed in Chapter 8; and the quotient space projections and other topics discussed in Sections 9.2–9.5.

4. **In which chapter(s) of your thesis can this material be found?**
   Chapters 6–9.

Candidate's signature: **Date:**

Contents

II Equivariant machine learning   200

List of figures

List of tables

Notation for Part I

$\angle rr'$	Angle between r and r' 127
$[g,x]_{\mathcal{U}(G,K)}$	A point in $\mathcal{U}(G,K)$ 117
#	The cardinality of a set 66
$\succeq$	Domination relation on $\mathcal{Q}(\overline{\mathcal{V}})$ 94
A	Commutative ring 74
A^*	Group of units of A 75
$a_d(s,s')$	Alternating word $ss's\cdots$ of length d 128
A_G	Subgroup of units of A associated to G 75
A_Γ	Artin group associated to the presentation diagram Γ 64
α_i	Surjective homomorphism of free groups 62
α_w	Inner automorphism associated to the element w 101
$A * B$	Free product of two groups 58
Aut	Automorphism group 55
B	Symmetric bilinear form 53
$\mathrm{Bs}(V,\mathcal{A})$	The barycentric subdivision of the abstract simplicial complex $(V,\mathcal{A})$ 160
$\mathrm{Ch}(A,\preceq)$	Set of chains in the poset $(A,\preceq)$ 157
χ_η	Lustig-Moriah invariant associated to the representation η 75
$C(\Omega,v_0)$	Core graph of (Ω,v_0) 169

$\mathrm{Cone}(V, \mathcal{A}))$	Cone on the abstract simplicial complex $(V, \mathcal{A})$ 160
δ_{ij}	Kronecker δ function 73
Dih_k	Dihedral group of order $2k$ 64
$\mathbb{E}^2$	Euclidean plane 149
EC	Extended conjugacy class 72
η	Representation of the group ring $\mathbb{Z}G$ 74
$\mathbb{F}$	Free group, either of rank n, $\mathbb{F}_n$; or on generating tuple X, $\mathbb{F}(X)$ 169
f_E	Map induced on the edge set by f 174
f_V	Map induced on the vertex set by f 174
$\Gamma(W, S)$	Presentation diagram for a Coxeter system 50
gcd	Greatest common divisor 59
${}^h g$	The (left) conjugate hgh^{-1} of g by h 77
G_Q	Quiver group associated to Q 86
Gr	Graph automorphism group 101
$\mathbb{H}^2$	Hyperbolic plane 59
$H(Q)$	The total sum of the height function on Q 93
h^Q	Height function on Q 93
H_{s_i}	Hyperplane fixed by s_i 53
$H_{s_i}^+$	Positive half-space bounded by H_{s_i} 53
$\mathcal{H}_X$	Hyperplane arrangement associated to X 129
I_S^A	Correction ideal for a ring A and generating tuple S 75
Inn	Inner automorphism group 101
K_s	Mirror of the complex K associated to s 116
K^X	Subcomplex of Σ containing K and bounded by $\mathcal{H}_X$ 130
$\mathrm{lk}(v)$	Link of the vertex v 107
$\ell_S = \ell$	Length function on W with respect to S 53
$L(W, S)$	Nerve of a Coxeter system 119

m_{ij}	Order of $s_i s_j$ 49
$\mathbb{M}_m(A)$	Ring of $m \times m$ matrices over A 74
$m_{rr'}$	Order of rr' 130
$\mu_v(Q)$	Mutation of the quiver Q at vertex v 84
$\mu_v^{(2)}(Q)$	Mutation modulo 2 of the quiver Q at vertex v 106
$\mu_w(\overline{\mathcal{V}}, \phi)$	Mutation of $(\overline{\mathcal{V}}, \phi)$ corresponding to the word w 101
$\mathcal{O}_\Gamma$	Complex of groups representing W_Γ 164
Ω	A cube complex 163
$\widehat{\Omega}$	Completion of a Γ-complex Ω 167
$\widehat{\Omega}^{\text{free}}$	Free completion of Ω 188
Ω_X	Rose graph associated to X 168
$O_n(B)$	Orthogonal group with respect to B 53
$\mathcal{O}^Q$	Orientation sign function on Q 93
∂_{x_i}	Derivation of a free group associated to the generator x_i 73
$\partial_S \mathbf{T}$	Matrix of derivations of T with respect to S 74
ϕ	Marking of a group, or map from a quiver group to a Weyl group 55, 88
φ	Euler's totient function 60
ψ_v	Isomorphism of quiver groups associated to mutation at v 88
Q	A quiver 84
$\check{Q}$	Quiver with all edges oriented towards v_0 94
$\widehat{Q}$	Quiver with all edges oriented away from v_0 94
(Q, ϕ)	Presentation quiver 89
$\mathcal{Q}(\Upsilon)$	Set of quivers with underlying weighted graph Υ 92
ρ	Faithful geometric representation of a group 53, 75
$R(W, S)$	Set of reflections in a Coxeter system 50

S	Coxeter generating tuple 49
$\mathbb{S}^2$	Sphere 149
Σ_r	Subcomplex of Σ fixed by r 123
Σ_r^+	Half-space of Σ with respect to Σ_r containing K 124
Σ_r^-	Half-space of Σ with respect to Σ_r not containing K 124
$\Sigma(W,S)$	The Davis complex of (W,S) 120
S_n	The group of permutations of n letters 104
$\mathrm{Sub}(\Omega)$	Cubical subdivision of Ω 163
$\mathcal{S}(W,S)$	The set of spherical subsets of S 119
$\mathcal{S}_{\supset\emptyset}$	The poset of non-empty spherical subsets of S 119
S_X	Abstract Coxeter generating tuple associated to X 130
$S(x)$	Elements of $s \in S$ such that x is in the mirror K_s 116
$\mathbf{t}$	A word representing the element t 74
$\mathcal{U}(G,K)$	Basic construction associated to G and K 117
$\mathcal{V}(W,S)$	Coxeter-Dynkin diagram for a Coxeter system 50
$\overline{\mathcal{V}}$	Weighted graph obtained from $\mathcal{V}$ 86
$(\overline{\mathcal{V}},\phi)$	Presentation Dynkin diagram 92
$\overline{\mathcal{V}}_T$	Weighted graph associated to the tuple T 104
W	Coxeter group 49
W^{ab}	Abelianisation of the Coxeter group W 51
$\overline{W}_X$	Coxeter group abstractly associated to X 35, 130
W_Γ	Coxeter group with presentation diagram Γ 50
(W,S)	Coxeter system 49
W_T	Special subgroup generated by T 50
$W(\mathcal{V})$	Coxeter group with Coxeter-Dynkin diagram $\mathcal{V}$ 50

W_X Reflection subgroup of W generated by X 126

X Finite tuple of group elements 126, 168

ξ Quotient of the ring A by the correction ideal I_S^A 75

$\mathbb{Z}G$ Integral group ring of G 74

Notation for Part II

$[c]$	Chamber containing the point $c \in C$ 238
$g \cdot x$	Action of g on x 206
$\| \cdot \|_\infty^X$	Supremum norm over the set X 205
$\| \cdot \|_p^\mu$	L^p norm with respect to the measure μ 204
$\mathbb{1}$	Vector $(1, \ldots, 1)$ 207
α	Continuous function to be approximated by a machine learning model 202
$\overline{\alpha}$	Map induced between quotient spaces from the map α 210
B	A base, ie an ordered subset $(b_1, \ldots, b_k)$ of N 227
β	Machine learning model, ie a variable function which can approximate α 202
$\overline{\beta}$	Machine learning model on a fundamental domain or quotient space 210
$b(G)$	The minimum size of a base for the group G 237
C	Set of permutations of the point $(1, 2, \ldots, n)$ 238
$\mathfrak{C}^G(X, \mathbb{R}^m)$	Class of G-invariant continuous functions $X \to \mathbb{R}^m$ 214
$\mathfrak{C}(X, \mathbb{R}^m)$	Class of continuous functions $X \to \mathbb{R}^m$ 205
Δ_i	Orbit of b_i under G_{i-1} 227
D_{train}	Set of training data 203
$D_{\text{train}}^{\text{aug}}$	Set of augmented training data 206

D_{train}^{π}	Set of training data after the projection π has been applied 210
ε	Perturbation vector 227
$\mathcal{F}$	Fundamental domain 210
$\overline{\mathcal{F}}$	Closure of the fundamental domain $\mathcal{F}$ 210
$\partial\mathcal{F}$	Boundary of the fundamental domain $\mathcal{F}$ 212
G	Group acting discretely by isometries 211
Γ_i	Part of the partition Π_i containing b_i 242
G_i	Point-wise stabiliser of the first i elements of a base 227
H_i	Product of symmetric groups over a partition of N 242
$\mathbb{I}$	Identity matrix 207
λ_i	Linear map between real vector spaces 203
$L_G^p(\mu, \mathbb{R}^m)$	Class of G-invariant functions $\mathbb{R}^n \to \mathbb{R}^m$ with finite L^p norm with respect to μ 213
$L^p(\mu, \mathbb{R}^m)$	Class of functions $\mathbb{R}^n \to \mathbb{R}^m$ with finite L^p norm with respect to μ 204
$\mathfrak{M}_{n,*,m}^G(\sigma)$	Class of functions of the form $\beta \circ \pi$ where $\beta \in \mathfrak{M}_{n,*,m}(\sigma)$ and π is a G-invariant projection 213
$\mathfrak{M}_{n_0,n_1,\dots,n_{\ell+1}}(\sigma)$	Class of functions given by a neural network with ℓ hidden layers and activation function σ 202
$\mathfrak{M}_{n,*,m}(\sigma)$	Class of functions given by a neural network with 1 hidden layer and activation function σ 204
μ	Linear map which averages over all but one coordinate of a tensor 228
N	Set of indices $\{1, \dots, n\}$ 226
(N, E)	Directed graph with vertex set N 233
ω	Edge labelling of a graph 233
ϕ	Map $X \to G$ used to define π 211
$\phi_\uparrow$	Map $X \to G$ used to define $\pi_\uparrow$ 227

π	Projection map 210
π_{Dir}	Projection onto a Dirichlet fundamental domain 252
$\pi_{\downarrow}$	Descending combinatorial projection 228
$\pi_{\downarrow\mathrm{av}}$	Descending averaging combinatorial projection 228
$\pi_{\uparrow}$	Ascending combinatorial projection 227
$\pi_{\uparrow\mathrm{av}}$	Ascending averaging combinatorial projection 228
π_X	Projection from X onto its quotient space 212
Π_i	G_i-invariant partition of N 242
$p(x)$	Polynomial invariant in $x_1, \ldots, x_n$ 207
$q(x)$	Polynomial equivariant in $x_1, \ldots, x_n$ 207
R	Set of orbit representatives 211
ρ	Bijection between S_n and C 239
$\mathbb{R}^n_{\mathrm{dist}}$	Set of points in $\mathbb{R}^n$ with distinct coordinates 238
$\mathbb{R}[x_1, \ldots, x_n]$	Ring of real polynomials in $x_1, \ldots, x_n$ 207
$\mathbb{S}^n$	Unit sphere of n dimensions 220
σ	Activation function 203
S_n	Group of permutations on n letters 225
$\mathrm{Stab}_G(x)$	Stabiliser of x in the group G 217
$\mathrm{Sym}(\Omega)$	Group of permutations of Ω 228
U_i	Right transversal for $\mathrm{Sym}(\Delta_i) \times \mathrm{Sym}(\Gamma_i - \Delta_i)$ in $\mathrm{Sym}(\Gamma_i)$ 242
$\Upsilon(G)$	Edge labelled directed graph encoding G 233
X	Input space for a machine learning problem 204
x'	Perturbed point obtained from x 227
X/G	Quotient space 212
Y	Output space for a machine learning problem 205

Chapter 0

Introduction

THE FIRST PART OF THIS thesis studies generators of Coxeter groups. We consider *Nielsen equivalence*, which is the natural notion of equivalence on the set of finite ordered generating sets of a group which comes from the universality of free groups. We discuss several special cases of the problem of classifying finite generating tuples of Coxeter groups. These include classifying certain reflection generating tuples of *Weyl groups* arising from the theory of *cluster algebras* (Chapter 3); reflection generating tuples in arbitrary Coxeter groups, with respect to some choice of 'reflections' in the group, (Chapter 4); and arbitrary generating tuples in the class of *right-angled* Coxeter groups (Chapter 5).

In Part II we discuss a second topic: a novel application of well-studied ideas from geometric group theory (namely fundamental domains and quotient spaces of group actions) to the problem of *equivariant machine learning*: a highly active area of research in computer science. We use the universal property of quotient spaces to give a unified framework for so-called *intrinsic* approaches to equivariant machine learning. In the case that the group is a subgroup G of the symmetric group S_n which acts on $\mathbb{R}^n$ by permuting coordinates, we use the geometry of S_n thought of as a Coxeter group of type A acting on its *Coxeter complex* (or equivalently via the Tits representation, or on its Davis complex) to define an algorithm which computes a fundamental domain for the action of G on $\mathbb{R}^n$, and a G-invariant projection map from $\mathbb{R}^n$ to this fundamental domain.

A feature which unifies the two parts of this thesis is a tendency towards al-

gorithmic approaches to problems. In Part II this is clear as we are designing and studying machine learning algorithms. In Part I we give algorithmic approaches to studying three related and progressively weaker notions of equivalence of generating tuples of Coxeter groups. First we look at *mutation equivalence* of generating tuples of Weyl groups; then we give algorithms to test when a tuple of reflections generates a given Coxeter group and when two such tuples are *reflection equivalent*; and finally we give algorithms to test when a tuple of elements generates a given right-angled Coxeter group and when two tuples are *Nielsen equivalent*. Although we do not discuss this in great detail, we have also written a software package to aid in studying Coxeter groups, see [102], in which some of these algorithmic approaches have been implemented.

0.1 Nielsen equivalence in Coxeter groups

Nielsen equivalence is a natural way to classify different generating tuples of a group G. Consider two generating tuples of the same size $S = (s_1, \ldots, s_n)$ and $T = (t_1, \ldots, t_n)$ for G, and let $\mathbb{F}_n = \langle x_1, \ldots, x_n \rangle$ be the free group of rank n. By the universal property of free groups, the inclusion $S \hookrightarrow G$ induces a unique surjective homomorphism $\phi_S \colon \mathbb{F}_n \to G$ such that $x_i \mapsto s_i$ for each $1 \leqslant i \leqslant n$. Such a map is called a *marking* of G. The tuples S and T are said to be *Nielsen equivalent* if there is an automorphism $\alpha \in \mathrm{Aut}(\mathbb{F}_n)$ such that $\phi_T \coloneqq \phi_S \circ \alpha$ maps $x_i \mapsto t_i$ for each $1 \leqslant i \leqslant n$—in particular, ϕ_T is the marking of G associated to the inclusion $T \hookrightarrow G$ (see Definition 1.15). Early work on Nielsen equivalence tended to use combinatorial techniques [55, 117] or algebraic techniques [78], although geometric techniques have also started to play a significant role [104, 76, 39]. Nielsen equivalence also has connections with (simple-)homotopy type classification, see for example [77].

The purpose of Part I of this thesis is to initiate the study of Nielsen equivalence in the class of Coxeter groups. These groups are very important in the field of geometric group theory, with very strong connections to areas like combinatorics; Lie theory; classical geometry; and even algebraic geometry and representation

theory. Geometrically, they can be defined as groups which act discretely on some space so that they are generated by a finite tuple of reflections. This translates into a more rigorous algebraic definition in terms of a presentation. A Coxeter group is a group admitting a presentation of the form

$$W = \langle s_1, \ldots, s_n \mid s_i^2, (s_i s_j)^{m_{ij}} \rangle, \text{ where } m_{ij} = m_{ji} \in \{2, 3, \ldots, \infty\}. \qquad (1)$$

Coxeter groups are closely related to some other classes of groups which have already been well studied with respect to Nielsen equivalence, for example surface groups [117, 76], and more generally Fuchsian groups [99, 79].

To distinguish Nielsen equivalence classes it is useful to have good invariants. Several invariants for Nielsen equivalence have been studied over the years, from the simple [90] to the complex [78], however there are well-known problems which arise when trying to construct invariants in groups with 2-torsion. Since Coxeter groups are generated by reflections (involutions), this presents difficulties. In Chapter 2, we discuss several approaches to the problem in the case of Coxeter groups. We discount the elementary approaches to Nielsen equivalence invariants, since none of these are usefully applicable in general. Instead, we follow the approach of Martin Lustig and Yoav Moriah in their study of Fuchsian groups [79]. For a group G, they define an invariant $\chi_\eta \colon G^n \to R$ associated to a representation $\eta \colon \mathbb{Z}G \to \mathbb{M}_m(A)$ of the integral group ring of G to a matrix ring over some abelian ring A. This invariant is valued in some quotient ring $R = A/I$ of A. We are able to characterise all such 1-dimensional invariants for Coxeter groups.

Theorem A (Theorem 2.24): *Let W be a Coxeter group with integral group ring $\mathbb{Z}W$ and let $\eta \colon \mathbb{Z}W \to A$ be a 1-dimensional representation of $\mathbb{Z}W$ to an abelian ring A. Then the Lustig-Moriah invariant χ_η factors through $\chi_{\eta^{ab}}$ where $\eta^{ab} \colon \mathbb{Z}W \to \mathbb{Z}W^{ab}$. In particular χ_η is valued in some quotient of $\mathbb{Z}_{m_0}$ where m_0 is the greatest common divisor of $\{m_{ij} \mid i \neq j\}$.*

In the case that $n = 2$, ie W is a dihedral group, this provides a complete Nielsen equivalence invariant for generating pairs. In general, however, this seems

to offer quite a coarse invariant and often $m_0 = 1$, making this invariant useless. We also consider higher dimensional invariants where η is a 'mixed' representation based on the Tits representation of $W \to \mathrm{GL}_n(\mathbb{R})$, but this invariably leads to single-valued invariants.

Thereafter, our study focusses on constructing equivalences between generating tuples, rather than finding invariants to show inequivalence. This falls into two main strands: classifying reflection generating tuples of arbitrary Coxeter groups up to a stronger notion of equivalence which we call *reflection equivalence*; and studying Nielsen equivalence in the restricted class of right-angled Coxeter groups (and their quasiconvex subgroups).

0.1.1   Reflection equivalence in Coxeter groups

Typically Coxeter groups are not studied as isolated groups, but as *Coxeter systems* (W, S) where S is a finite tuple of generators, with respect to which W admits a presentation of the form given in (1). Then the elements of S are called the *simple reflections* of the Coxeter system, and the set of their conjugates

$$R = \{wsw^{-1} \mid s \in S \text{ and } w \in W\} \tag{2}$$

is called the set of *reflections* of the Coxeter system.

In general, a Coxeter group may admit many different Coxeter systems which are inequivalent up to automorphism, and may have different sets of reflections and even have different ranks (ie sizes of the tuple S)—see the discussion of *rigidity* in Section 2.1. Moreover the algebraic rank of a Coxeter group, ie the minimum size of any generating tuple, may be very different from their *Coxeter rank* (the minimum size of any Coxeter generating tuple). As an example, all Coxeter groups which admit *finite irreducible* Coxeter systems (see Definition 1.5) have algebraic rank 2, but can have arbitrarily large Coxeter ranks.

This demonstrates that studying generating tuples of reflections is *highly dependent* on the choice of Coxeter system for a Coxeter group, while Nielsen equivalence is independent of the choice of Coxeter system. Nevertheless, the study of

Coxeter systems rather than Coxeter groups is an extremely important and rich field of research. This motivates us to separate out questions about classifying generating tuples into two areas:

Question 0.1. When are two generating tuples of a Coxeter group Nielsen equivalent?

Question 0.2. When are two generating tuples of reflections of a Coxeter group with respect to a given Coxeter system equivalent up to a suitable notion of equivalence?

This second question is the subject of Chapters 3 and 4. It does not make sense to study reflection generating tuples up to Nielsen equivalence since any generating tuple of reflections is Nielsen equivalent to many generating tuples containing non-reflections. As an example, if $T = (t_1, t_2)$ is a generating tuple of reflection for a Coxeter group with fixed choice of Coxeter system (W, S), then it is Nielsen equivalent to $T' = (t_1 t_2, t_1)$, but $t_1 t_2$ is a rotation or a translation, geometrically speaking, and is not conjugate to any element of S.

Instead we consider a different notion of equivalence which is a strengthening of Nielsen equivalence and does preserve the set of reflections. For technical reasons, we allow a reflection generating tuple to include the identity as a generator. Then the equivalence relation generated by the transformations of the form

1. $(t_1, \ldots, t_n) \mapsto (t_{\sigma(1)}, \ldots, t_{\sigma(n)})$ for some $\sigma \in S_n$,

2. $(t_1, \ldots, t_i, \ldots, t_n) \leftrightarrow (t_1, \ldots, 1, \ldots, t_n)$ if $t_i = t_j$ for some $j \neq i$,

3. $(t_1, \ldots, t_i, \ldots, t_n) \mapsto (t_1, \ldots, t_j t_i t_j^{-1}, \ldots, t_n)$ for some $j \neq i$.

The third kind of transformation is called a *partial conjugation*. Since reflections are involutions, we could replace t_j^{-1} with t_j, and it follows that this transformation is a self-inverse. These transformations preserve the property of being a generating tuple, and if t_i is a conjugate of some $s \in S$, then so is $t_j t_i t_j^{-1}$. We call the notion of equivalence this gives *reflection equivalence*, see Definition 2.6.

Quiver mutations and Weyl groups

Another motivation for studying this notion of equivalence is that it arises naturally, and quite unexpectedly, out of the theory of mutations of quivers—oriented and edge-labelled graphs. This applies only to a special class of finite Coxeter groups called Weyl groups, which are classified by their *Coxeter-Dynkin diagrams* (see Definition 1.2)—an unoriented edge-labelled graph associated to the presentation (1). A quiver is an *oriented* and edge-labelled graph, and there is a transformation which can be applied to quivers which is called *mutation*, see Definition 3.2,. These transformations are particularly nicely behaved if the quiver is obtained by orienting a Coxeter-Dynkin diagram associated to a Weyl group—such quivers are called *finite type*.

Given a quiver which can be transformed into a(n oriented) Coxeter-Dynkin diagram $\mathcal{V}$ by a sequence of mutations, it is possible to associate a tuple of reflection generators to the the Weyl group classified by $\mathcal{V}$ (in fact one can associate a whole presentation to such a quiver which looks like (1) with finitely many additional relations) [6]. When a mutation is performed, the effect is to change the associated tuple of reflection generators by a finite sequence of partial conjugations. We call the equivalence relation which this induces *mutation equivalence* see Definition 3.12.

A priori, mutation equivalence is a stronger equivalence relation than reflection equivalence. It is almost a tautology to say that any reflection generating tuple associated to a finite type quiver is reflection equivalent to the standard one, see Proposition 3.16. In Chapter 3, we study to what extent mutation equivalence is in fact stronger than reflection equivalence. In particular we prove the following.

Theorem B (Theorem 3.36 and Corollary 3.37)**:** *Let W be a Weyl group (with its standard Coxeter system), then any reflection generating tuple associated to a quiver is mutation equivalent to a reflection generating coming from the Coxeter-Dynkin diagram (up to diagram automorphism). In particular, for Coxeter-Dynkin diagrams of type A_n, B_n, D_{2k+1}, or E_n, all such reflection generating tuples are mutation equivalent.*

The main part of the proof is in Proposition 3.18 which allows us to arbitrar-

ily change the orientation on a quiver whose underlying graph is a tree without affecting the presentation associated to the quiver. This proof admits a nice topographical interpretation in which we define a *height function* (see Definition 4.58) on the quiver which encodes the orientations on the edges, and then we perform a sequence of mutations at so-called *prominent vertices*, which are likened to *erosion* and *elevation*.

Reflection equivalence in arbitrary Coxeter systems

The rather specialised case of generating tuples of Weyl groups coming from mutations of Coxeter-Dynkin diagrams serves principally as a motivation for studying reflection generating tuples in arbitrary Coxeter systems. In this level of generality, quiver mutation is not *a priori* a viable tool to study reflection equivalence. The reason is that arbitrary quivers are not well-behaved under mutation in the same way that finite type quivers are, meaning it not clear how to associate presentations of arbitrary Coxeter groups to quivers so that these presentations transform nicely under mutations.

Instead we approach the problem from a geometric viewpoint. Given a Coxeter system (W, S), there is a CW-complex Σ on which it acts faithfully called the *Davis complex*, see Section 4.1. This complex has the Cayley graph of (W, S) as its 1-skeleton, and can be given a $\mathrm{CAT}(0)$ metric such that the action is discrete and by isometries. It also has a fixed choice of fundamental domain for the action of W called the *fundamental chamber* and denoted K.

Any reflection $t \in R$ (recall (2)) has fixed set $\Sigma_t \subset \Sigma$ which has a collar neighbourhood. $\Sigma - \Sigma_t$ has two connected components and t swaps these components. Thus t acts geometrically as a 'reflection' in Σ, and Σ_t behaves like a reflection hyperplane.

We can represent a tuple of reflections in (W, S) by a hyperplane arrangement

$$T = (t_1, \ldots, t_n) \mapsto \{\Sigma_{t_i} \mid 1 \leqslant i \leqslant n\} = \mathcal{H}_T,$$

and then the partial conjugation which transforms t_i into $t_j t_i t_j^{-1}$ can be interpreted geometrically by replacing Σ_{t_i} with its reflection $t_j \Sigma_{t_i}$ in Σ_{t_j}.

Before using this geometric viewpoint to classify reflection generating tuples up to reflection equivalence, we need to understand when a tuple of reflections correspond to a Coxeter generating tuple. Let T be a finite tuple of reflections, and write W_T for the subgroup of W they generate. It is always true that W_T is a Coxeter group, but (W_T, T) may not be a Coxeter system in the sense that T does not give rise to a presentation of the form (1). An alternative way to put this is to consider the Coxeter group abstractly generated by T with presentation

$$\overline{W}_T := \langle T \mid t^2,\ (tt')^{m_{tt'}},\ \text{for all } t, t' \in T \rangle,$$

where $m_{tt'}$ is the order of tt' in W. (W_T, T) will fail to be a Coxeter system if the map $\overline{W}_T \to W_T \colon t \mapsto t$ is not injective.

We first prove a criterion for when (W_T, T) is a Coxeter system, which is a reinterpretation of a criterion due to Matthew Dyer [42]. Our proof is independent of that original result. To state it we need to define what an *outlier* is.

Let T be a finite tuple of reflections, and W_T the group it generates. The hyperplane arrangement associated to T divides Σ into components. Consider the component K_T of $\Sigma - \mathcal{H}_T$ which contains the fundamental chamber K. We say T contains *no outliers* if for every proper subset T' of T, $\Sigma - \mathcal{H}_{T'}$ has a component which strictly contains K_T as a subset, see Definition 4.28 for an equivalent definition. If T has no outliers, then for any $t, t' \in T$, if Σ_t and $\Sigma_{t'}$ meet in Σ, then they do so in the boundary of K_T.

THEOREM C (Equivalent to Theorem 4.34): *The pair (W_T, T) is a Coxeter system if and only if T has no outliers and for any distinct $t, t' \in T$, either Σ_t and $\Sigma_{t'}$ do not meet; or the dihedral angle between them, as measured in K_T, is π/m for some integer m.*

It is possible to use partial conjugations to turn any tuple of reflections into one containing no outliers, but in general the resulting tuple does not satisfy the conditions on the angles. We introduce a new transformation which takes two reflections $t, t' \in T$ such that Σ_t and $\Sigma_{t'}$ meet at an angle $k\pi/m$ where $\gcd(k, m) = 1$, and replaces t' with another reflection t'' in W_T such that the angle between Σ_t and $\Sigma_{t''}$ is π/m. This is, in general, not a reflection equivalence, but when combined

with the criterion above allows us to prove the following.

THEOREM D (Theorem 4.40 and Corollaries 4.41 and 4.42): *Let T be a finite tuple of reflections in (W, S), then there is an algorithm which produces a tuple of reflections $\widetilde{T}$ that generates W_T such that $(W_T, \widetilde{T})$ is a Coxeter system; computes the index $[W : W_T]$; and hence tests whether or not T generates W.*

Restricting ourselves only to partial conjugations, this algorithm yields the following classification of reflection generating tuples of (W, S).

THEOREM E (Theorem 4.44): *Let T be a finite tuple of reflections which generate (W, S), then T is reflection equivalent to a generating tuple of reflections which contains no outliers, and such that the angles between any two Σ_t and $\Sigma_{t'}$ which meet are non-obtuse.*

In particular, it follows that if W is a Weyl group, then all reflection generating tuples are reflection equivalent, see Corollary 4.45.

The Theorem leaves open the possibility that some generating tuples of reflections are inequivalent, and this is indeed the case. The simplest example is that of the dihedral group of order 10, which admits a Coxeter system $(\mathrm{Dih}_5, (s, s'))$. Then $(s, s'ss')$ is a reflection generating tuple which is not reflection equivalent to (s, s'). However, if we allow ourselves to use one extra generator, by replacing $T = (t_1, \ldots, t_k)$ with $(t_1, \ldots, t_k, 1)$, then things become much simpler.

This process of increasing the size of a generating tuple is called *performing a stabilisation* of T, see Definition 1.17. More generally we will say that a reflection generating tuple $T' = (t'_1, \ldots, t'_{k'})$ is a *stabilisation* of T if T' can be obtained from T by performing $(k' - k)$ stabilisations. After performing $\#S = n$ stabilisations, all reflection generating tuples become reflection equivalent to a stabilisation of S for trivial reasons since we can find reflection equivalences as follows

$$T = (t_1, \ldots, t_k) \xrightarrow{\text{stabilise}} (t_1, \ldots, t_k, \overbrace{1, \ldots, 1}^{n})$$
$$\xrightarrow[\text{equivalence}]{\text{reflection}} (t_1, \ldots, t_k, s_1, \ldots, s_n)$$
$$\xrightarrow[\text{equivalence}]{\text{reflection}} (\underbrace{1, \ldots, 1}_{k}, s_1, \ldots, s_n).$$

A rigorous justification of this is given in Lemma 2.8. We are able to show that in fact performing a *single* stabilisation is sufficient.

THEOREM F (Theorem 4.47)**:** *Let T be a generating tuple of reflections for (W, S), then after performing one stabilisation, T is reflection equivalent to some stabilisation of S (note that the number of stabilisations is determined by #T and #S, in particular it will be #T + 1 − #S in this case).*

Theorems E and F show that in general, reflection generating tuples of Coxeter systems behave in the same way they do in the reasonably well-known case of dihedral groups, compare with Theorem 2.1.

0.1.2   Nielsen equivalence in right-angled Coxeter groups

There is, of course, more to study about reflection equivalence, for example giving an explicit description of the different equivalence classes of reflection generating tuples for certain (sub-classes of) Coxeter systems, but Theorem E goes a long way towards providing a comprehensive answer to Question 0.2. In Chapter 5 we turn to the first question, regarding Nielsen equivalence in general.

As we observed at the start of Section 0.1.1, in general, Coxeter systems are ill-suited to studying Nielsen equivalence in the associated Coxeter group since Coxeter groups do not have a canonically defined Coxeter system (even up to automorphism), and the Coxeter rank of a Coxeter group can be significantly larger than the algebraic rank. This poses a problem because almost all of the tools to study Coxeter groups depend on, or at least are influenced by, a choice of Coxeter system.

For this reason we focus on the case of *right-angled Coxeter groups* (RACGs). These are the Coxeter groups which admit a presentation of the form given in (1) in which all m_{ij}'s are either 2 or ∞, see Definition 1.5. Geometrically this means that all reflection hyperplanes are either parallel, or orthogonal. At first glance this appears to be a restriction to an extremely simple class of groups, however the class of these groups and their subgroups turns out to be a rich one with many important applications in geometric group theory and beyond, see for example

[57].

For our purposes, what makes RACGs nice to work with is that they are *rigid*, meaning that they do have a canonical choice of Coxeter system up to automorphism (and sometimes even up to inner automorphism, see Theorem 2.10); and also their Coxeter rank coincides with their algebraic rank, which can be seen easily by mapping to their abelianisation. Thus a Coxeter system (W, S) for a RACG gives a somewhat canonical minimal generating tuple for W.

Returning briefly to reflection generating tuples with respect to a choice of Coxeter system (W, S), as with Weyl groups it is possible to conclude from Theorem E that every reflection generating tuple of W is reflection (and hence Nielsen) equivalent to some stabilisation of S.

In fact we expand our remit to include *quasiconvex* subgroups of RACGs, subgroups which sit inside their host group nicely with respect to the geometry of the group (in particular, the geometry of the Davis complex of (W, S)), see Definition 5.11.

Again our approach is geometrical in flavour, but instead of working in the Davis complex we manipulate certain cube complexes whose edges are labelled by elements of S. Given a finite generating tuple X for a quasiconvex subgroup G of W, we can build an S labelled rose graph Ω_X (ie 1-dimensional cube complex) by subdividing and labelling each petal according to the elements of S.

We make heavy use of the work of Pallavi Dani and Ivan Levcovitz [31], who construct finite sequences of cube complexes to G starting with Ω_X and ending with a so-called *completion* of G, $\widehat{\Omega}_X$, see Definition 5.3. This completion, or at least a certain core subgraph, is uniquely determined by G, independent of the choice of X. This immediately gives an algorithm to check if a given tuple of elements generates a certain quasiconvex subgroup in parallel with Theorem D, see Theorem 5.14.

There are two natural ways to interpret the labelled cube complexes Ω used to construct a completion, one geometric, and the other algebraic. Geometrically, one can take their universal cover $\widetilde{\Omega}$ which maps G-equivariantly into the Davis complex Σ of (W, S). The procedure building the completion applies operations

which make the map $\widetilde{\Omega} \to \Sigma$ closer to an injection whose image is convex. Algebraically, the labelling defines a map from the fundamental group of Ω to $G \leqslant W$, and the moves can be interpreted as performing sequences of *Tits moves* (algebraic manipulations of words using the relations in (1), see Definition 1.10) on words representing the elements of G.

Notice that in the case that Ω has free fundamental group, this second interpretation gives a marking of G. Since Ω_X is a graph, we start with a space with free fundamental group. Constructing a completion sequence as before, but making sure never to create a cube complex with non-free fundamental group, gives a sequence of markings which differ by an automorphism (or possibly a surjection) of free groups. This corresponds to a sequence of Nielsen equivalences (or possibly *reductions*, which are the inverse operation of stabilisation, see Definition 1.17) between generating tuples for G.

Even though the completion sequences defined in [31] are finite for quasiconvex subgroups, it is not at all clear that these *free* completion sequences are finite. The way we gain control over free completion sequences is by carefully constructing them in parallel with non-free completion sequences. In this way we can prove the following.

Theorem G (Theorem 5.23)**:** *Let X be a finite tuple of elements in a RACG (W,S) Then the following are equivalent:*

1. *X produces a finite free completion sequence $\Omega_X \to \cdots$*

2. *X produces a finite non-free completion sequence $\Omega_X \to \cdots$*

3. *X generates a quasiconvex subgroup of W*

This gives a practical and algorithmic method to simplify a given generating tuple for G. It also gives a test for when a generating tuple of W is Nielsen equivalent to a standard generating tuple.

Theorem H (Theorem 5.30)**:** *Let $\widehat{\Omega}_X^{free}$ be a standard free completion of the rose graph Ω_X associated to some finite generating tuple of W, X. If $\widehat{\Omega}_X^{free}$ retracts onto a graph then X is Nielsen equivalent to a stabilisation of S.*

This falls short of proving that all generating tuples of W are Nielsen equivalent to stabilisations of each other. Nevertheless, we conjecture that this is the case.

Conjecture I (Conjecture 5.33): *Let (W, S) be a Coxeter system for a RACG, then every generating tuple for W is Nielsen equivalent to a stabilisation of S.*

The observation that all *reflection* generating tuples are Nielsen equivalent to some stabilisation of S lends some credence to this claim. We have also implemented free completion sequences in Mathematica [102] and verified the conjecture on many randomly generated generating tuples of a few different RACGs.

Two possible approaches to proving this Conjecture follow from our work. The first is to show that any generating tuple is Nielsen equivalent to some tuple of reflections. The second is to prove that any free completion coming from a generating tuple retracts onto a graph.

0.2   A geometric approach to group equivariant machine learning

In Part II we discuss a novel approach to supervised group equivariant machine learning using classical ideas from geometric topology. A very broad class of machine learning problems can be phrased mathematically as follows. Given a continuous function $\alpha \colon X \to Y$ between connected manifolds X and Y, we want to approximate α as well as possible (with respect to some suitable norm on the space of continuous functions $X \to Y$) in a class of functions $\mathfrak{M}$ which can be easily handled by a computer. The choice of this class $\mathfrak{M}$ is called the *machine learning model* or *architecture*, and the method to find a function $\beta \in \mathfrak{M}$ which closely approximates α is called the *machine learning algorithm*. A machine learning algorithm is *supervised* if it involves taking some large (finite) sample $X_0 \subset X$ on which the function α is known, and using the data $D_{\text{train}} = \{(x, \alpha(x)) \mid x \in X_0\}$ to train the machine learning model.

A classic example of a supervised machine learning model and algorithm is a

neural network, which assumes that X and Y are real vector spaces and takes $\mathfrak{M}$ to be (roughly speaking) the class of piece-wise linear maps $X \to Y$. Then given the training data D_{train} it evaluates the difference between the current choice of β and α on X_0 and uses a version of the gradient descent algorithm with respect to this cost function to modify the linear parts of β to decrease the cost.

We approach the problem of finding more accurate and efficient machine learning models and algorithms in the more general case that X and Y are Riemannian manifolds, and there is a group G which acts on X and Y discretely by isometries such that the map α we want to approximate is equivariant with respect to these G actions. Our main approach involves a data pre-processing set where we define a G-invariant map $\pi \colon X \to X$. G-invariance implies that $\pi(x)$ lies in the same orbit as x, so there is a function $\phi \colon X \to G$ such that $\pi(x) = \phi(x) \cdot x$. We then replace the training data by $D_{\text{train}}^{\pi} = \{(\pi(x), \phi(x) \cdot \alpha(x)) \mid x \in X_0\}$ (see Figure 7.1). Since this only changes the input data, we can use this new training data as input for any machine learning algorithm and model, giving our approach a lot more flexibility than many of the current approaches which work only for neural networks, say, see for example [116, 59, 81, 25, 27].

Suppose we have trained a model $\overline{\beta} \colon \pi(X) \to Y$ on the data D_{train}^{π}, then we can turn this into a G-equivariant map $\beta \colon X \to Y$ by defining

$$\beta(x) = \phi(x)^{-1} \cdot \overline{\beta}(\pi(x)) = \phi(x)^{-1} \cdot \overline{\beta}(\phi(x) \cdot x).$$

The problem now comes in choosing and computing the G-invariant map π. In order that π should respect the geometry of X and the action of G by isometries, we propose choosing π to be a *projection onto (the closure of) a fundamental domain* $\mathcal{F}$ for G acting on X. We give two methods of finding such a projection. The first is based on the idea of a *Dirichlet fundamental domain*: let $x_0 \in X$ be a point whose stabiliser lies in the kernel of the action of G on X, and define

$$\overline{\mathcal{F}} = \{x \in X \mid d(x_0, x) \leqslant d(x_0, g \cdot x) \; \forall g \in G\},$$

where d is the metric on X induced by the Riemannian metric; then the fundamental domain $\mathcal{F}$ is the interior of this set, see Definition 9.1. Since $\mathcal{F}$ is defined

via a minimisation problem with respect to the metric d, a natural way to compute (or in reality approximate) $\pi\colon X \to \overline{\mathcal{F}}$ is to perform discrete gradient descent in the Cayley graph of G with respect to some choice of generating set S.

The second method is more specialised in its application. Assume $X = \mathbb{R}^n$, on which the symmetric group S_n acts by permuting the coordinates, and that G acts on X via a map $G \to S_n$ onto some subgroup of S_n. We give an explicit combinatorial description in Section 8.1.1 of a projection $\pi\colon X \to X$ onto a fundamental domain.

This is based on [36], in which an explicit description of *right transversals* for subgroups of S_n is given, ie sets of unique right coset representatives, see Definition 8.21. Notice that the action of S_n on X is essentially the same as its action on its Davis (or Coxeter) complex if we think of S_n as a Weyl group of type A. We can use the theory of chambers and galleries, discussed in [16], to modify the description of transversals mention above to give a transversal which corresponds to a fundamental domain. Using this, we can describe an explicit projection map.

To the best of our knowledge, this is the first explicit description of fundamental domains for arbitrary permutation groups acting on $\mathbb{R}^n$, and also of a projection map onto the fundamental domain. In fact our procedure depends on a choice of a *base* for G, an ordered subset B of $\{1, 2, \ldots, n\}$, such that the point-wise stabiliser of B in G is trivial, see Definition 8.2. A base always exists since for any G one can take $B = (1, 2, \ldots, n-1)$. Different choices for B lead to different fundamental domains, so our method can be quite flexible. We also give several variations of the method which may be better suited for different machine learning applications. In Section 8.2 we use the representation of subgroups of S_n in [67] to give an implementable algorithm for computing π and analyse the complexity of this algorithm.

THEOREM J (Theorems 8.10, 8.13 and 8.15): *Given a subgroup G of S_n and a choice of base B of size $k < n$, there is an algorithm which can compute $\pi(x)$ for $x \in X$ in two steps. The first one-off step computes initial data (which does not depend on x), and takes $O(k^2 n^3)$ time and $O(n^2 \log n)$ space. The second step is applied to each x, and takes $O(k^2 n^2)$ time and $O(n^2 \log n)$ space.*

If G is a so-called primitive *subgroup of S_n then we can compute an efficient base with a bound on the size k (depending only on n). Then the first step can be computed in $O(n^4 (\log \log n)^2)$ time, and the second step can be computed in $O(n^3 (\log \log n)^2)$ time. For some special subgroups, for example $G = S_n$, A_n, or $\mathbb{Z}_n$ then this drops to $O(n^2)$ time.*

We have implemented our approach for several example machine learning problems, and compared the accuracy of the resulting machine learning architectures with benchmarks from the literature. Our main example is that of computing the first Hodge number of complex *complete intersection Calabi-Yau* manifolds (CICYs), which have been studied in [61, 18, 17, 43]. These are varieties defined as follows. Consider the space $U = \mathbb{CP}^{n_1} \times \cdots \times \mathbb{CP}^{n_\ell}$, and a collection of k polynomials p_i defined on U, then the zero-set of each defines a hypersurface $H_i = \{z \in Z \mid p_i(z) = 0\}$. Write $\deg_i(p_j)$ for the degree of p_j in the coordinates on $\mathbb{CP}^{n_i}$. If

$$\sum_{j=1}^{k} \deg_i(p_j) = n_i + 1 \text{ for all } i,$$

and each hypersurface is non-degenerate, then the intersection $V = \bigcap_{j=1}^{k} H_j \subset U$ is a CICY, and V is determined up to diffeomorphism by the matrix $D(V) = \left(\deg_i(p_j) \right)_{ij}$.

Using the dataset in [54] of CICYs whose matrix $D(V)$ is of size up to 12×15, we compared our methods to others in the literature for predicting the first Hodge number $h^{1,1}(V)$ from $D(V)$. In this case, permuting the columns of $D(V)$ corresponds to permuting the polynomials p_i, and permuting the rows corresponds to reordering the factors in U. Neither of these changes the diffeomorphism type of V, so the problem is invariant under an action of $S_{12} \times S_{15}$ on $\mathbb{R}^{12} \otimes \mathbb{R}^{15}$. We found using a projection onto a Dirichlet fundamental domain yielded the highest accuracy, see Table 7.2. Our methods also lead to improvements over the baseline accuracies in the other applications we tried, see Tables 7.1 and 7.3.

In addition to these concrete experiments, we also present a unified framework for so-called *intrinsic* approaches to G-equivariant machine learning based on the universal property of quotient spaces. We discuss how other methods in the literature fit into this framework, and also compare these methods to ours

qualitatively. When the group G is very small, simpler methods such as *augmentation* are just as good, if not better than our approach, but for larger groups (say $S_{12} \times S_{15}$) several of these methods break down as they become computationally intractable. Our method has the advantages that it is efficient, works even for very large groups, and does not depend on a specific machine learning model, such as neural networks.

One might be concerned that these pre-processing steps impact the expressiveness of the machine learning model, however, at least in the case of neural networks and projections onto a fundamental domain we show that this is not the case by proving two versions of the *Universal Approximation Theorem*.

THEOREM K (Theorem 7.6): *Let $\pi\colon \mathbb{R}^n \to \mathbb{R}^n$ be a projection onto a fundamental domain and μ a finite measure on $\mathbb{R}^n$. Then any G-invariant function $\mathbb{R}^n \to \mathbb{R}^m$ which has finite L^p norm with respect to μ can be approximated arbitrarily well (with respect to the L^p norm) by $\beta = \overline{\beta} \circ \pi$ where $\overline{\beta}\colon \mathbb{R}^n \to \mathbb{R}^m$ is a neural network.*

THEOREM L (Theorem 7.8): *Let $\pi\colon \mathbb{R}^n \to \mathbb{R}^n$ be a projection onto a fundamental domain $\overline{\mathcal{F}}$ and X a G-invariant compact subset of $\mathbb{R}^n$ such that $\pi(X) \subset \mathcal{F}$. Then any G-invariant continuous function $X \to \mathbb{R}^m$ can be approximated arbitrarily well (with respect to the supremum norm) by $\beta = \overline{\beta} \circ \pi$ where $\overline{\beta}\colon \mathbb{R}^n \to \mathbb{R}^m$ is a neural network.*

0.3   Structure of the thesis

In Chapter 1 we introduce some of the background on Coxeter group and Nielsen equivalence separately which is then used throughout the rest of Part I. In the first Section of Chapter 2 we give a more specific overview of Nielsen equivalence in the context of Coxeter groups and introduce the main questions we seek to answer. In the second Section of that Chapter we discuss invariants for Nielsen equivalence in Coxeter groups. This Section does not play a significant role in the rest of the thesis.

The three main Chapters then go on to discuss Weyl groups and quiver mutations (Chapter 3), reflection equivalence (Chapter 4), and Nielsen equivalence

Chapter 0

Part I

Chapter 1

Chapter 2
2.1 2.2

Chapter 3   Chapter 4   Chapter 5
              4.5

Part II

Chapter 6

Chapter 7
7.4
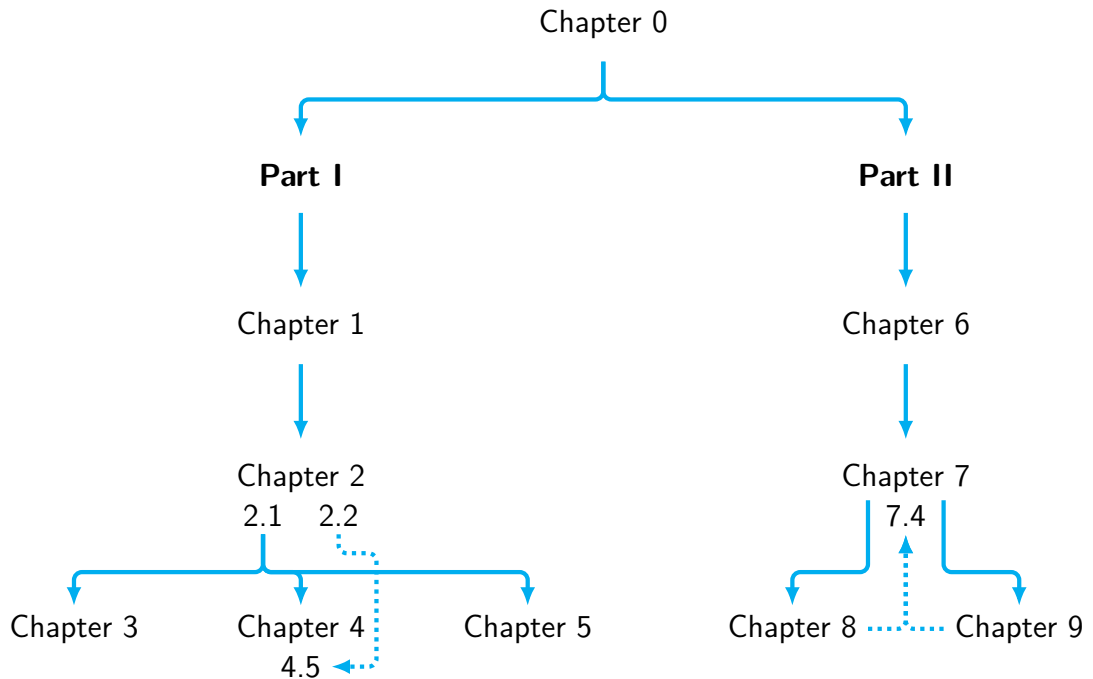
Chapter 8 ⋯⋯ Chapter 9

Figure 1: Structure and dependencies of the Chapters in this thesis.

in right-angled Coxeter groups (Chapter 5). Each of these Chapters can be read more or less independently of one another.

In Part II we introduce the background on equivariant machine learning in Chapter 6. Our work falls into two parts: first a general mathematical approach to the problem discussed in Chapter 7 followed by specific approaches and algorithms which realise this approach in certain settings. These are discussed in Chapters 8 and 9 which can be read independently of each other.

Part I:

Nielsen equivalence in Coxeter groups

By abandoning…the square and rectangle as a basis for composition, we have increased the possibilities for invention of all kinds.

Carmelo Arden Quinn (1913–2010)

Uruguayan artist and co-founder of the Madí art movement

Chapter 1

Background

IN THIS CHAPTER WE PROVIDE an overview of some background on Nielsen equivalence and Coxeter groups. In the first Section we discuss Coxeter groups, starting with their definition and some of the basic terminology and theory. As part of this we review two well-known solutions to the word problem and summarise the key ideas in [31]. In this paper Pallavi Dani and Ivan Levcovitz introduce certain sequences of labelled cube complexes which can be associated to subgroups of *right-angled* Coxeter groups. We modify these in Chapter 5 to study Nielsen equivalence.

In Section 1.2, we introduce Nielsen equivalence itself in terms of both generating tuples of elements of a group and markings of that group. We outline the history of the study of this topic, finishing with a heuristic overview of the topological approach to studying Nielsen equivalence inspired by John Stallings work on graphs and free groups via folds.

In the final Section we summarise what may at first appear to be an unrelated topic: quiver mutations. Arising out of the fields of combinatorics and representation theory in the context of cluster algebras, this has a somewhat mysterious connection with certain finite Coxeter groups called Weyl groups. This forms the motivation for our study of a variation of Nielsen equivalence, *reflection equivalence*, as well as an application.

1.1   Coxeter groups

Coxeter groups are a very important class of finitely generated groups possessing both a rich combinatorial and geometric theory. Named after Harold SM Coxeter who classified the finite Coxeter groups [30], the credit for studying them in general goes to Jacques Tits in the 1960s [106, 107]. Certain finite Coxeter groups called Weyl groups play a central role in the theory and classification of Lie groups and algebras, which is why Coxeter groups are pivotal in Bourbaki's treatment of this subject [12]. Tits was principally interested in Coxeter groups because of their importance in the construction of *Tits buildings*—spaces constructed to study exceptional groups of Lie type [1]. A good introduction to the geometric theory of Coxeter groups can be found in Chapters I–III of [16].

Coxeter groups generalise the idea of a reflection group. *Spherical, affine, and hyperbolic* type Coxeter groups are a rich source of examples of discrete groups acting on spaces of constant curvature. Since all isometries of spherical, Euclidean, and hyperbolic spaces can be decomposed as a sequence of reflections, many discrete groups of isometries acting on one of these spaces, especially in dimension 2, are isomorphic to a finite index subgroup of a suitable Coxeter group.

Coxeter groups are also very important from a combinatorial point of view. They admit a simple solution to the word problem due to Tits (see Section 1.1.2), which is the direct algebraic counterpart of the completion sequences we define in Section 5.1.3. They also admit a partial ordering called the *Bruhat order*, which is of interest in order theory [10] as well as in the study of *Schubert varieties*—certain algebraic subsets of flag manifolds indexed by elements of an associated Coxeter group, see for example [9, 20]. Closely related, Coxeter groups play an essential role in the theory of *Kazhdan-Lusztig polynomials* which connect to representation theory. A good reference for the combinatorial theory of Coxeter groups is [10]. There are also surprising connections [6] between Coxeter groups and the theory of cluster algebras [47, 48] which we exploit in Chapter 3.

Coxeter groups are also important in the field of geometric group theory. All Coxeter groups are *automatic* and it is believed they may all be *biautomatic*. Coxeter

groups are examples of CAT(0) groups, as they act on the *Davis complex* which, with its piece-wise Euclidean metric, is CAT(0) [32]. In the case of right-angled Coxeter groups, the Davis complex is a cube complex, and cube complexes associated with these groups are central to the work of Frédéric Haglund and Dani Wise on *special cube complexes* [57]. Coxeter groups are also very closely connected to *Artin groups* and work on Coxeter groups forms a foundation for much of the work on Artin groups.

1.1.1   Definition and basic properties

Most of the following is very well-known, for details consult for example [12, 16, 66]. The groups, which we refer to throughout most of the thesis as *Coxeter groups*, should more properly be called *Coxeter-Tits* groups. Harold SM Coxeter is rightly credited for classifying the finite Coxeter groups in 1935 [30], however the credit for initiating the systematic study of all Coxeter groups goes to Jacques Tits in the 1960s [106, 107].

Definition 1.1. Let W be a group, and suppose there is a finite subset of involutions $S = \{s_1, \ldots, s_n\}$ of W which generates the group. Define $m_{ii} = 1$ and for each $1 \leqslant i, j \leqslant n$ let m_{ij} be the order of $s_i s_j$ in W. Since each s_i is a different involution, it follows that $m_{ij} = m_{ji} \in \{2, 3, \ldots, \infty\}$. If W is equal to the group with presentation

$$\langle s_1, \ldots, s_n \mid s_i^2, \ (s_i s_j)^{m_{ij}} \text{ for all } 1 \leqslant i, j \leqslant n \rangle,$$

then W is called a **Coxeter group**. The presentation above is called a **Coxeter presentation** for W and the pair (W, S) is called a **Coxeter system**. Given a Coxeter group W, we call the minimum number of generators in a Coxeter system for W the **Coxeter rank** of W.

Conversely, given such a presentation, each of the in S represents a distinct element of W and has order 2. Note that neither the Coxeter system (W, S), nor the rank n, is an isomorphism invariant of W. It is often useful to summarise the data of a Coxeter system in the form of a graph. There are two standard conventions for this graph, and we employ both.

Definition 1.2. The **presentation diagram** of a Coxeter system (W, S) is the labelled graph $\Gamma = \Gamma(W, S)$ with vertex set S in which distinct vertices s_i and s_j are joined by an edge labelled m_{ij} if $m_{ij} < \infty$. By convention we omit edge labels equal to 2. The Coxeter group with presentation diagram Γ is written W_Γ.

On the other hand, the **Coxeter-Dynkin diagram** of (W, S) is also a labelled graph $\mathcal{V} = \mathcal{V}(W, S)$ which again has vertex set S and in which two distinct vertices s_i and s_j are joined by an edge labelled m_{ij} if $m_{ij} > 2$. This time the convention is that edges labelled 3 have their label omitted. The Coxeter group corresponding to the Coxeter-Dynkin diagram $\mathcal{V}$ is denoted by $W(\mathcal{V})$.

Note that either of these diagrams determine both the Coxeter group W, as well as a choice of Coxeter system. By writing W_Γ or $W(\mathcal{V})$, we are implying the systems $(W_\Gamma, V\Gamma)$ or $(W(\mathcal{V}), V\mathcal{V})$ respectively. In a presentation diagram for a Coxeter group, generators which do not appear together in any relations are not joined by an edge and so the connected components correspond to a maximal free-product decomposition of the group. On the other hand, in a Coxeter-Dynkin diagram, generators which commute are not connected by edges, so the connected components correspond to a direct product decomposition.

Definition 1.3. Let (W, S) be a Coxeter system, and $T \subset S$, then the subgroup of W generated by T is denoted W_T and is called a **special subgroup**. A conjugate of a special subgroup is called a **parabolic subgroup**.

It turns out that W_T is a Coxeter group with Coxeter system (W_T, T). In Section 1.1.2, we define a faithful representation of W through which the generators act by affine reflections on a real vector space and the action is discrete on a certain subset of that space.

Definition 1.4. Anticipating this geometric realisation of W, define the set

$$R = R(W, S) = \{wsw^{-1} \mid s \in S, \ w \in W\},$$

to be the set of **reflections** of the Coxeter system (W, S).

There are several special classes of Coxeter groups which we record here.

Definition 1.5. Let (W, S) be a Coxeter system, with Coxeter-Dynkin diagram $\mathcal{V}$, and numbers m_{ij} as defined previously. Then (W, S) is

1. **irreducible** if $\mathcal{V}$ is connected, ie (W, S) does not decompose as a direct product of special subgroups.

2. **spherical** if it is finite. The finite irreducible Coxeter systems have been classified. Their Coxeter-Dynkin diagrams are given in Table 1.1, and all spherical Coxeter systems are finite direct products to these groups.

3. a **Weyl group** if it is spherical and irreducible of type A, B, D, E, F, or G as shown in Table 1.1.

4. **affine** if it is infinite and acts discretely by isometries on some Euclidean space, and the generators S act by affine reflections.

5. **hyperbolic** if it is infinite and acts by isometries on some hyperbolic space, generated by reflections.

6. **even** if all m_{ij}'s are even or ∞ for $i \neq j$.

7. **right-angled** if $m_{ij} \in \{2, \infty\}$ for all $i \neq j$. We abbreviate right-angled Coxeter groups as RACGs from now on.

When trying to compute Nielsen equivalence invariants in Section 2.2, it is helpful to project a Coxeter group to its abelianisation. Fortunately, understanding W^{ab} is straightforward. Let (W, S) be a Coxeter system. Define an equivalence relation $\sim$ on S as the transitive closure of $s_i \sim s_j$ if m_{ij} is odd (including if $i = j$ and $m_{ii} = 1$). The following is straightforward to deduce from the Coxeter presentation of (W, S); alternatively, a proof may be found in Lemma 3.6 of [13].

PROPOSITION 1.6: *With notation as above, let k be the number of equivalence classes in S. Then $W^{ab} \cong \mathbb{Z}_2^k$ and for any pair of generators $s, s' \in S$ the following are equivalent.*

1. *They are in the same equivalence class*

2. *They have the same image in W^{ab}*

3. *They are conjugate in W*

A_n for $n \geqslant 1$

B_n for $n \geqslant 2$

D_n for $n \geqslant 4$

E_6

E_7

E_8

F_4

G_2

H_3

H_4

$I_2(m)$ for $m = 5$ or $m \geqslant 7$
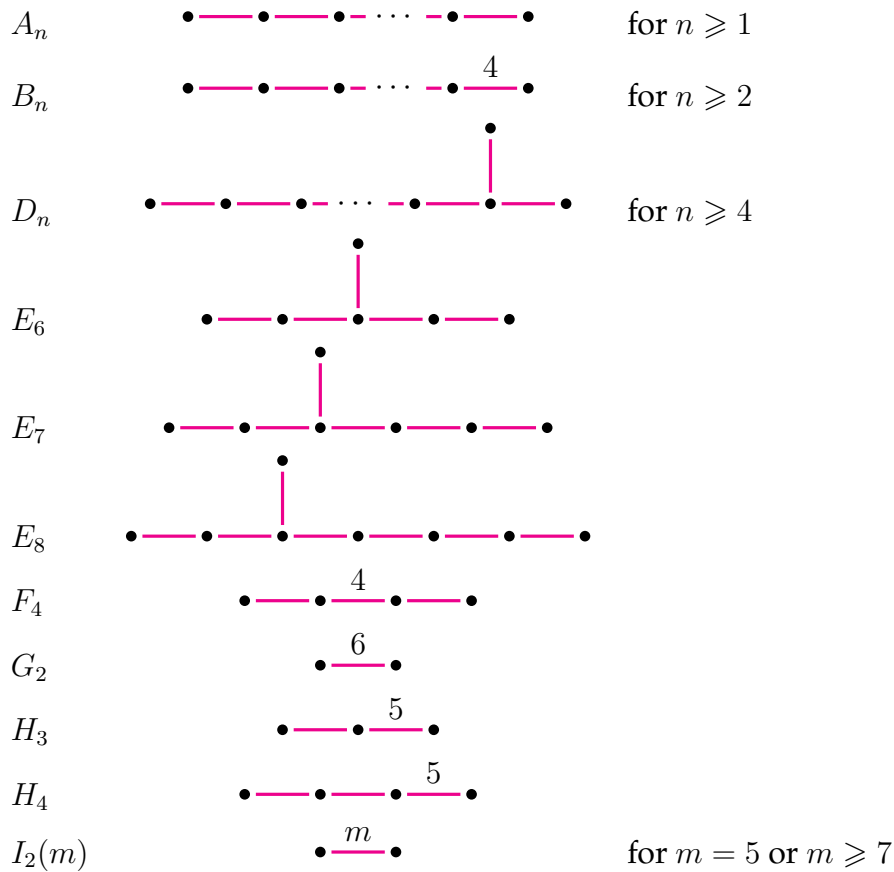
Table 1.1: Coxeter-Dynkin diagrams corresponding to finite irreducible Coxeter systems. The subscript number in the name of each indicates the rank.

It follows that W and W^{ab} have the same rank if and only if W is even. When we consider reflection equivalence in Chapter 4, understanding which generators are conjugate is useful.

1.1.2 The word problem

The word problem is a decision problem first introduced by Max Dehn in [33]. It asks whether, given a group G and a set of generators S for G, there is an algorithm to decide whether or not two words over $S \cup S^{-1}$ represent the same element in G. In general, the word problem is undecidable in groups, however, for Coxeter groups there are many solutions. Here we mention two: the first comes from a faithful linear representation of W, which is perhaps the most efficient and which we implement in [102]; the second is a combinatorial algorithm, which underpins the nice properties of the *completion sequences* which we use in Chapter 5, see

Theorem 1.11.

Definition 1.7. Let (W, S) be a Coxeter system of rank n, and let $\{e_1, \ldots, e_n\}$ be the standard basis of $\mathbb{R}^n$. Define a symmetric bilinear form B on $\mathbb{R}^n$ by

$$B(e_i, e_j) := -\cos\left(\frac{\pi}{m_{ij}}\right).$$

Define $\rho : W \to O_n(B)$ on the generators by $\rho(s_i)e_j = e_j - 2B(e_i, e_j)e_i$.

THEOREM 1.8 (Tits representation, see Section V.4 in [12]): *The map ρ defines a faithful representation of W on $\mathbb{R}^n$. The action preserves the form B and each generator s_i acts by an affine reflection in the hyperplane $H_{s_i} = \{v \in \mathbb{R}^n \mid B(e_i, v) = 0\}$. Write $H_{s_i}^+$ for the half-space with respect to H_{s_i} which contains e_i. Then the set $C_0 = \bigcap_{i=1}^n H_{s_i}^+$ (called the* fundamental chamber*) is open, connected, and non-empty. The action of W on the interior of the set $\bigcup_{w \in W} w\overline{C_0}$ (called the Tits cone), is discrete.*

Now two words $t_1 \cdots t_k$ and $t_1' \cdots t_{k'}'$ in the alphabet S represent the same element of W if and only if

$$\rho(t_1) \cdots \rho(t_k) = \rho(t_1') \cdots \rho(t_{k'}'). \tag{1.1}$$

This can be implemented into a computable algorithm as follows. First, if we assume that the representation is in fact over the ring of integers $\mathbb{Z}$, then the two sides of (1.1) can be computed exactly and compared. More generally, the representation will be over some finite field extension $\mathbb{Q}(a_1, \ldots, a_m)$ of $\mathbb{Q}$ generated by the entries of $\{\rho(s) \mid s \in S\}$. Moreover, there will be a bound on the size of the denominator of any rational coefficients (given in simplest form)—this is achieved by multiplying through by a suitable integer replaces these rational coefficients with integral ones. Overall then, it is possible to rephrase the problem over the polynomial ring $\mathbb{Z}[x_1, \ldots, x_m]$. This can be solved in a similar way to the case over $\mathbb{Z}$. For details, consult [75].

To talk about the combinatorial solution we need another definition.

Definition 1.9. Let (W, S) be a Coxeter system. Define the **length function** on W with respect to S, $\ell_S = \ell : W \to \mathbb{N}$ as the length of the shortest word over S which represents a given element. Such a shortest word is called a **reduced expression**.

Given a word $t_1 \cdots t_k$ over S, it is always possible to manipulate it by using the relations in the Coxeter presentation of (W, S) to turn it into a reduced expression representing the same element.

Definition 1.10. Given a word $t_1 \cdots t_k$ over S, a **Tits move** is one of the following rewriting procedures:

M1) Replace a subword of the form $\overbrace{s_i s_j s_i \cdots}^{\text{length } m_{ij}}$ by $\overbrace{s_j s_i s_j \cdots}^{\text{length } m_{ij}}$.

M2) Delete a subword of the form $s_i s_i$ for some $s_i \in S$.

Notice that neither of these moves increases the length of the word.

THEOREM 1.11 (Theorem 3 in [107])**:** *Given any word over S representing an element $w \in W$, there is a finite sequence of Tits moves which turns the word into a reduced expression for w. Moreover, all reduced expressions for w can be obtained from this reduced expression by a sequence of M1 moves (which do not change the length of the word).*

The combinatorial algorithm which solves the word problem involves performing all possible (M1) moves on a given word. If at some point an (M2) move can be performed, this is used to shorten the word. After finitely many steps, no more (M2) moves will be possible, and one ends up with a finite list of all minimal length words representing the same element as the starting word. Repeating this process for a second word, one can compare reduced expressions—the words represent the same element if and only if their lists of reduced expressions are the same. There is a particularly strong version of this Theorem which holds for right-angled Coxeter groups (RACGs).

COROLLARY 1.12 (Proposition 2.2 in [31])**:** *Let (W, S) be a right-angled Coxeter system, and $t_1 \cdots t_k$ a non-reduced expression over S. Then there are indices $i \neq j$ such that $t_i = t_j$ and $t_{i+1}, \ldots, t_{j-1}$ all commute with $t_i = t_j$. Hence $t_1 \cdots t_{i-1} t_{i+1} \cdots t_{j-1} t_{j+1} \cdots t_k$ is a shorter word representing w.*

Finally, we will record a special property of spherical Coxeter groups with respect to the length function.

Proposition 1.13 (Exercise 22 on page 40 of [12])**:** *Let (W, S) be a spherical Coxeter system with length function ℓ. Then W contains an element w_0 such that $\ell(w_0) \geqslant \ell(w)$ for all $w \in W$. This element is unique, an involution, and conjugation by w_0 permutes the elements of S.*

1.2   Nielsen equivalence

The problem of Nielsen equivalence concerns classifying tuples of generators of groups and has been studied in combinatorial group theory throughout the second half of the last century. Henceforth, we work with generating *tuples* of groups rather than generating *sets*. Nielsen equivalence is named after Jakob Nielsen who in 1924 proved the following result about free groups [91].

Theorem 1.14 (Nielsen's Theorem)**:** *Let $\mathbb{F}_n$ be the free group of rank n generated by the tuple $(x_1, \dots, x_n)$. Then the group of automorphisms $\mathrm{Aut}(\mathbb{F}_n)$ is finitely generate by the following automorphisms:*

  T1) $x_i \mapsto x_{\sigma(i)}$, for some permutation $\sigma \in S_n$

  T2) $x_i \mapsto x_i^{-1}$, for some fixed i, and all other generators are unchanged

  T3) $x_i \mapsto x_i x_j$, for some fixed $i \neq j$, and all other generators are unchanged

Definition 1.15. Each of the automorphisms (T1)–(T3) is called an **elementary Nielsen transformation**. Any finite sequence of elementary Nielsen transformations (ie an automorphism on $\mathbb{F}_n$) is called a **Nielsen transformation**.

By the universal property of free groups, given a finitely generated group G and a generating tuple $S = (s_1, \dots, s_n)$, there is a unique surjective homomorphism $\phi : \mathbb{F}_n \to G$ such that $\phi(x_i) = s_i$.

Definition 1.16. A surjective homomorphism $\phi \colon \mathbb{F}_n \to G$ from a free group to a group G is called a **marking** of G. The image of the standard generating tuple of $\mathbb{F}_n$ is the generating tuple this marking **represents**.

Elementary Nielsen transformations of $X = (x_1, \dots, x_n)$ descend in the obvious way to transformations of $S = \phi(X)$. These are also called elementary Nielsen

transformations. We will talk about Nielsen equivalence of X and S interchangeably.

Definition 1.17. Two generating tuples S and S' for a group G are **Nielsen equivalent** if there is a Nielsen transformation (ie a sequence of elementary Nielsen transformations) which turns S into S'. It follows from Nielsen's Theorem that this defines an equivalence relation on the set of generating tuples of G of fixed size n.

If we replace a generating tuple $S = (s_1, \ldots, s_n)$ of G by replacing $(s_1, \ldots, s_n, 1)$ we will say we have **performed a stabilisation** of S. More generally we will say that a generating tuple $S' = (s'_1, \ldots, s'_{n'})$ is a **stabilisation** of S if S' can be obtained from S by performing $(n' - n)$ stabilisations, possibly followed by an elementary Nielsen transformation of type (T1).

Conversely, a generating tuple of G is called **reducible** if it is Nielsen equivalent to one which contains the identity.

Remark 1.18 (Nielsen equivalence after stabilisations). Suppose G has *algebraic rank* k (ie the minimum number of elements needed to generate the group is k), and $S = (s_1, \ldots, s_n)$ and $S' = (s'_1, \ldots, s'_{n'})$ are generating tuples with $n' \leqslant n$. Then, after performing k stabilisations, S becomes automatically Nielsen equivalent to a stabilisation of S'. This is because there is a generating tuple $S'' = (s''_1, \ldots, s''_k)$ of size k and we can find Nielsen equivalences such that

$$
\begin{aligned}
S = (s_1, \ldots, s_n) \quad &\overset{\text{stabilise}}{\longmapsto} \quad (s_1, \ldots, s_n, \overbrace{1, \ldots, 1}^{k}) \\[2pt]
&\overset{\substack{\text{Nielsen} \\ \text{equivalence}}}{\longmapsto} \quad (s_1, \ldots, s_n, s''_1, \ldots, s''_k) \\[2pt]
&\overset{\substack{\text{Nielsen} \\ \text{equivalence}}}{\longmapsto} \quad (\overbrace{1, \ldots, 1}^{n}, s''_1, \ldots, s''_k) \\[2pt]
&\overset{\substack{\text{Nielsen} \\ \text{equivalence}}}{\longmapsto} \quad (s'_1, \ldots, s'_{n'}, \overbrace{1, \ldots, 1}^{n-n'}, s''_1, \ldots, s''_k) \\[2pt]
&\overset{\substack{\text{Nielsen} \\ \text{equivalence}}}{\longmapsto} \quad (s'_1, \ldots, s'_{n'}, \overbrace{1, \ldots, 1}^{n-n'+k}) \\[2pt]
&\overset{\text{stabilise}}{\longleftarrow} \quad (s'_1, \ldots, s'_{n'}) = S'.
\end{aligned}
$$

Where we first express the generators in each of S, S', and S'' as words over each other, and then performing suitable (T2) and (T3) transformations to create or delete these words in the appropriate entries.

Notice that by the Nielsen's Theorem again, any Nielsen transformation of S lifts to an automorphism of $\mathbb{F}_n$, which gives an alternative definition of Nielsen equivalence. Let S and S' be two generating tuples of G of equal size, and let $\phi, \phi' : \mathbb{F}_n \to G$ be the two surjections coming from the universal property. Then S and S' are Nielsen equivalent if and only if there is an automorphism α of $\mathbb{F}_n$ such that $\phi = \phi' \circ \alpha$.

More generally, suppose S' has size $m < n$ and $\phi' : \mathbb{F}_m \to G$ is the corresponding surjection. If there is a surjective homomorphism $\alpha : \mathbb{F}_n \to \mathbb{F}_m$ such that $\phi = \phi' \circ \alpha$, then S is Nielsen equivalent to S' after $(n - m)$ stabilisations and, in particular, is reducible.

The goal in general is to classify all generating tuples of a given finitely generated group up to Nielsen equivalence. For example, in the free abelian group $\mathbb{Z}^k$, a generating tuple of size n can be represented as an $(n \times k)$ $\mathbb{Z}$-matrix. Elementary Nielsen transformations correspond to row operations. Moreover, putting this matrix in row echelon form shows that any generating tuple of $\mathbb{Z}^k$ is either reducible (if $n > k$) or Nielsen equivalent to that standard generating tuple. The general case for finitely generated abelian groups can be treated in a similar way using [35], yielding the following result.

THEOREM 1.19 (Theorem 1.1 in [94]): *Let G be a finitely generated abelian group, so that it can be decomposed as $\mathbb{Z}_{m_1} \times \cdots \times \mathbb{Z}_{m_n} \times \mathbb{Z}^r$ where $1 < m_{i+1} | m_i$ for $1 \leqslant i < n$. All generating tuples of fixed rank greater than $(n + r)$ are Nielsen equivalent, and every generating tuple of rank $(n + r)$ is Nielsen equivalent to exactly one generating tuple of the form*

$$(\overbrace{1, \ldots, 1, k}^{n \text{ terms}}, \overbrace{1, \ldots, 1}^{r \text{ terms}}),$$

where $1 \leqslant k \leqslant m_n/2$ and $\gcd(k, m_n) = 1$.

1.2.1   Previous work on Nielsen equivalence

The most important early result on Nielsen equivalence is Grushko's Theorem
[55] proved in 1940, which treats the case of free products of groups. The original
proof is very combinatorial, however John Stallings later gave a short geometric
proof of the same result [104].

Theorem 1.20 (Grushko's Theorem)**:** *Let $G = A * B$ be a free product, then any gen-
erating tuple for G is Nielsen equivalent to one of the form $(s_1, \ldots, s_n, s'_1, \ldots, s'_m)$, where
$(s_1, \ldots, s_n)$ generates A, and $(s'_1, \ldots, s'_m)$ generates B.*

Nielsen equivalence has been studied in earnest since at least the 1960s. Until
the millennium methods tended to be combinatorial. For example, in [38], Martin
Dunwoody showed that any tuple of $n + 1$ generators of a finite solvable group of
rank n is reducible. Certain examples of rank 2, one-relator groups were studied
in [28, 64].

One class of groups which have been very well-studied are *Fuchsian groups*, ie
groups which nicely on $\mathbb{H}^2$, see the Definition below. In Theorem 6 of [117], Heiner
Zieschang proved that all minimal generating tuples of closed surface groups of
genus greater than 3 are Nielsen equivalent. Starting with work of Gerhard Rosen-
berger in the 1970s (for example [100, 99]), the effort to classify all generating tu-
ples of arbitrary Fuchsian groups has been the subject of several papers of Martin
Lustig and Yoav Moriah beginning with [77, 80]. In this effort they developed
a heavy-duty Nielsen equivalence invariant which is K-theoretic and based on
Reidemister-Whitehead torsion [78]. The state of the art is presented in [79]. For the
purpose of comparing with the case of Coxeter groups later, we state their main
result—for simplicity we restrict to the genus 0 case.

Definition 1.21. A (genus 0) **Fuchsian group** is a group G which admits a presen-
tation for the form

$$\langle s_1, \ldots, s_\ell \mid s_1^{\gamma_1}, \ldots, s_\ell^{\gamma_\ell}, s_1 s_2 \cdots s_\ell \rangle,$$

where $\gamma_i \geqslant 2$ are called the **exponents** of G. We can give an equivalent geometric
definition of (genus 0) Fuchsian groups as groups which act properly discontin-

uously and co-compactly by orientation preserving isometries on the hyperbolic plane $\mathbb{H}^2$ and are generated by a set of finite order elements.

To see the equivalence of these definitions, let $P \subset \mathbb{H}^2$ be a fundamental domain for G acting on $\mathbb{H}^2$. P is a convex finite sided polygon and we can label the vertices cyclically by $p_1, \ldots, p_\ell$. Proper discontinuity of the action implies that the internal angle at each of the vertices p_i are of the form $2\pi/\gamma_i$ for some $\gamma_i \geqslant 2$ an integer. Then G is generated by the set of rotations $\{s_i \mid 1 \leqslant i \leqslant \ell\}$, where s_i fixes p_i and rotates $\mathbb{H}^2$ by $2\pi/\gamma_i$. This produces the presentation given above.

The exponents turn out to be isomorphism invariants for Fuchsian groups. The relation $s_1 s_2 \cdots s_l$ allows us to write any one of the generators in terms of the others, and so any tuple of the form

$$U = \left(s_1^{u_1}, \ldots, s_{j-1}^{u_{j-1}}, s_{j+1}^{u_{j+1}}, \ldots, s_\ell^{u_\ell} \right)$$

generates G, where $\gcd(u_i, \gamma_i) = 1$ for all i. Outside a small class of exceptional cases, such a generating tuple is minimal; moreover, Rosenberger showed in [99] that for non-exceptional Fuchsian groups, every generating tuple of G is Nielsen equivalent to a tuple of this form. Then Lustig and Moriah proved the following.

Theorem 1.22 (Theorem 1.2 in [79])**:** *Let G be a Fuchsian group with presentation as above. Assume that if G has an even number of exponents equal to 2, then at least 5 exponents are greater than 2; and that if G has an odd number of exponents equal to 2, then at least 7 of the exponents are greater than 2. Given a second generating tuple*

$$V = \left(s_1^{v_1}, \ldots, s_{k-1}^{v_{k-1}}, s_{k+1}^{v_{k+1}}, \ldots, s_\ell^{v_\ell} \right)$$

with $\gcd(v_i, \gamma_i) = 1$ and j not necessarily equal to k, formerly define $u_j = v_k = 1$. Then U and V are Nielsen equivalent if and only if $u_i = \pm v_i \ (mod\ \gamma_i)$ for each $1 \leqslant i \leqslant \ell$.

In Section 2.2, we discuss the method they use to distinguish inequivalent generating tuples. In Chapter 4, we show that for minimal reflection generating tuples of Coxeter systems, an analogous result holds.

It is reasonable to ask how many Nielsen equivalence classes of generating tuples groups can have in general, given the well-known fact that up to *Tietze transformations*, all presentations of a finitely generated group are the same. Based on [93, 45], Martin Evans showed that, for any $n \geqslant 3$, there is a group G of rank n which has non-reducible generating tuples of size m, for any $m > n$ [44]. Considering just minimal rank generating tuples, the cyclic group $\mathbb{Z}_k$ has $\varphi(k)/2$ inequivalent generation tuples of size 1 for $k \geqslant 3$. Here φ is Euler's totient function, which is unbounded as k increases. The same thing happens for minimal generating tuples of Fuchsian groups [79]. There are even groups with infinitely many Nielsen equivalence classes of generating pairs, for example [89].

More recent approaches to Nielsen equivalence tend to use to more geometric methods, often inspired by Stallings' proof of Grushko's Theorem [104] and his seminal paper [105]. For example, in [76], Larsen Louder generalised Zieschang's work to show that any generating tuple of a closed surface group of genus greater than or equal to 2 is either reducible or Nielsen equivalent to the standard generating tuple. In a similar vein, Ederson Dutra has studied the case of Fuchsian groups by working with the orbifold $\mathbb{H}^2/G$ [39, 40].

Similar methods were employed by Ilya Kapovich and Richard Weidmann to show that in a generic small cancellation group, there are generating tuples which are *not* Nielsen equivalent after one stabilisation [69]. Moreover, in the special case of one-ended torsion-free δ-hyperbolic groups, there are generating tuples which are *not* Nielsen equivalent after $(n - 1)$ stabilisations, where n is the rank of the group [68] (compare with Remark 1.18).

They also considered groups G which act by isometries on a δ-hyperbolic metric space X. They proved that, except when G is both free and the orbit maps $G \to X \colon g \mapsto gx$ are quasi-isometric embeddings for all $x \in X$, the generating tuples of fixed rank are 'uniformly short' up to Nielsen equivalence [70]. This result was proved independently by Goulnara Arzhantseva [2]. Staying in the world of δ-hyperbolic groups, Spencer Dowdall and Samuel Taylor recently proved a Grushko-like result for hyperbolic extensions of hyperbolic groups [37]. Their work generalised a result of Juan Souto [103] studying Nielsen equivalence in the

fundamental groups of mapping tori of pseudo-Anosov maps of hyperbolic surfaces.

Nielsen equivalence has some perhaps surprising applications. Lustig and Moriah have applied their study of Fuchsian groups and Nielsen equivalence invariants to produce infinitely many examples of complexes which are homotopy, but not simply-homotopy, equivalent [77]; and to classify Heegaard splittings of Seifert fibred spaces up to isotopy and homeomorphisms [80]. In [24], William Chen used the cardinalities of Nielsen equivalence classes of generating pairs of finite groups to study solutions to the Markov equation, $x^2 + y^2 + z^2 - 3xyz = 0$ over the finite field $\mathbb{Z}_p$. As another application, Darryl McCullough characterises which groups have so-called *large* actions on orientable surfaces with a single orbit of points with non-trivial stabilisers. They are non-abelian two-generator groups, and the equivalence classes of these actions correspond to the Nielsen equivalence classes of generating pairs for the group which is acting, see [85].

Perhaps the most well-known application is to give practical algorithms for randomly sampling elements uniformly from a finite group. In principle, the position of a random walk on the Cayley graph of a finite group (with respect to some generating set) ends up being uniformly distributed after sufficiently many steps, but, in practice, this number of steps is too large to use. Instead, a more practical method involves starting with a generating tuple and performing a random sequence of (T3) Nielsen transformations, see [35] and references therein.

1.2.2 Stallings folds

In this Section, we translate the algebraic description of Nielsen transformation and reductions in terms of markings into a topological one. Let G be a finitely generated group, and declare a generating tuple for G, S, to be the *standard generating tuple*. Suppose we wanted to prove that all finite generating tuples for G are either reducible or Nielsen equivalent to the standard generating tuple S. Then given an arbitrary finite generating tuple X for G and corresponding marking $\phi_0 : \mathbb{F}(X) := \mathbb{F}_{n_0} \to G$, we need to construct a finite sequence of surjective homomorphisms $\{\alpha_i : \mathbb{F}_{n_i} \to \mathbb{F}_{n_{i+1}}\}_{i=0}^{N}$ between free groups which realise a sequence

of Nielsen transformations (if α_i is an isomorphism) and reductions (if $\mathbb{F}_{n_{i+1}}$ has strictly lower rank than $\mathbb{F}_{n_i}$), which transform X into S.

More precisely, each initial subsequence of the maps $\alpha_i \circ \cdots \circ \alpha_0$ induces a marking $\phi_i : \mathbb{F}_{n_{i+1}} \to G$ of G, by mapping $(\alpha_i \circ \cdots \circ \alpha_0)(x)$ to $\phi_0(x)$ for each $x \in X$. Then we need to construct the sequence so that ϕ_N is the marking representing S.

$$
\begin{array}{ccccccccc}
\mathbb{F}_{n_0} & \xrightarrow{\alpha_0} & \mathbb{F}_{n_1} & \xrightarrow{\alpha_1} & \cdots & \xrightarrow{\alpha_{i-1}} & \mathbb{F}_{n_i} & \xrightarrow{\alpha_i} & \cdots & \xrightarrow{\alpha_{N-1}} & \mathbb{F}_{n_N} \\
\phi_0 \downarrow & & \phi_1 \downarrow & & & & \phi_i \downarrow & & & & \phi_N \downarrow \\
G & \xrightarrow{=} & G & \xrightarrow{=} & \cdots & \xrightarrow{=} & G & \xrightarrow{=} & \cdots & \xrightarrow{=} & G
\end{array}
$$

To translate this into topology, we can replace each of the groups with cell complexes whose fundamental groups are the corresponding groups, and the homomorphisms by cellular maps which induce the corresponding homomorphisms on the level of fundamental groups.

$$
\begin{array}{ccccccccc}
\Omega_0 & \xrightarrow{f_0} & \Omega_1 & \xrightarrow{f_1} & \cdots & \xrightarrow{f_{i-1}} & \Omega_i & \xrightarrow{f_i} & \cdots & \xrightarrow{f_{N-1}} & \Omega_N \\
g_0 \downarrow & & g_1 \downarrow & & & & g_i \downarrow & & & & g_N \downarrow \\
\mathcal{O} & \xrightarrow{=} & \mathcal{O} & \xrightarrow{=} & \cdots & \xrightarrow{=} & \mathcal{O} & \xrightarrow{=} & \cdots & \xrightarrow{=} & \mathcal{O}
\end{array}
\tag{1.2}
$$

Typically we choose Ω_i to be either a graph, or a space which retracts onto a graph, so these is no homological obstruction to representing representing homomorphisms $\pi_1 \Omega_i \to \pi_1 \mathcal{O}$ by cellular maps. Since ϕ_N is the marking representing the generating tuple S, it makes sense to choose $\mathcal{O}$ to be the presentation complex associated to $\langle S \mid R \rangle$, and to aim to have g_N be the inclusion of the 1-skeleton into $\mathcal{O}$, ie Ω_N is the rose graph with edges corresponding to the elements of S.

By setting things up in this way, g_N is in some sense the simplest map of topological spaces which induces the surjection $\mathbb{F}(S) \to G$ on fundamental groups. Starting then with the map g_0 representing an arbitrary marking, the aim is to simplify this map as much as possible. It is not always possible to simplify the maps monotonically with respect to some measure of complexity, however (see for example [76]).

The easiest way to simplify the map $g_i : \Omega_i \to \mathcal{O}$ is if it fails to be locally injective. If it is not locally injective when restricted to the 1-skeleton at a vertex v, then there are two edges e and e' which meet at v and are mapped to the same

edge in $\mathcal{O}$. Letting f_i be the quotient map which identifies these edges, we get a new simpler map $\Omega_{i+1} = \Omega/e \sim e'$. As long as the fundamental group of Ω_{i+1} is still free, f_i represents a Nielsen transformation. Identifying edges in this way was introduced in [105] and is called a *Stallings fold*.

There are essentially four different possibly types of Stallings fold, depending on how the endpoints of e and e' are identified prior to the fold. These four cases are shown in Figure 5.3. Analysing precisely which of these types of fold occur in a given setting can be an extremely useful tool, see for example its application in studying the accessibility of finitely presented groups in [8].

$$\text{(a)} \qquad \text{(b)} \qquad \text{(c)} \qquad \text{(d)}$$
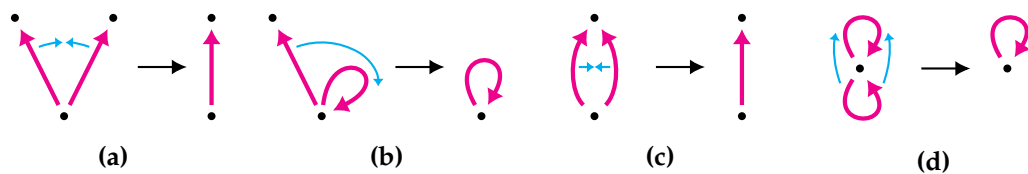
Figure 1.1: The four types of Stallings folds.

If g_i is locally injective on the 1-skeleton, but not on the whole of Ω_i, then it must identify higher dimensional cells and the idea of folding generalises to *cell identification*.

To keep track of the maps g_i, it helps to label the edges of $\mathcal{O}$ by S and orient them. Pulling back the labelling and orientation to Ω_i then records all of the information of the map g_i. Naturally, this general approach we have outlined is adapted to the individual case one is interested in. For example, when studying surfaces in [76], Louder lets $\mathcal{O}$ be a closed surface cellulated as a square complex (ie a 2-dimensional cube complex) with a so-called $\mathcal{VH}$-structure.

Chapter 2

Preliminaries

Studying Nielsen equivalence in Coxeter groups has the potential to open the door to studying other classes of related groups. For example, the theory of Artin groups (see [21] and references therein) often parallels that of Coxeter groups because any Coxeter group W_Γ with presentation diagram Γ can be viewed as a quotient of the corresponding Artin group A_Γ with the same presentation diagram. For this reason, classifying generating tuples in Coxeter groups provides an invariant for generating tuples of Artin groups via the surjection $A_\Gamma \to W_\Gamma$ (see Section 2.2). This correspondence is perhaps particularly true of RACGs and right-angled Artin groups. By the work of Haglund and Wise [57] this could also help in studying the fundamental groups of special cube-complexes.

We discuss several aspects of the problem of Nielsen equivalence in Coxeter groups including finding suitable invariants; the special case of generating tuples consisting purely of reflections; and how the completion sequences of Dani and Levcovitz can be applied to study Nielsen equivalence in RACGs. Prior to that, we begin our investigation by considering the case of rank 2 Coxeter groups, which is already well-known (the rank 1 case $\mathbb{Z}_2$ being trivial).

Theorem 2.1: *Let Dih_k for $k \geqslant 2$ be the dihedral group of order $2k$ which is isomorphic to the Coxeter group $W(I_2(k)) = \langle s_1, s_2 \mid s_1^2, s_2^2, (s_1 s_2)^k \rangle$. For concreteness, view Dih_k as the group of symmetries of a regular k-gon with s_1 and s_2 reflections in lines which meet at an angle π/k. Then any generating pair is Nielsen equivalent to (s_1, t) where t is a reflection such that $s_1 t$ is a rotation by $2\pi\ell/k$ for some $1 \leqslant \ell < k/2$ with $\gcd(\ell, k) = 1$*

(see Figure 2.1). Any two distinct such generating pairs are inequivalent. Any generating tuple of rank at least 3 is Nielsen equivalent to $(s_1, s_2, 1, \ldots, 1)$.
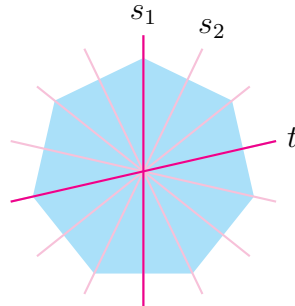
Figure 2.1: A generating pair for a dihedral group.

The first part of this follows from Proposition 4.5 in [78], but the whole statement is provable by elementary methods.

In Chapter 4, we prove that up to the natural notion of equivalence, all reflection generating tuples of arbitrary Coxeter systems satisfy the natural generalisation of this, see Theorems 4.44 and 4.47.

2.1 Ranks, rigidity, and reflection equivalence

The first problem which arises when studying generating tuples of Coxeter groups is that, in general, a Coxeter group W does not have a canonical choice of associated Coxeter system (W, S). There are examples (the first non-trivial example was given in [88], see Example 2.14) of Coxeter groups with two different Coxeter systems which have non-isomorphic diagrams. It is also possible for a Coxeter group to admit two Coxeter systems with different ranks. The most well known examples are the dihedral groups Dih_{2k} for k odd. In this case, two different presentation diagrams are given in Figure 2.2. A relatively straightforward way to determine all possible ranks of Coxeter systems for a given Coxeter group is given in [87]. This just requires looking at the presentation diagram for one of its Coxeter systems.

Nevertheless, if we fix the set of reflections in a Coxeter group, we can still maintain control on the *size* of Coxeter generating tuples of reflections, even if we

cannot control the corresponding presentation diagram.
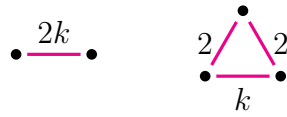
Figure 2.2: Two presentation diagrams corresponding to distinct Coxeter systems for the dihedral group Dih_{2k} where $k \geqslant 3$ is odd.

THEOREM 2.2 (Lemma 3.7 and Theorem 3.8 in [13]): *Let (W, S) be a Coxeter system, and $S' \subset R(W, S)$ a set of reflections such that (W, S') is another Coxeter system for W. Then $R(W, S') = R(W, S)$ and $\#S' = \#S$.*

For many Coxeter groups, the *algebraic rank* can be significantly less than the minimum rank of one of their Coxeter systems, their *Coxeter rank*. For example, Table 1.1 shows there are irreducible finite Coxeter systems with arbitrarily large ranks, but it has been shown that all of these groups have algebraic rank 2 [29]. It can be shown that this does not happen in certain classes of Coxeter groups. For example, rank n Coxeter groups with even Coxeter systems have abelianisation $\mathbb{Z}_2^n$ (see Proposition 1.6), therefore we can conclude that they also have algebraic rank n. More generally, in [19], the authors survey some classes of Coxeter groups where the algebraic rank is known, and prove that if a Coxeter system (W, S) has rank n and for all $i \neq j$, $m_{ij} \geqslant 6 \times 2^n$, then the algebraic rank of W is also n.

2.1.1   Nielsen and reflection equivalence

There can be significant a disparity between the rank of a Coxeter system and the algebraic rank of the corresponding Coxeter group, so it makes sense to separate the problem of Nielsen equivalence into two parts.

Question 2.3. For a Coxeter group W, when are two generating tuples of the same size Nielsen equivalent, and when is a non-minimal generating tuple reducible?

Question 2.4. For a Coxeter system (W, S) with set of reflections R (recall Definition 1.4), when are two generating tuples with elements in R the same up to a suitable notion of equivalence? If (W, S') is another Coxeter system for W such

that $R(W, S) = R(W, S')$, are S and S' equivalent to one another (by Theorem 2.2 S and S' must have the same cardinality)?

The reason why we do not use Nielsen equivalence in the second question is that it is the inappropriate notion of equivalence for reflection generating tuples. Nielsen transformations do not, in general, preserve the property of a generator being a reflection (applying a transformation of type (T3), for example, always yields a non-reflection). Instead we define a weaker notion of equivalence which preserves the set of reflections.

Definition 2.5. Let $(x_1, \ldots, x_n)$ be a generating tuple for the free group $\mathbb{F}_n$, then an **elementary partial conjugation** is a transformation of the form

T4) $x_i \mapsto x_j x_i x_j^{-1}$ for some fixed $i \neq j$, and all other generators unchanged.

A transformation of type (T4) can be expressed as a sequence of elementary Nielsen transformations of types (T2) and (T3), see Theorem 1.14. Therefore, it is an example of a Nielsen transformation which can be reinterpreted in terms of markings $\phi \colon \mathbb{F}_n \to W$. The induced transformation on the generators of W, in the case that $\phi(x_i) = s_i$ and $\phi(x_j) = s_j$ are reflections, can be rewritten as $s_i \mapsto s_j s_i s_j$ since s_j is an involution. By definition, this preserves the conjugacy class of s_i, and so transforms a generating tuple of reflections into another generating tuple of reflections.

In order to be able to deal with stabilisations and reductions of generating tuples of reflections we need to allow the identity 1 to appear as a generator and introduce one more transformation to communicate between reflections and 1 (since 1 is in a conjugacy class on its own).

Definition 2.6. Let (W, S) be a Coxeter system with set of reflections R. Consider the set of **reflection markings** $\phi \colon \mathbb{F}(x_1, \ldots, x_n) \to W$ such that $\phi(x_i) \in R \cup \{1\}$ for all $1 \leqslant i \leqslant n$. We call the tuple $(\phi(x_i))_i$ a **reflection generating tuple**. Consider the following very special case of a (T3) elementary Nielsen transformation.

T3*) $x_i \mapsto x_i x_j$ for some fixed $i \neq j$ such that either $\phi(x_i) = 1$ or $\phi(x_i) = \phi(x_j)$, and all other generators unchanged.

Two reflection markings are called **reflection equivalent** if one can be transformed into the other by a sequence of (T1), (T3*), and (T4) transformations.

Performing stabilisations on, and **stabilisations of**, reflection generating tuples are defined in the same way as in Definition 1.17. When working exclusively with reflection generating tuples, we call a reflection generating tuple **reducible** if it is reflection equivalent a stabilisation of some other reflection generating tuple.

We are interested in answering Question 2.4 with respect to reflection equivalence. We have already motivated studying this question in light of the importance of reflections in the study of Coxeter systems. To further motivate the definition of reflection equivalence: apart from being a straightforward weakening of Nielsen equivalence which preserves reflections, we have seen that it arises naturally in a completely independent context. Theorem 3.9 gives a way to transform certain reflection generating tuples of Weyl groups associated to quivers by mutations. These transformations are always compositions of partial conjugations and we study them in Chapter 3.

In Chapter 4 we discuss an algorithm to determine whether a given tuple of reflections from a Coxeter system (W, S) generates W. Here, however, we can state a necessary condition which also leads to a straightforward bound on the number of stabilisations which need to be performed on a reflection generating tuple to guarantee that it is reflection equivalent to a stabilisation of S.

LEMMA 2.7 (Follows from the proof of Lemma 6.4 in [111])**:** *Let (W, S) be a Coxeter system with $\#S = n$. If $T = (t_1, \ldots, t_\ell)$ is a finite tuple of reflections which generates W, then there is some permutation $\sigma \in S_\ell$ such that for $1 \leqslant i \leqslant n$, $t_{\sigma(i)}$ is conjugate to s_i.*

Now we can prove the following.

LEMMA 2.8: *Let (W, S) be a Coxeter system with $\#S = n$, and suppose $T = (t_1, \ldots, t_\ell)$ and $T' = (t'_1, \ldots, t'_{\ell'})$ are generating tuples with $\ell' \leqslant \ell$. Then, after performing n stabilisations, T becomes reflection equivalent to a stabilisation of T'.*

Proof. The proof follows the same pattern as Remark 1.18: starting with T we

perform n stabilisations. We need to show that there is a reflection equivalence

$$(t_1, \ldots, t_\ell, \overbrace{1, \ldots, 1}^{n}) \mapsto (t_1, \ldots, t_\ell, s_1, \ldots, s_n).$$

By Lemma 2.7, there is some $\sigma \in S_\ell$ such that for $1 \leqslant i \leqslant n$, $t_{\sigma(i)}$ is conjugate to s_i. By writing $t_{\sigma(i)} = w_i s_i w_i^{-1}$ for some $w_i \in W$, we can perform n (T3*) transformations to obtain

$$(t_1, \ldots, t_\ell, \overbrace{1, \ldots, 1}^{n}) \mapsto (t_1, \ldots, t_\ell, w_1 s_1 w_1^{-1}, \ldots, w_n s_n w_n^{-1}).$$

For $1 \leqslant i \leqslant n$, we can write w_i as a word over T. By performing a sequence of (T4) transformations on $w_i s_i w_i^{-1}$ according to the inverse of the word representing w_i, we are left with s_i. Repeating this for each i, this yields

$$(t_1, \ldots, t_\ell, w_1 s_1 w_1^{-1}, \ldots, w_n s_n w_n^{-1}) \longmapsto (t_1, \ldots, t_\ell, s_1, \ldots, s_n).$$

A similar argument works for each of the other steps in Remark 1.18. $\square$

Theorem 2.2 guarantees that n is the minimal number of stabilisations required for this kind of argument to work. Comparing this with Remark 1.18, the algebraic rank of a Coxeter group may often be significantly lower than the size of S, so, *a priori*, it seems much harder to make reflection generating tuples reflection equivalent via stabilisations. In fact, we show in Theorem 4.47 that one can do significantly better—in any Coxeter system a *single* stabilisation always suffices.

2.1.2   Rigidity of Coxeter groups

The most alluring situation in which to study the equivalence of generating tuples of Coxeter groups is when Questions 2.3 and 2.4 overlap, ie when the algebraic rank of W is attained by one of its Coxeter systems. Especially if W admits only one Coxeter system up to isomorphism. This brings in the notion of *rigidity* of Coxeter groups. A good introduction to this topic is [13], or for a detailed exploration consult [4].

Definition 2.9. Let W be a Coxeter group, and (W, S) a Coxeter system for W. The Coxeter *system* (W, S) is **reflection rigid** if every Coxeter generating tuple S' in $R(W, S)$ determines the same presentation diagram as S, and **strongly reflection rigid** if additionally for any such S', there is an inner automorphism of W sending S to S'.

On the other hand, the Coxeter *group* W is **rigid** if any two Coxeter systems for W, (W, S) and (W, S') determine the same presentation diagram (equivalently if S and S' differ by an automorphism of W), and **strongly rigid** if this automorphism can be chosen to be inner.

It follows immediately from the definition that if W is strongly rigid then all Coxeter generating tuples are reflection equivalent, since they all differ by an inner automorphism which lifts to an inner automorphism of the free group.

We have already remarked that Coxeter groups with even Coxeter systems have algebraic rank equal to the rank of any (and hence every) even Coxeter system. For the special case of RACGs, David Radcliffe proved that all RACGs are rigid [95]. The following simple criterion for which RACGs are strongly (reflection) rigid based on their presentation diagram was given by Noel Brady et al.

THEOREM 2.10 (Theorem 4.10 in [13])**:** *A RACG W_Γ is strongly reflection rigid if and only if for each vertex of Γ, the subgraph induced by all vertices not connected to that vertex, is connected. Moreover W_Γ is strongly rigid if and only if, in addition, each vertex is the intersection of all maximal complete subgraphs of Γ containing that vertex.*

We have already seen that, in general, finite Coxeter groups are not rigid (see Figure 2.2) however for finite Coxeter groups we have the following.

THEOREM 2.11 (Theorem 3.10 in [13])**:** *Let (W, S) be a Coxeter system for a finite Coxeter group W, then it is reflection rigid.*

In general, a Coxeter system need not be even reflection rigid and one way to see this is using so-called *digram twists*, first defined in [13] in Definition 4.4. Let Γ be the presentation diagram for a Coxeter system $(W, V\Gamma)$. Suppose S, T are disjoint subsets of $V\Gamma$ such that

1. W_S is spherical, and

2. Each vertex in $V\Gamma - (S \cup T)$ which is connected to a vertex of T is also connected to a vertex of S by an edge labelled 2.

Let w_0 be the longest element in W_S (see Proposition 1.13) and define a new generating tuple of reflections $\widetilde{V} = (\tilde{v})_{v \in V\Gamma}$ where $\tilde{v} = v$ if $v \in V\Gamma - T$, and otherwise $\tilde{v} = {}^{w_0}v$.

THEOREM 2.12 (Theorem 4.5 in [13]): *The pair $(W, \widetilde{V})$ is a Coxeter system. Its presentation diagram $\widetilde{\Gamma}$ can be obtained from Γ as follows:*

1. *For each edge in Γ joining a vertex s in S to a vertex t in T, add an edge with the same label between $\widetilde{w_0 s} = {}^{w_0}s$ and $\tilde{t} = {}^{w_0}t$ (recall from Proposition 1.13 that conjugation by w_0 permutes the elements of S).*

2. *For any other edge in Γ, which joins vertices v_1 and v_2 say, add an edge with the same label between $\tilde{v}_1$ and $\tilde{v}_2$.*

Definition 2.13. The presentation diagram $\widetilde{\Gamma}$ is said to be obtained from Γ by **twisting** T by w_0 around S, and the overall effect is called a **diagram twist**.

Example 2.14 ([88]). The presentation diagram on the right in Figure 2.3 is obtained by twisting T around S in the diagram on the left, and so represent the same Coxeter group.
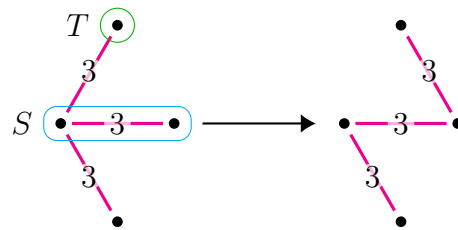
Figure 2.3: A diagram twist.

It follows immediately from the definition that, if Γ is a presentation diagram for a Coxeter group and $\widetilde{\Gamma}$ is obtained by a diagram twist, then $\widetilde{V}$ is reflection equivalent to $V\Gamma$. Therefore, if a Coxeter group is strongly rigid up to diagram twists (compare to Theorems 5.4 and 5.7 in [13]) then all its Coxeter generating tuples are reflection equivalent.

2.2 Invariants

There are two parts to any classification question. Constructing a sequence of Nielsen transformations, for example via a sequence of spaces as described in Section 1.2.2, proves that a given generating tuple is either reducible or Nielsen equivalent to the *standard generating tuple*, but cannot prove that a generating tuple is *not* reducible, or *not* equivalent to the standard generating tuple. To prove this requires the use of suitable invariants. Finding suitable invariants is hard in general. Here we mention a few known approaches to the problem which we can try to apply to the case of Coxeter groups.

2.2.1 Previous approaches to invariants

The first observation is that if $\psi : G \to H$ is a surjective homomorphism and Nielsen equivalence in H is already understood, then if the images of two generating tuples of G under ψ are inequivalent in H, then they must be inequivalent in G. The converse, however, does not hold: just because the images of two generating tuples in H are equivalent in H, does not mean the original generating tuples are equivalent in G. This approach is most commonly applied when H is the abelianisation of G and has the same rank, since Nielsen equivalence in finitely generated abelian groups is well-understood (see Theorem 1.19).

For groups of algebraic rank 2, a better option is called the *Higman invariant*. This is a strengthening of Higman's Lemma [90], which states that for a generating pair (x_1, x_2) of a group G, the order of the commutator $[x_1, x_2]$ is a Nielsen equivalence invariant.

Definition 2.15. The **extended conjugacy class** of an element $g \in G$, EC(g), the union of the conjugacy classes of g and g^{-1}.

Lemma 2.16 (Higman invariant)**:** *Then given a generating pair (x_1, x_2) of G, the extended conjugacy class of the commutator $[x_1, x_2]$ is a Nielsen equivalence invariant.*

Proof. This can easily be seen by applying the elementary Nielsen transformations:

T1) If $(x_1, x_2) \mapsto (x_2, x_1)$, then the inverse of $[x_2, x_1]$ is in $\mathrm{EC}([x_2, x_1])$ and

$$[x_2, x_1]^{-1} = \left(x_2 x_1 x_2^{-1} x_1^{-1}\right)^{-1} = x_1 x_2 x_1^{-1} x_2^{-1} = [x_1, x_2] \in \mathrm{EC}([x_1, x_2]).$$

T2) If $(x_1, x_2) \mapsto (x_1^{-1}, x_2)$, then the conjugate of the inverse of $[x_1^{-1}, x_2]$ by x_1 lies in $\mathrm{EC}([x_1^{-1}, x_2])$ and

$$\left([x_1^{-1}, x_2]^{-1}\right)^{x_1} = \left([x_1^{-1}, x_2]^{x_1}\right)^{-1}$$
$$= \left(x_1 x_1^{-1} x_2 x_1 x_2^{-1} x_1^{-1}\right)^{-1} = [x_1, x_2] \in \mathrm{EC}([x_1, x_2]).$$

T3) If $(x_1, x_2) \mapsto (x_1 x_2, x_2)$, then

$$[x_1 x_2, x_2] = x_1 x_2 x_2 x_2^{-1} x_1^{-1} x_2^{-1} = [x_1, x_2] \in \mathrm{EC}([x_1, x_2]). \qquad \square$$

A much more powerful—albeit computationally more complicated—invariant was developed by Lustig and Moriah in [78] based on Reidemister-Whitehead torsion. While the general invariant is K-theoretic, given a linear representation of a group a simplified version of their invariant based on the theory of Fox derivatives [49] can be used. They applied this in their study of Fuchsian groups [77, 80, 79]. In the remained of this Section we outline this simpler approach. We start by recalling the basics of Fox calculus.

Definition 2.17. Let G be a group, and $\mathbb{Z}G$ its integral group ring. Denote by $\varepsilon : \mathbb{Z}G \to \mathbb{Z}$ the **augmentation map**, ie the ring homomorphism induced by sending all elements of G to 1. A map $D : \mathbb{Z}G \to \mathbb{Z}G$ is called a **derivation** if it satisfies

$$D(u + v) = D(u) + D(v) \text{ and } D(uv) = D(u)\varepsilon(v) + uD(v) \text{ for all } u, v \in \mathbb{Z}G.$$

The derivations of $\mathbb{Z}G$ form a right $\mathbb{Z}G$-module. Ralph H Fox showed that if $G = \mathbb{F}_n$ is the free group generated by $X = (x_1, \dots, x_n)$, then this $\mathbb{Z}\mathbb{F}_n$-module is generated by the n derivations $\{\partial_{x_i} \mid 1 \leqslant i \leqslant n\}$, which are determined by their values on the generators of $\mathbb{F}_n$ by

$$\partial_{x_i} x_j = \delta_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise} \end{cases}.$$

An important property of derivations in the free group ring, from the point of view of Nielsen equivalence, is that they satisfy a version of the chain rule. Let

$Y = (y_1, \ldots, y_n)$ be Nielsen equivalent to X so that it is another free basis for $\mathbb{F}_n$. Then if we define the derivation $\partial_{y_i} y_j = \delta_{ij}$, these new derivations satisfy a *chain rule*

$$\partial_{x_i} u = \sum_{j=1}^{n} \partial_{y_j} u \cdot \partial_{x_i} y_j. \tag{2.1}$$

Definition 2.18. Given free bases X and Y of $\mathbb{F}_n$, define the **Jacobian matrix** to be $\partial_X Y := (\partial_{x_i} y_j)_{ij}$. In addition, for $u \in \mathbb{Z}\mathbb{F}_n$ we write $\partial_X u = (\partial_{x_i} u)_i$.

We can check is invertible over $\mathbb{Z}\mathbb{F}_n$ with inverse $\partial_Y X = (\partial_{y_i} x_j)_{ij}$:

$$\partial_X Y \partial_Y X = (\partial_{x_i} y_j)_{ij} (\partial_{y_i} x_j)_{ij} = \left(\sum_{k=1}^{n} \partial_{x_k} y_j \cdot \partial_{y_i} x_k \right)_{ij}$$

$$\overset{(2.1)}{=} (\partial_{y_i} y_j)_{ij} = (\delta_{ij})_{ij},$$

which is the identity. Moreover, with this notation the chain rule can be expressed by the matrix equation $\partial_X u = \partial_Y u \cdot \partial_X Y$. This Jacobian matrix captures the Nielsen transformation between X and Y.

Now let G have rank n, let $\langle S \mid R \rangle$ be the 'standard' presentation for G as chosen in Section 1.2.2 with $\#S = n$. In what follows, the distinction between a *word* over some generating tuple of G and the *group element* it represents is particularly important; we emphasise this by denoting words using bold letters, for example **t**. Let $\phi : \mathbb{F}_n \to G : x_i \mapsto s_i$ be the marking corresponding to S. This marking induces a surjective ring homomorphism $\mathbb{Z}\mathbb{F}_n \to \mathbb{Z}G$ which is also denoted ϕ. For a (freely reduced) word **t** over S, let $\tilde{t}$ be its lift to $\mathbb{F}_n$, and define

$$\partial_S \mathbf{t} := (\phi(\partial_{x_1} \tilde{t}), \ldots, \phi(\partial_{x_n} \tilde{t})).$$

More generally, if $T = (t_1, \ldots, t_n)$ is a generating tuple of G, and each generator t_i is represented by a word $\mathbf{t}_i$, then we get a Jacobian-like matrix $\partial_S \mathbf{T} = (\phi(\partial_{x_i} \tilde{t}_j))_{ij}$.

This matrix depends on the choice of words representing elements of T, but this dependence is controlled for by the *correction term* below. To extract a usable Nielsen equivalence invariant from $\partial_S \mathbf{T}$ we use a representation $\eta : \mathbb{Z}G \to \mathbb{M}_m(A)$ into a matrix ring over a commutative ring A. Notice that this representation induces a map $\mathbb{M}_n(\mathbb{Z}G) \to \mathbb{M}_{mn}(A)$ (where where we replace each entry in an element of $\mathbb{M}_n(\mathbb{Z}G)$ by an A-matrix, and then 'forget' this sub-matrix structure).

Definition 2.19. Let $A_G \leqslant A^*$ be the subgroup of the group of units of A which is generated by $\det(\eta(\pm s))$ for all $s \in S$ (since $\pm s$ is a unit in $\mathbb{Z}G$ and η is a ring homomorphism, $\eta(\pm s)$ must be a unit, and so have determinant in A^*).

Also define the **correction ideal** I_S^A to be the ideal of A generated as follows. For $R \subset \mathbb{F}_n$, the set of relations for G in the standard presentation, consider the set $\{\eta(\phi(\partial_{x_i} r)) \mid 1 \leqslant i \leqslant n \text{ and } r \in R\} \subset \mathbb{M}_m(A)$. Then I_S^A is generated by all entries in the matrices in this set.

Finally, denote by $\xi \colon A \to A/I_S^A$ be the quotient map, then $\xi(A_G)$ is a subgroup of the group of units of A/I_S^A.

THEOREM 2.20 (Straightforward generalisation of Corollary 2.10 in [79])**:** *With notation as above, define*

$$\chi_\eta \colon G^n \to A/I_S^A \colon T \mapsto \xi(\det(\eta(\phi(\partial_S \mathbf{T})))).$$

This function is well-defined in the sense that it does not depend on the choice of lift of T to $\mathbf{T} \in (\mathbb{F}_n)^n$. If T and S are Nielsen equivalent then $\chi_\eta(T) \in \xi(A_G)$.

Definition 2.21. We call the Nielsen equivalence invariant χ_η constructed in this way the **(m-dimensional) Lustig-Moriah invariant** associated to the representation $\eta \colon \mathbb{Z}G \to \mathbb{M}_m(A)$.

We briefly summarise how Lustig and Moriah apply this to study Fuchsian groups, which are closely related to two dimensional hyperbolic Coxeter groups. Recall from Definition 1.21 that a (genus 0) Fuchsian group is given by a presentation of the form

$$G = \langle s_1, \ldots, s_l \mid s_1^{\gamma_1}, \ldots, s_\ell^{\gamma_\ell}, s_1 \cdots s_\ell \rangle.$$

This group acts on the hyperbolic plane $\mathbb{H}^2$. They lift the standard representation $G \to \mathrm{PSL}_2(\mathbb{C})$ to $\rho \colon G \to \mathrm{SL}_2(\mathbb{C})$ (when this is possible) such that

$$\rho(s_1) = \begin{pmatrix} e^{2\pi i/\gamma_1} & 0 \\ 0 & e^{-2\pi i/\gamma_1} \end{pmatrix} \tag{2.2}$$

Then they 'mix' this with the quotient $G \to \mathbb{Z}_p = \langle t \mid t^p \rangle$ for some suitable $p \geqslant 3$ which divides both γ_1 and γ_2 (much of their work goes into arranging matters so

that this is possible). Then they define a representation $\eta : G \to \mathrm{SL}_2(\mathbb{C}\mathbb{Z}_p)$ over the complex group ring $A = \mathbb{C}\mathbb{Z}_p$ via

$$\eta(s_1) = \rho(s_1) \begin{pmatrix} t & 0 \\ 0 & t^{-1} \end{pmatrix}, \quad \eta(s_2) = \begin{pmatrix} t^{-1} & 0 \\ 0 & t \end{pmatrix},$$

and $\eta(s_i) = \rho(s_i)$ for each $3 \leqslant i \leqslant \ell$. Then they could apply Theorem 2.20. Instead of mapping directly onto the quotient ring by ξ, they define an element denoted $\Pi(a, b, r) \in \mathbb{C}\mathbb{Z}_p$ which is used to annihilate I_S^A; for details see the proof of Lemma 6.4 in [79].

2.2.2   Nielsen equivalence invariants for Coxeter groups

The example of dihedral groups shows that in general we cannot hope that all minimal generating tuples of an arbitrary Coxeter group are equivalent, therefore, an invariant to distinguish classes is required. The simplest approach is to map to a suitable quotient. The abelianisation of any Coxeter group is the direct product of some number of copies of $\mathbb{Z}_2$, see Proposition 1.6. This does not give a useful invariant since in these groups all generating tuples are reducible or equivalent to the standard one, as shown in Theorem 1.19.

The other natural candidates for quotients might be other Coxeter groups, and the only Coxeter groups we understand *a priori* are the dihedral groups, however Theorem 2.1 shows that these cannot give a useful invariant either, since all non-minimal generating tuples are Nielsen equivalent.

Some Coxeter groups whose Coxeter rank is greater than 2 nevertheless have algebraic rank 2, for example, all irreducible finite Coxeter groups have algebraic rank 2. Therefore one could try to apply the Higman invariant to study minimal generating sets of these groups.

As none of the elementary approaches to invariants work in general, we turn to those developed by Lustig and Moriah. The simplest choice of representation is to use $\mathbb{Z}W \to A$ where $A = \mathbb{Z}W^{\mathrm{ab}}$ or $\mathbb{Z}\mathbb{Z}_2$ (here we equate A with $\mathbb{M}_1(A)$). In the first case $A_W = \{\pm \overline{w}\}_{\overline{w} \in W^{\mathrm{ab}}}$ which equals $\mathbb{Z}_2 \times W^{\mathrm{ab}} = (\mathbb{Z}W^{\mathrm{ab}})^*$ by Theorem 6 in [63]; or in the second case $A_W = \{\pm 1, \pm \overline{s}\}$ where $\mathbb{Z}_2 = \{1, \overline{s}\}$.

2.2.3   The dihedral case

We can illustrate this invariant in the case of dihedral groups.

Notation 2.22. For elements g, h in a group, we denote the conjugate of g by h, hgh^{-1}, by $^h g$.

Let $W = W(I_2(k))$ with Coxeter generating pair $S = (s_1, s_2)$. There is a surjective homomorphism $W \to \{\pm 1\}$ induced by mapping S to -1. Any generating pair $T = (t_1, t_2)$ must contain an element which gets mapped to -1, which for dihedral groups must be a reflection. Possibly after a $(T3)$ transformation we can assume T is a pair of reflections, and after an overall conjugation, one of these is either s_1 or s_2.

Without loss of generality, let $t_1 = s_1$. Then we claim we can choose $t_2 = {}^w s_2$, where $w = (s_2 s_1)^p$ for some $0 \leqslant p < k/2$. Indeed, t_2 is a reflection so can be written as $^w s_1$ or $^w s_2$ for some $w \in W$. If k is odd then s_1 and s_2 are conjugate and $^w s_1$ can be rewritten as $^{w'} s_2$ for some w'. If k is even then W surjects onto $\mathbb{Z}_2 \times \mathbb{Z}_2$ by mapping s_1 and s_2 to the generators of the first and second factors respectively. If $T = (s_1, {}^w s_1)$ then the image of $\langle T \rangle$ under this surjection is $\mathbb{Z}_2 \times \{1\}$, so T does not generate W, a contradiction.

Considering $t_2 = {}^w s_2$ we may as well assume that $\ell(t_2) = 2\ell(w) + 1$, ie there is no cancellation, so that we can write $w = \cdots s_1 s_2 s_1$. If w has even length, then it has the form $(s_2 s_1)^p$ for some $p \geqslant 0$. If w has odd length then we can apply a Nielsen transformation to replace t_2 with $^{s_2} t_2 = {}^{s_2 w} s_2$, and $s_2 w = (s_2 s_1)^p$ for some $p \geqslant 0$. If $p \geqslant k/2$ then it is straightforward to check that repeated applications of the relations s_i^2 and $(s_1 s_2)^k$ allows us to rewrite the word $^{(s_2 s_1)^p} s_2$ as $^{(s_2 s_1)^{p'}} s_2$ for some $0 \leqslant p' < k/2$.

We compute $\phi(\partial_S \mathbf{T})$ using $\tilde{t}_1 = x_1$ and $\tilde{t}_2 = (x_2 x_1)^p x_2 (x_1 x_2)^p = x_2 (x_1 x_2)^{2p}$:

$$\phi(\partial_S \mathbf{T}) = \begin{pmatrix} 1 & 0 \\ s_2 \left(1 + s_1 s_2 + \cdots + (s_1 s_2)^{2p-1}\right) & 1 + s_2 s_1 + \cdots + (s_2 s_1)^{2p} \end{pmatrix}.$$

We use the representation $\eta : \mathbb{Z}W \to \mathbb{Z}\mathbb{Z}_2 = A$ in this case, setting $\eta(s_i) = \bar{s}$. For k odd this is the abelianisation, and for k even this is a quotient of the abelianisation

$\mathbb{Z}_2^2$ turns out not to matter. Applying η and taking the determinant we get

$$\det(\eta(\phi(\partial_S \mathbf{T}))) = \det \begin{pmatrix} 1 & 0 \\ 2p\overline{s} & 2p + 1 \end{pmatrix} = 2p + 1.$$

Finally we must compute the correction ideal. The Coxeter presentation for W with respect to S is

$$\langle s_1, s_2 \mid s_1^2, s_2^2, (s_1 s_2)^k \rangle,$$

So I_S^A is generated by

$$\eta(\phi(\partial_{x_1} x_1^2)) = \eta(1 + s_1) = 1 + \overline{s} = \eta(1 + s_2) = \eta(\phi(\partial_{x_2} x_2^2)),$$
$$\eta(\phi(\partial_{x_2} x_1^2)) = \eta(0) = 0 = \eta(0) = \eta(\phi(\partial_{x_1} x_2^2)),$$
$$\eta(\phi(\partial_{x_1} (x_1 x_2)^k)) = \eta\left(1 + s_1 s_2 + \cdots + (s_1 s_2)^{k-1}\right) = k,$$
$$\eta(\phi(\partial_{x_2} (x_1 x_2)^k)) = \eta\left(s_1 \left(1 + s_2 s_1 + \cdots + (s_2 s_1)^{k-1}\right)\right) = k\overline{s}.$$

One can then check that $I_S^A = \{a + b\overline{s} \mid a = b \pmod{k}\}$, and so quotienting $\xi \colon A \to A/I_S^A$ we gent an invariant valued in $A/I_S^A = \mathbb{Z}_k$.

In this case $A_W = \{\pm 1, \pm \overline{s}\} \to \{\pm 1\} = \xi(A_W)$, so S and T are inequivalent unless $2p + 1 = \pm 1 \pmod{k}$, ie unless

$$p = \begin{cases} 0 \text{ or } \frac{k}{2} - 1 & k \text{ even,} \\ 0 & k \text{ odd.} \end{cases}$$

Notice that $T = (t_1, t_2) = (s_1, {}^{(s_2 s_1)^p} s_2) = (s_1, s_2(s_1 s_2)^{2p})$ is Nielsen equivalent to $(t_1, t_1 t_2) = (s_1, (s_1 s_2)^{2p+1})$, where this second generator is a rotation through angle $2\pi(2p + 1)/k$. If $2p + 1 \geqslant k/2$ then replacing $t_1 t_2$ by its inverse $t_2 t_1$ gives a rotation by $2\pi(k - (2p + 1))/k$, and so running over $0 \leqslant p < k/2$ we see that T is Nielsen equivalent to (s_1, r) where r is a rotation by $2\pi\ell/k$ for $1 \leqslant \ell < k/2$ where $\gcd(\ell, k) = 1$ (if the gcd is not 1, then T does not generate W).

Now, two of these generating sets are equivalent if and only if either $p = 0$ or $p = k/2 - 1$ in the even case. If $p = 0$ then $\ell = 1$ and $t_2 = s_2$ meaning that $T = S$ is the standard generating set. If k is even then $p = k/2 - 1$ implies $2p - 1 \geqslant k/2$, but again this corresponds to $\ell = 1$. This time $t_2 = s_1 s_2 s_1$. This gives another proof of the first part of Theorem 2.1.

What this means is that in the case of dihedral groups, this approach gives a complete invariant for Nielsen equivalence of minimal generating tuples.

2.2.4 The general case

In arbitrary Coxeter groups we cannot perform such a detailed analysis since what we do in the case of dihedral groups is very heavily based on the fact that all elements in $W(I_2(k))$ can be represented by some reduced alternating word $s_1 s_2 s_1 \cdots$ or $s_2 s_1 s_2 \cdots$. Nevertheless, we can still assess how strong of an invariant this method gives.

Let (W, S) be a Coxeter system with T another generating tuple of the same size as S. Index the elements of S as s_{ij} such that any other generator $s_{i'j'}$ has $i' = i$ if and only if it is conjugate to s_{ij}. Using Proposition 1.6, we can denote the generators of W^{ab} by $\bar{s}_i$ such that the map $W \to W^{\mathrm{ab}}$ maps each $s_{ij} \mapsto \bar{s}_i$.

Now we want to compute the correction ideal I_S^A for $A = \mathbb{Z}W^{\mathrm{ab}}$. Letting x_{ij} be the generator of $\mathbb{F}_n$ such that $\phi(x_{ij}) = s_{ij}$. For each relation s_{ij}^2 we lift this to the word x_{ij}^2, then

$$\eta(\phi(\partial_{x_{ij}} x_{ij}^2)) = \eta(1 + s_{ij}) = 1 + \bar{s}_i, \tag{2.3}$$

$$\eta(\phi(\partial_{x_{i'j'}} x_{ij}^2)) = \eta(0) = 0, \text{ if } i \neq i' \text{ or } j \neq j'.$$

Fix a pair of distinct generators s_{ij} and $s_{i'j'}$; to simplify notation write $m = m_{ij,i'j'}$. Then for $m < \infty$, we lift the relation $(s_{ij} s_{i'j'})^m$ to the word $(x_{ij} x_{i'j'})^m$ and compute as follows. If $i \neq i'$, then m is necessarily even and

$$\eta(\phi(\partial_{x_{ij}} (x_{ij} x_{i'j'})^m)) = \eta(1 + s_{ij} s_{i'j'} + \cdots + (s_{ij} s_{i'j'})^{m-1}) = \frac{m}{2}(1 + \bar{s}_i \bar{s}_{i'}), \tag{2.4}$$

$$\eta(\phi(\partial_{x_{i'j'}} (x_{ij} x_{i'j'})^m)) = \eta(s_{ij}(1 + s_{ij} s_{i'j'} + \cdots + (s_{ij} s_{i'j'})^{m-1})) = \frac{m}{2}(\bar{s}_i + \bar{s}_{i'}), \tag{2.5}$$

$$\eta(\phi(\partial_{x_{i''j''}} (x_{ij} x_{i'j'})^m)) = \eta(0) = 0, \text{ if } i''j'' \notin \{ij, i'j'\}.$$

Otherwise, if $i = i'$ then $\eta(s_{ij}) = \eta(s_{i'j'}) = \bar{s}_i$ and

$$\eta(\phi(\partial_{x_{ij}} (x_{ij} x_{ij'})^m)) = \eta(1 + s_{ij} s_{ij'} + \cdots + (s_{ij} s_{ij'})^{m-1}) = m, \tag{2.6}$$

$$\eta(\phi(\partial_{x_{ij'}} (x_{ij} x_{ij'})^m)) = \eta(s_{ij}(1 + s_{ij} s_{ij'} + \cdots + (s_{ij} s_{ij'})^{m-1})) = m\bar{s}_i, \tag{2.7}$$

$$\eta(\phi(\partial_{x_{i''j''}} (x_{ij} x_{ij'})^m)) = \eta(0) = 0, \text{ if } i''j'' \notin \{ij, ij'\}.$$

LEMMA 2.23: *These generate the ideal*

$$I_S^A = \left\{ \sum_{\overline{w} \in W^{ab}} a_{\overline{w}} \overline{w} \;\middle|\; \sum_{\overline{w} \in W^{ab}} (-1)^{\ell(\overline{w})} a_{\overline{w}} = 0 \;(\text{mod } m_0) \right\},$$

where $\ell(\overline{w})$ is the length of $\overline{w}$ with respect to the length function on W^{ab} thought of as a Coxeter group (see Definition 1.9), and m_0 is the greatest common divisor of all the entries in the $m_{ij,i'j'}$'s for $ij \neq i'j'$.

Proof. Call the set in the statement I, which is an ideal. Each of the non-zero generators in equations (2.3)–(2.7) lie in I, so $I_S^A \subset I$.

Define the **complexity** of a non-zero element $x = \sum a_{\overline{w}} \overline{w}$ in I to be the the pair

$$c(x) := (M = \max\{\ell(\overline{w}) \mid a_{\overline{w}} \neq 0\}, N = \#\{\overline{w} \mid \ell(\overline{w}) = M\}),$$

and we order complexities lexicographically. First, given $x \in I$ we find $y \in I_S^A$ such that $c(x - y) = (0, 1)$ which implies that $x - y \in \mathbb{Z}$. Assume that $c(x)$ is greater than $(0, 1)$, and let $\overline{u}$ be maximal in length such that $a_{\overline{u}} \neq 0$ in x. Choose $\overline{s}_i$ such that $\ell(\overline{s}_i \overline{u}) < \ell(\overline{u})$ and call $\overline{v} = \overline{s}_i \overline{u}$. Then using the generator (2.3) for $\overline{s}_i$ we know that $y_1 = a_{\overline{u}} \overline{v}(1 + \overline{s}_i) = a_{\overline{u}} \overline{v} + a_{\overline{u}} \overline{u} \in I_S^A$. Now replace x with $x_1 = x - y_1$. In x_1, the coefficient of $\overline{u}$ is zero, and the only other coefficient which differs from x is that of $\overline{v}$ which satisfies $\ell(\overline{v}) < \ell(\overline{u})$. Hence $c(x_1)$ is strictly less than $c(x)$.

Repeating this finitely many times we arrive at $x_n = x - (y_1 + \cdots + y_n)$ having $c(x_n) = (0, 1)$. Since x_n still lies in I, this means $x_n = km_0$ for some $k \in \mathbb{Z}$.

Doing this same process where x is each of the generators in equations (2.4)–(2.7), we can show that I_S^A contains every finite $m_{ij,i'j'}$ for $ij \neq i'j'$. As an example, consider (2.4). Then

$$\frac{m}{2}(1 + \overline{s}_i \overline{s}_{i'}) - \frac{m}{2}\overline{s}_{i'}(1 + \overline{s}_i) - \frac{m}{2}(-1)(1 + \overline{s}_{i'}) = m.$$

It follows that I_S^A contains $m_0 \mathbb{Z}$, and hence x_n. Our original x is now a sum of two elements of I_S^A, proving that $I = I_S^A$. $\qquad\square$

Quotienting A by I_S^A gives a map $\xi \colon \mathbb{Z}W^{ab} \to \mathbb{Z}_{m_0}$, where each $\overline{s}_i$ in W^{ab} is

mapped to -1. We could have replaced the representation η at the outset with $\mathbb{Z}W \to \mathbb{Z}_{m_0}$ without any loss to the power of the invariant (we already saw a hint of this in the dihedral case). With all this in mind, the next Theorem follows immediately from the universal property of abelianisation.

THEOREM 2.24: *Let $\eta\colon \mathbb{Z}W \to A$ be a 1-dimensional representation of $\mathbb{Z}W$ to an abelian ring A. The Lustig-Moriah invariant χ_η factors through $\chi_{\eta^{ab}}$ where $\eta^{ab}\colon \mathbb{Z}W \to \mathbb{Z}W^{ab}$. In particular χ_η is valued in some quotient of $\mathbb{Z}_{m_0}$.*

The image of A_W under ξ is $\{\pm 1\}$, so given a generating tuple T with the same size as S we can conclude that if $\chi_{\eta^{ab}}(T) \neq \pm 1$ then T and S are not Nielsen equivalent. The upshot is that at best this method yields an invariant which takes values in the set of cosets $\mathbb{Z}_{m_0}/\{\pm 1\}$. For the dihedral group $W(I_2(k))$, $m_0 = k$ and there were only $\varphi(k)$ Nielsen equivalence classes of generating pairs (where φ is Euler's totient function), and so this provided a complete invariant. In general, however, m_0 can often be 1, making this invariant useless; and even if $m_0 > 1$, this is likely to be quite a coarse invariant. We apply this invariant in Example 4.53.

2.2.5 Higher dimensional invariants

It follows from Theorem 2.24 that in order to produce finer invariants from Lustig and Moriah's method we need to use a representation of $\mathbb{Z}W$ which is at least two dimensional (and whose image is not abelian). We can try to model their approach to Fuchsian groups using a mixed representation based on the group representation $G \to \mathrm{PSL}_2(\mathbb{C})$.

For Coxeter groups, the analogue is to mix the Tits representation, ρ from Theorem 1.8 with the quotient of W to an abelian group. Recall the image of ρ lies in the orthogonal group $O_n(B)$ of $\mathbb{R}^n$ with respect to the symmetric bilinear form B. Some form of mixing is necessary because otherwise $A = \mathbb{R}$, and the only non-trivial ideal which $I_S^{\mathbb{R}}$ could be, is $\mathbb{R}$ itself, leading to a single-valued invariant.

For Fuchsian groups, Lustig and Moriah restrict to the case that they could find a quotient $G \to \mathbb{Z}_p$ for some $p \geqslant 3$ (the case of Fuchsian groups with 2 torsion causes problems, see Section 1 in [79] and references therein). In Coxeter groups

we have no choice but to map $W \to \mathbb{Z}_2^k$ for some k. For simplicity, let us assume that (W, S) is an even Coxeter system (so that W had algebraic rank equal to the Coxeter rank), and we can take $k = n$ by mapping to W^{ab}. Index the elements of S as s_i for $1 \leqslant i \leqslant n$, and denote their images in W^{ab} by $\bar{s}_i$.

For every generator s_i, the image $\rho(s_i)$ is conjugate in $\mathrm{GL}_n(\mathbb{R})$ to the diagonal matrix $\mathrm{Diag}(-1, 1, \ldots, 1)$. Define a mixed representation of $\mathbb{Z}W$ by

$$\eta \colon \mathbb{Z}W \to \mathbb{M}_n(\mathbb{R}W^{\mathrm{ab}}) \colon s_i \mapsto \bar{s}_i \rho(s_i).$$

A routine computation shows that this does indeed define a representation.

Unfortunately this does not yield a useful invariant. To see this, we can start to compute I_S^A. Fix $1 \leqslant i \leqslant n$, then $\eta(\phi(\partial_{x_i} x_i^2)) = \eta(1 + s_i) = \mathbb{I} + \bar{s}_i \rho(s_i)$. If $m_{ij} = 2$ for all $j \neq i$, then this matrix is $\mathrm{Diag}(1 + \bar{s}_i, \ldots, \overset{i^{\mathrm{th}} \text{ entry}}{1 - \bar{s}_i}, \ldots, 1 + \bar{s}_i)$, and otherwise there are some additional off-diagonal entries lying in $\mathbb{R}\bar{s}_i - \{0\}$. In any case, I_S^A contains $1 + \bar{s}_i$ and $1 - \bar{s}_i$, and hence the whole of $\mathbb{R}\{1, \bar{s}_i\}$. Doing this for each i, it follows that $I_S^A = A$, and so the resulting invariant is single-valued. This problem seems to be very resistant to trying variations on the theme of a mixed representation of ρ with $\mathbb{Z}W \to \mathbb{Z}W^{\mathrm{ab}}$—we, at least, have been unable to overcome it. It seems unlikely therefore that this approach can lead to any kind of useful Nielsen equivalence invariant based on the Tits representation.

Remark 2.25 (Fuchsian groups). This issue does not arise for Fuchsian groups because every generator is mapped by the representation to a matrix conjugate to (2.2), and so writing ζ_i for some primitive γ_i-root of unity, the correction ideal is generated in part by

$$\eta(\phi(\partial_{x_i} x_i^{\gamma_i})) = \eta(1 + s_i + \cdots + s_i^{\gamma_i - 1})$$
$$= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \begin{pmatrix} \zeta_i & 0 \\ 0 & \zeta_i^{-1} \end{pmatrix} + \cdots + \begin{pmatrix} \zeta_i^{\gamma_i - 1} & 0 \\ 0 & \zeta_i^{-(\gamma_i - 1)} \end{pmatrix} = 0.$$

See the proof of Lemma 6.4 in [79] for details.

Chapter 3

Weyl groups and quiver mutations

IN THIS CHAPTER, WE LOOK at what could be thought of as an application of reflection equivalence based on the surprising connection between theory of *quiver mutations*, a tool from the theory of cluster algebras, and presentations of Weyl groups with generating sets consisting of reflections. We discuss this this before studying reflection equivalence in Chapter 4 because the material here does not depend on any of the results in that Chapter, and uses completely distinct tools.

In the first Section of this Chapter, we summarise the main definitions and results we need from [6, 48] on quivers and their associated presentations. A quiver is simply a directed and edge-labelled simple graph; a mutation of a quiver is a combinatorial transformation which turns one quiver into another with the same vertex set. We can associate a Coxeter-like group presentation to a quiver in which each generator has order 2. Each mutation corresponds to an explicit isomorphism between the groups whose presentation is associated to the quivers.

If a quiver is obtained by orienting and modifying the labelling of the Coxeter-Dynkin diagram of a Weyl group in a certain way, the Coxeter-like presentation is exactly the Coxeter presentation of the Weyl group. Thus, quiver mutations give a way to produce other presentations for these Weyl groups. Moreover, the isomorphisms associated to the mutations can be expressed as a sequence of partial conjugations—ie mutations correspond to certain reflection equivalences.

The aim of this Chapter is to investigate the equivalence relation on the set of reflection generating tuples of Weyl groups associated to quivers which is induced

by mutation. We show that all such generating tuples are mutation equivalent to a *standard one* (see Definition 3.11). In many cases all standard generating tuples are mutation equivalent, see Theorem 3.36 and Corollary 3.37.

It is not true that every reflection generating tuple of a Weyl group arises from a quiver. However, at least in the case of type A Weyl groups, we are able to generalise mutation to *mutation modulo* 2, after which every minimal reflection generating tuple can be associated to a quiver, and all of these are equivalent under mutation modulo 2, see Corollary 3.44.

Section 3.3.1 is devoted to stating and proving Proposition 3.18, which is the main tool for the subsequent main results later in the Chapter. This Proposition allows us to alter the orientation on quivers whose underlying graph is a tree without affecting the associated group. The proof of this formed the basis of a collaborative art project with Melissa Rodd [98].

3.1 Quivers and group presentations

Definition 3.1. A **quiver** is a finite, edge-labelled, directed graph which contains no loops or multiple edges. Each edge is labelled by a positive integer called the **weight** of the edge; edge labels of 1 are suppressed. Forgetting the orientations on the quiver Q yields its **underlying weighted graph** and the **underlying graph** is the unlabelled graph one gets by also forgetting the weights on all of its edges. We call a quiver **treelike** if its underlying graph is a tree.

There is a way to combinatorially transform a quiver to produce a new quiver. First introduced in [48], the formulation here is taken from [53].

Definition 3.2. Let Q be a quiver with vertex set V, and let $v \in V$ be a vertex. Then $\mu_v(Q)$ is the quiver with the same vertex set V and edges defined as follows.

1. For each oriented path $u \bullet \xrightarrow{m} \bullet \xrightarrow{n} \bullet w$ of length 2 through v in Q, we add an edge $u \bullet \xrightarrow{mn} \bullet w$

2. Reverse the orientation of all edges incident at v

3. If any double edges have been created, remove them as shown in Figure 3.1
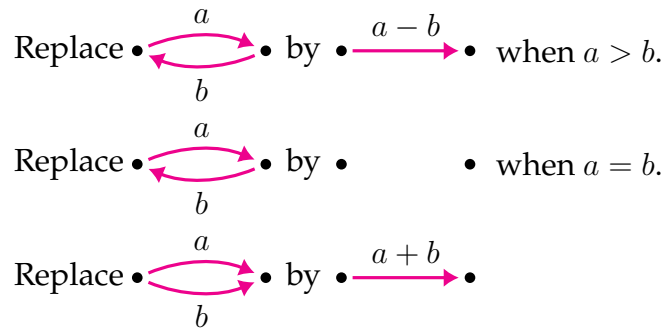
Figure 3.1: Removing double edges when mutating a quiver.

Then $\mu_v(Q)$ is said to be obtained by **mutation at vertex** v.

It follows from the definition that $\mu_v(\mu_v(Q)) = Q$.

Remark 3.3 (Mutations and connectivity). A simple observation is that not only do mutations not change the vertex set, they also do not change the connected components. That is because if an edge is either added or removed between u and w by the mutation μ_v, then by the definition of μ_v, u and w must be neighbours of v both before and after mutation. Almost everything we do works equally for disconnected quivers, however for simplicity we consider only the connected case.

Example 3.4. Consider the following sequence of quivers which illustrates a mutation at the vertex v.
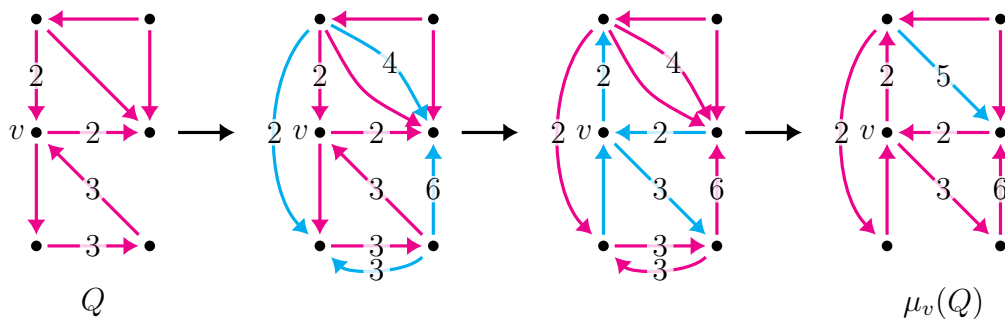
Figure 3.2: An example of a quiver mutation.

Definition 3.5. Two quivers Q and Q' with vertex set V are said to be **mutation equivalent** if there is a sequence of quivers $Q_0, Q_1, \ldots, Q_n$, and a sequence of vertices $v_0, v_1, \ldots, v_{n-1} \in V$ such that $Q = Q_0$, $Q' = Q_n$, and, for each $1 \leqslant i \leqslant n$, $Q_i = \mu_{v_{i-1}}(Q_{i-1})$. Since mutations are reversible, it follows that mutation equivalence is an equivalence relation.

Let $\mathcal{V}$ be a Coxeter-Dynkin diagram of type A, B, D, E, F, or G, and let $\overline{\mathcal{V}}$ be the weighted graph obtained from $\mathcal{V}$ by replacing each weight 3 edge by a weight 1 edge; each weight 4 edge by a weight 2 edge; and each weight 6 edge with a weight 3 edge.

A quiver Q is **mutation-Dynkin** if it is mutation equivalent to some quiver Q' whose underlying weighted graph is $\overline{\mathcal{V}}$; then the **type** of Q is $\mathcal{V}$. We say that Q' is a **quiver obtained from** $\mathcal{V}$.

THEOREM 3.6 (Proposition 9.7 in [48])**:** *If Q is mutation-Dynkin, then all edges have weight* 1, 2, *or* 3, *and all chordless cycles have one of the forms shown in Figure 3.3 (a cycle is chordless if it does not backtrack and the subgraph induced by the vertices of the cycle contains no edges which are not in the cycle). In particular, they are oriented.*
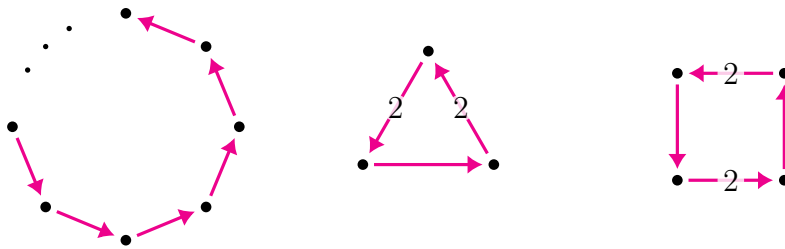
Figure 3.3: The possible chordless cycles in a mutation-Dynkin quiver.

Given a mutation-Dynkin quiver we can associate to it a group as follows.

Definition 3.7 (Introduction of [6])**.** Let Q be a mutation-Dynkin quiver with vertex set V, then define the **quiver group** G_Q associated to Q to be generated by the vertices V, subject to the relations:

1. v^2 for all $v \in V$

2. $(vu)^{m_{vu}}$ for all pairs $\{v, u\} \subset V$ with $v \neq u$, where

$$
m_{vu} = \begin{cases}
2 & \text{if } v\,\bullet \qquad\quad \bullet\,u \\
3 & \text{if } v\,\bullet\!\longrightarrow\!\bullet\,u \\
4 & \text{if } v\,\bullet\!-\!2\!\rightarrow\!\bullet\,u \\
6 & \text{if } v\,\bullet\!-\!3\!\rightarrow\!\bullet\,u
\end{cases}
$$

3. $(v_1 v_2 \cdots v_{n-1} v_n v_{n-1} \cdots v_2)^2$ for each chordless cycle $\overset{v_1}{\bullet}\!\longrightarrow\!\overset{v_2}{\bullet}-\ \cdots\ \rightarrow\!\overset{v_n}{\bullet}\!\longrightarrow\!\overset{v_1}{\bullet}$
   with either all edge weights equal to 1, or the weight of $v_n\,\bullet\!\longrightarrow\!\bullet\,v_1$ is 2

If Q is obtained from a Dynkin diagram $\mathcal{V}$, then the presentation defined above is the Coxeter presentation, so we have an isomorphism $G_Q \cong W(\mathcal{V})$. Note that G_Q only depends on the underlying weighted graph of Q, not on its orientation. Compare the proof of the following with, for example, Lemma 4.1 in [6].

LEMMA 3.8: *Let Q and Q' be two mutation-Dynkin quivers with the same underlying graph. Then an isomorphism of the underlying weighted graphs of Q and Q', restricted to the vertex sets $VQ \to VQ'$, induces an isomorphism $G_Q \to G_{Q'}$.*

Proof. It is sufficient to show that G_Q has the same relations if we change the orientations on the edges of Q. The relations in (1) do not depend on the edges so are left unchanged. Suppose Q contains and edge $v\bullet\!\!-\!\!w\!\!\rightarrow\!\!\bullet u$ which leads to the relation $(vu)^{m_{vu}}$. The if we replace the edge with its reverse $v\bullet\!\!\leftarrow\!\!w\!\!-\!\!\bullet u$ this gives the relation $(uv)^{m_{uv}}$. Since w has not changed, $m_{uv} = m_{vu}$, and the relation $(uv)^{m_{uv}} = (uv)^{m_{vu}}$ implies the relation $(uv)^{-m_{vu}} = (v^{-1}u^{-1})^{m_{vu}} = (vu)^{m_{vu}}$ where the last equality comes from the relations of type (1). Therefore in the presence of relations of type (1), relations of type (2) do not depend on the orientations on the edges.

Finally, suppose Q contains an oriented cycle $\overset{v_1}{\bullet}\!\!\rightarrow\!\!\overset{v_2}{\bullet}\!\!-\cdots\rightarrow\!\!\overset{v_n}{\bullet}\!\!\rightarrow\!\!\overset{v_1}{\bullet}$ with either all edge weights equal to 1, or the weight of $v_n\bullet\!\!\rightarrow\!\!\bullet v_1$ is 2. If the orientations of one of these edges is reversed in Q', then they all must be, since Theorem 3.6 guarantees all chordless cycles are oriented. After flipping all of the orientations and performing a cyclic permutation we get a cycle of the form

$$\overset{v_n}{\bullet}\!\!\rightarrow\!\!\overset{v_{n-1}}{\bullet}\!\!-\cdots\rightarrow\!\!\overset{v_1}{\bullet}\!\!\rightarrow\!\!\overset{v_n}{\bullet}$$

with either all edge weights equal to 1, or the weight of $v_1\bullet\!\!\rightarrow\!\!\bullet v_n$ is 2. The resulting relation of type (3) is $(v_n v_{n-1} \cdots v_2 v_1 v_2 \cdots v_{n-1})^2$. Writing this out (with suggestive parentheses)

$$(v_n v_{n-1} \cdots v_2)(v_1 v_2 \cdots v_{n-1} v_n v_{n-1} \cdots v_2 v_1 v_2 \cdots v_{n-1}),$$

the relation implies all of its cyclic conjugates are also relations. In particular,

$$(v_1 v_2 \cdots v_{n-1} v_n v_{n-1} \cdots v_2 v_1 v_2 \cdots v_{n-1})(v_n v_{n-1} \cdots v_2),$$

which equals $(v_1 v_2 \cdots v_{n-1} v_n v_{n-1} \cdots v_2)^2$, the relation of type (3) in G_Q. $\qquad\square$

The surprising result is that these groups associated to mutation-Dynkin quivers transform nicely under quiver mutations.

THEOREM 3.9 (Theorem 5.4 in [6]): *Let Q be mutation-Dynkin of type $\mathcal{V}$ and let v be a vertex of Q, then there is an isomorphism of groups $G_Q \cong G_{\mu_v(Q)}$, defined on each generator $u \in VQ$ by*

$$\psi_v : G_Q \to G_{\mu_v(Q)} : u \mapsto \begin{cases} vuv^{-1} & \text{if } v\bullet\!\!\longrightarrow\!\!\bullet u \\ u & \text{otherwise} \end{cases}.$$

In particular, $G_Q \cong W(\mathcal{V}) \cong G_{\mu_v(Q)}$.

Note that since Q and $\mu_v(Q)$ have the same vertex set, G_Q and $G_{\mu_v(Q)}$ have the same tuples as their formal generating tuples (though the elements of these tuples are different, in general, as elements of the groups).

3.2 Presentation quivers and mutation equivalence

Quiver mutations are not a viable tool to study reflection equivalence in Weyl groups, because the proof that all generating tuples of reflections associated to quivers are reflection equivalent is almost tautological, see Proposition 3.16. Instead, we can think of the equivalence relation on reflection generating tuples which arises out of quiver mutations as an *a priori* stronger relation than refection equivalence. We make this question precise in Questions 3.14 and 3.15 below.

Assumption 3.10. For the rest of this Chapter, fix a Coxeter-Dynkin diagram $\mathcal{V}$ of type A, B, D, E, F, or G and let V be a set with the same size as $V\mathcal{V}$. Fix an ordering on V, and we assume that every quiver Q in the sequel has vertex set V.

Recall that the diagram $\mathcal{V}$ specifies $W = W(\mathcal{V})$ not only up to isomorphism, but also a specific Coxeter system, ie $(W, V\mathcal{V})$. In light of Theorem 3.9 we can use mutation-Dynkin quivers to represent certain markings of a Weyl group where the generators are reflections.

Definition 3.11. Let Q be a mutation-Dynkin quiver of type $\mathcal{V}$ and $\phi\colon G_Q \to W(\mathcal{V})$ an isomorphism from the quiver group of Q such that $\phi(V)$ is a subset of the set

of reflections of $W(\mathcal{V})$. We call the pair (Q, ϕ) a **presentation quiver** of **type** $\mathcal{V}$. (Q, ϕ) is **treelike** if Q is treelike; it is **standard** if the Q is obtained from $\mathcal{V}$ (see Definition 3.5), and ϕ maps $V \to V\mathcal{V}$, ie maps the generators of G_Q to the Coxeter generators of $W(\mathcal{V})$.

Specifying a presentation quiver (Q, ϕ) is equivalent to specifying a reflection generating tuple $X = \phi(V)$ (where the order is induced by the order on V) together with a presentation for $W(\mathcal{V})$ of the form given in Definition 3.7. If (Q, ϕ) is standard then $\phi(V)$ equals $V\mathcal{V}$ up to a graph automorphism, and the presentation is the usual Coxeter presentation of $W(\mathcal{V})$.

Theorem 3.9 allows us to extend mutations of quivers to mutations of presentation quivers.

Definition 3.12. Given a presentation quiver (Q, ϕ) and $v \in V$, let $Q' = \mu_v(Q)$ (note $VQ' = V$ by the definition of a mutation), and let $\psi_v \colon G_Q \xrightarrow{\sim} G_{Q'}$ be the isomorphism given by Theorem 3.9. Then define the **mutation** of (Q, ϕ) to be $\mu_v(Q, \phi) := (Q', \phi \circ \psi_v^{-1})$, and we write $\mu_v(\phi) = \phi \circ \psi_v^{-1}$.

Two presentation quivers (Q, ϕ) and (Q', ϕ') are **mutation equivalent** if there is a sequence of presentation quivers $(Q_0, \phi_0), (Q_1, \phi_1), \ldots, (Q_n, \phi_n)$, and a sequence of vertices $v_0, v_1, \ldots, v_{n-1} \in V$ such that $(Q, \phi) = (Q_0, \phi_0)$, $(Q', \phi') = (Q_n, \phi_n)$, and for each $1 \leqslant i \leqslant n$, $(Q_i, \phi_i) = \mu_{v_{i-1}}(Q_{i-1}, \phi_{i-1})$.

We define $\mu_v(\phi)$ in this way so that the following diagram commutes:

$$Q \xrightarrow{\mu_v} Q'$$

$$
\begin{array}{ccc}
G_Q & \xrightarrow{\psi_v} & G_{Q'} \\
\downarrow{\scriptstyle \phi} & & \downarrow{\scriptstyle \mu_v(\phi)} \\
W(\mathcal{V}) & \xrightarrow{\text{id}} & W(\mathcal{V})
\end{array}
$$

LEMMA 3.13: *Let (Q, ϕ) be a presentation quiver and $v \in V$, then $\mu_v^4(Q, \phi) = (Q, \phi)$. Consequently mutation equivalence is an equivalence relation on presentation quivers.*

Proof. The second claim follows from the first. Mutations of quivers are involutions so $\mu_v^2(Q) = Q$, while μ_v^2 acts on ϕ by conjugating every neighbour of v in

Q by v. In fact, one can check that $\mu_v^2(Q, \phi) = (Q, \phi \circ \alpha_v)$ where α_v is the inner automorphism of G_Q associated to v. Since v has order 2 as an element of G_Q, $\mu_v^4(\phi) = \phi$. $\hfill\square$

Mutation equivalence of presentation quivers induces an equivalence relation on generating tuples of reflections for $W(\mathcal{V})$ coming from mutation-Dynkin quivers. We call this relation *mutation equivalence* as well. *A priori* this is a proper subset of the set of reflection generating tuples of size $\#V\mathcal{V}$, but on this subset, again *a priori*, mutation equivalence is a stronger relation than reflection equivalence.

As an example, let (Q, ϕ) be a standard presentation quiver, and α an inner automorphism of $W(\mathcal{V})$. Then the generating tuples associated to (Q, ϕ) and $(Q, \alpha \circ \phi)$ are reflection equivalent, however it is not at all clear that they are mutation equivalent. This is because mutations do not allow freedom over which partial conjugations are performed, since these are dictated by the orientation on Q. In addition, mutations usually change the underlying weighted graph of Q, and it is not immediately clear how to engineer a sequence of mutations which return some mutation of Q, to a quiver whose underlying weighted graph is $\overline{\mathcal{V}}$, in a way which retains control over how ϕ changes (see Definition 3.5 for the definition of $\overline{\mathcal{V}}$).

This motivates the following questions.

Question 3.14. Can every reflection generating tuple with size $\#V\mathcal{V}$ of a Weyl group $W(\mathcal{V})$ be represented by a presentation quiver of type $\mathcal{V}$?

Question 3.15. When are two reflection generating tuples which are represented by presentation quivers of type $\mathcal{V}$ mutation equivalent? In particular, if two such tuples are reflection equivalent, are they also mutation equivalent?

At the end of this Chapter (Section 3.5), we look at the first question for $\mathcal{V} = A_n$, and we see that the answer is no, however we can generalise our definitions to work with arbitrary reflection generating tuples. Until then, we focus on answering when two presentation quivers of type $\mathcal{V}$ are mutation equivalent. The fact that mutation equivalence is at least as strong as reflection equivalence follows immediately from the definition; in fact, we can prove that any two presentation quivers of type $\mathcal{V}$ are reflection equivalent.

PROPOSITION 3.16: *Let (Q_1, ϕ_1) and (Q_2, ϕ_2) be two presentation quivers of type $\mathcal{V}$, then the generating tuples they represent are reflection equivalent.*

Proof. Any two generating tuples represented by standard presentation quivers differ by a graph automorphism and the orientations on the quiver. This means the generating tuples differ by a permutation, and so are equivalent under a transformation of type (T1). Thus, it suffices to show that any generating tuple coming from a presentation quiver (Q, ϕ) is reflection equivalent to a generating tuple coming from a standard generating tuple.

By definition, Q is mutation equivalent to some orientation of $\overline{\mathcal{V}}, Q'$. Therefore, after these mutations, (Q, ϕ) is mutation equivalent to (Q', ϕ') where $\phi' \colon G_{Q'} \to W(\mathcal{V})$ and $(G_{Q'}, V)$ is a Coxeter system for $G_{Q'} \cong W(\mathcal{V})$. After choosing some isomorphism between the underlying weighted graph of Q' and $\overline{\mathcal{V}}$ (the choice is arbitrary), we can interpret ϕ' as an automorphism of $W(\mathcal{V})$ which preserves the set of reflections. Then by Proposition 3.34, this automorphism can be expressed as a composition of an inner automorphism and a graph automorphism.

We can apply a sequence of (T4) moves to the image of ϕ' to undo the inner automorphism and replace (Q', ϕ') with (Q', ϕ'') where $\phi'' \colon V \to V\mathcal{V}$, and this presentation quiver is standard. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

3.3   Defining mutations of unoriented graphs

The main step in understanding when two presentation quivers are mutation equivalent to each other is to gain control over when and how we can perform mutations on (Q, ϕ) which behave as we want them to on ϕ. The impediment to this is the orientation on Q which dictates everything about $\mu_v(\phi)$. We want to be able to perform mutations on (Q, ϕ) to change the orientation on Q arbitrarily, but which leave ϕ unchanged.

If we can do this, then we can change the orientations on the edges which meet a vertex v so that the isomorphism ψ_v from Theorem 3.9 consists of exactly the partial conjugations we want. It is reasonable to hope this may sometimes be possible. This is because the definition of the group G_Q associated to a quiver

Q only depends on the underlying weighted graph of Q. We can simplify the amount of information we need to keep track of by defining mutations on unoriented weighted graphs.

We can make this more formal as follows. Let Υ be a finite weighted graph with no edge loops or multiple edges and let $\mathcal{Q}(\Upsilon)$ be the set of all quivers whose underlying weighted graph is Υ. If every $Q \in \mathcal{Q}(\Upsilon)$ is mutation-Dynkin then there is a canonical isomorphism between G_Q and $G_{Q'}$ for any $Q, Q' \in \mathcal{Q}(\Upsilon)$ which is induced by the identity map on Υ, see Lemma 3.8. Denote by G_Υ the group (up to canonical isomorphism) associated to every quiver in $\mathcal{Q}(\Upsilon)$. Now fix an isomorphism $\phi \colon G_\Upsilon \to W(\mathcal{V})$ such that $\phi(V)$ lies in the set of reflections. This ϕ gives rise to a presentation quiver (Q, ϕ) for each $Q \in \mathcal{Q}(\Upsilon)$. Here we are suppressing the canonical isomorphisms to simplify notation.

Consider the graph which has vertex set $\{(Q, \phi) \mid Q \in \mathcal{Q}(\Upsilon)\}$ and an edge between any pair of presentations which are related by a single mutation—in particular, these mutations change the orientations on the quivers without changing either the underlying weighted graph or the map ϕ. We want to show that this graph is connected. Then we can freely change the orientation on Q without affecting the rest of the presentation quiver, giving us the control we need to perform any 'non-trivial' mutation.

Note that it can only be the case if Υ contains no cycles, otherwise $\mathcal{Q}(\Upsilon)$ contains quivers with unoriented cycles which cannot be mutation-Dynkin by Theorem 3.6. Since Υ contains no cycles, its underlying graph must be a forest. For simplicity, set us assume that Υ is connected. It follows from the classification of finite Coxeter groups (see Table 1.1) that if Υ is a tree and some quiver in $\mathcal{Q}(\Upsilon)$ is mutation-Dynkin, then $\Upsilon = \overline{\mathcal{V}}$ for some diagram of type A, B, D, E, F, or G.

If we can show that the graph described above is connected, then we can define mutations of pairs (Υ, ϕ) without the need for orientations. In preparation for this, and restricting to the case that $\Upsilon = \overline{\mathcal{V}}$, we make the following definition.

Definition 3.17. A **presentation Dynkin diagram** is a pair $(\overline{\mathcal{V}}, \phi)$ such that ϕ is an isomorphism $G_{\overline{\mathcal{V}}} \to W(\mathcal{V})$ which maps V to the set of reflections. We call $(\overline{\mathcal{V}}, \phi)$ a **standard** presentation Dynkin diagram of type $\mathcal{V}$ if $\phi(V) = V\mathcal{V}$.

3.3.1   Forgetting orientations on treelike quivers

We now prove that given a presentation quiver (Q, ϕ) such that the underlying weighted graph of Q is $\overline{\mathcal{V}}$, we can freely change the orientation on Q by a sequence of mutations which do not affect the underlying graph, or the map ϕ. Our proof does not rely on any special properties of mutation-Dynkin quivers, only on the fact that the quiver is tree-like—we phrase the following Proposition to reflect that.

PROPOSITION 3.18: *Let (Q_1, ϕ_1) and (Q_2, ϕ_2) be two treelike presentation quivers with the same underlying weighted graph, and $\phi_1 = \phi_2$, in other words (Q_1, ϕ_1) and (Q_2, ϕ_2) differ only by the orientations on Q_1 and Q_2. Then (Q_1, ϕ_1) and (Q_2, ϕ_2) are mutation equivalent by a sequence of mutations which do not change either the underlying weighted graph of Q_1 or ϕ_1 at any stage.*

Assumption 3.19. Choose a distinguished vertex $v_0 \in V$. In this Section all quivers share this same underlying graph and distinguished vertex.

Definition 3.20. Let Q be a treelike quiver with distinguished vertex v_0, then for any vertex $v \in V$ there is a unique unoriented path between v_0 and v which does not backtrack, denote this by $[v_0, v]$. The **height** of a vertex $v \in V$ in Q is

$$h^Q(v) := \sum_{e \in [v_0, v]} \mathcal{O}^Q(e)$$

where the sum is taken over all edges e in the path $[v_0, v]$, and

$$\mathcal{O}^Q(e) := \begin{cases} 1 & \text{if } e \text{ is oriented towards } v_0 \\ -1 & \text{if } e \text{ is oriented away from } v_0 \end{cases}$$

Note that this definition depends on the choice of v_0. However, since we have fixed this choice at the start we are suppressing it in the notation to make our expressions visually simpler. We then write

$$H(Q) := \sum_{v \in V} h^Q(v)$$

A vertex $v \in V$ is **prominent** in Q if $h^Q(v) > h^Q(u)$ for all vertices u adjacent to v in Q.

Let Q' be another quiver, we say that Q **dominates** Q', and write $Q \succeq Q'$, if $h^Q(v) \geqslant h^{Q'}(v)$ for all $v \in V$.

Note that $Q \succeq Q'$ implies $H(Q) \geqslant H(Q')$. It is useful to visualise a quiver by extending h^Q to a graph $h^Q : Q \to \mathbb{R}$ over Q, by mapping edges linearly between the heights of their endpoints. Then a vertex is prominent if it is a local maximum of this graph, and one quiver dominates another if the graph of one is higher than the graph of the other, see Figure 3.4.

Definition 3.21. Denote by $\check{Q}$ the quiver in which all edges are oriented towards v_0, and by $\widehat{Q}$ the quiver in which all edges are oriented away from v_0.
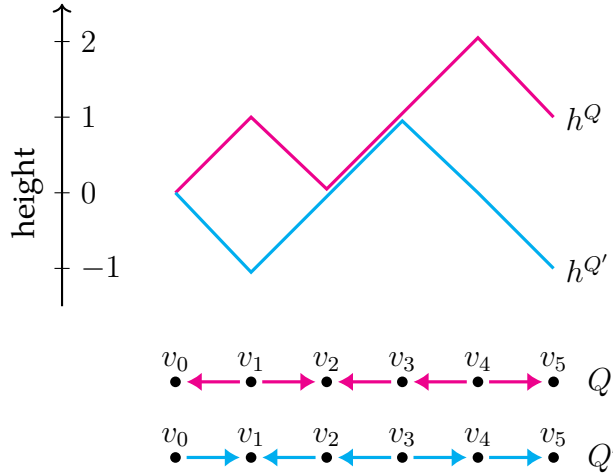
Figure 3.4: Two quivers and their corresponding graphs. The vertices v_1 and v_4 are prominent in Q, which dominates Q'.

LEMMA 3.22: *The relation $\preceq$ defines a partial ordering on $\mathcal{Q}(\overline{\mathcal{V}})$. For any quiver Q in this poset, $\widehat{Q} \preceq Q \preceq \check{Q}$, meaning this poset is a bounded lattice. Additionally, every chain has length bounded by $(H(\check{Q}) - H(\widehat{Q}))/2$.*

Proof. We need to show that if $Q \preceq Q'$ and $Q' \preceq Q$ then $Q = Q'$. Let e be an edge in the underlying graph, and let $\iota(e)$ be the endpoint closest to v_0, and $\tau(e)$ its other endpoint. By the hypothesis $h^Q(\iota(e)) = h^{Q'}(\iota(e))$ and $h^Q(\tau(e)) = h^{Q'}(\tau(e))$ so $h^Q(\tau(e)) - h^Q(\iota(e)) = h^{Q'}(\tau(e)) - h^{Q'}(\iota(e))$. But by the definition of h^Q and $h^{Q'}$ this implies $\mathcal{O}^Q(e) = \mathcal{O}^{Q'}(e)$, as required.

For all $v \in V$,

$$h^{\widehat{Q}}(v) = \sum_{e \in [v_0, v]} (-1) \leqslant \sum_{e \in [v_0, v]} \mathcal{O}^Q(e) \leqslant \sum_{e \in [v_0, v]} (+1) = h^{\check{Q}}(v)$$

and so $H(\widehat{Q}) \leqslant H(Q) \leqslant H(\check{Q})$ for all Q. Moreover, if $Q' \prec Q$ then $H(Q') < H(Q)$, so all chains have length bounded by $H(\check{Q}) - H(\widehat{Q})$. It follows from equations (3.1) and (3.3) below, that chain length is bounded by $(H(\check{Q}) - H(\widehat{Q}))/2$. $\square$

Definition 3.23. Following Section 9.1 of [48], we say that a a mutation is **shape preserving** if it does not change the underlying weighted graph of the quiver.

It follows from the definition that the mutation at vertex v is shape preserving if and only if either all edges incident to v are oriented towards v, or all are oriented away from v. Assume that the mutation μ_v of Q is shape preserving, and interpret it as a mutation of a presentation quiver (Q, ϕ). By Theorem 3.9, $\psi_v : G_Q \to G_{\mu_v(Q)} = G_Q$ is the identity map if and only if all edges incident to v are oriented away from v, or equivalently, if v is prominent.

We can therefore rephrase Proposition 3.18 as follows. Any two treelike quivers with the same underlying graph are mutation equivalent via a sequence of mutations at prominent vertices. Note that this is a stronger version of the last part of Proposition 9.2 in [48]. The effect of mutating a quiver at a prominent vertex v is merely to reorient all of the edges incident to v so that they point towards v. In Figure 3.4, mutating Q at the prominent vertex v_1 replaces the peak in the height graph with a valley.

In order to prove this, we consider what happens with the height function when we perform mutations at prominent vertices. We consider three operations. The first we call *gradient ascent*, which allows us to find a prominent vertex in some quiver which dominates another. We call the second *erosion* and it allows us to replace a prominent vertex with a 'valley' while preserving the property of the eroded quiver dominating another. Finally we have *elevation* which corresponds to eroding v_0, but has the effect of raising the height of the rest of the quiver.

LEMMA 3.24 (Gradient ascent)**:** *Let* $u \bullet \!\! \longleftarrow \!\! \bullet v$ *be an edge in Q, then* $h^Q(v) = h^Q(u) + 1$. *If* $Q \succ Q'$, *then there is a vertex* $\hat{v} \in V - \{v_0\}$ *such that* $\hat{v}$ *is prominent in Q and* $h^Q(\hat{v}) > h^{Q'}(\hat{v})$.

Proof. There are two cases:

if $u \in [v_0, v]$ then   otherwise, if $v \in [v_0, u]$ then

$$h^Q(v) = \sum_{e \in [v_0, v]} \mathcal{O}^Q(e) \qquad\qquad h^Q(v) = \sum_{e \in [v_0, v]} \mathcal{O}^Q(e)$$

$$= \sum_{e \in [v_0, u]} \mathcal{O}^Q(e) + \mathcal{O}^Q(u \bullet\!\!\longleftarrow\!\!\bullet v) \qquad\qquad = \sum_{e \in [v_0, u]} \mathcal{O}^Q(e) - \mathcal{O}^Q(v \bullet\!\!\longrightarrow\!\!\bullet u)$$

$$= 1 + h^Q(u), \qquad\qquad\qquad\qquad\qquad = h^Q(u) + 1$$

If $Q \succ Q'$ there is some vertex $u \in V$ such that $h^Q(u) > h^{Q'}(u)$. If u is prominent, set $\hat{v} = u$. Otherwise there is some neighbour v of u such that $u \bullet\!\!\longleftarrow\!\!\bullet v$ in Q. We claim $h^Q(v) > h^{Q'}(v)$. Indeed

$$h^Q(v) = h^Q(u) + 1 > h^{Q'}(u) + 1 \geqslant h^{Q'}(v)$$

where the second inequality comes from the fact $h^{Q'}(v) = h^{Q'}(u) \pm 1$, depending on the orientation of the edge between u and v in Q'.

Now, if v is prominent, set $\hat{v} = v$. Otherwise we repeat this process starting with v. We find a prominent vertex after a finite number of steps since the underlying graph of Q is a *finite* tree. Necessarily $\hat{v} \neq v_0$ since by definition $h^Q(v_0) = 0 = h^{Q'}(v_0)$. $\square$

LEMMA 3.25 (Erosion): *Let $\hat{v} \in V - \{v_0\}$ be prominent in Q, then $Q \succ \mu_{\hat{v}}(Q)$. If $Q \succ Q'$ let $\hat{v} \in V$ be the vertex from Lemma 3.24, then $Q \succ \mu_{\hat{v}}(Q) \succeq Q'$.*

Proof. For $v \in V$,

$$h^{\mu_{\hat{v}}(Q)}(v) = \begin{cases} h^Q(v) & \text{if } v \neq \hat{v} \\ h^Q(v) - 2 & \text{if } v = \hat{v} \end{cases} \tag{3.1}$$

so indeed $Q \succ \mu_{\hat{v}}(Q)$.

Now suppose, for a contradiction, that $\mu_{\hat{v}}(Q) \not\succeq Q'$. Then for all $v \in V - \{\hat{v}\}$, $h^{\mu_{\hat{v}}(Q)}(v) = h^Q(v) \geqslant h^{Q'}(v)$ so necessarily $h^{\mu_{\hat{v}}(Q)}(\hat{v}) < h^{Q'}(\hat{v}) < h^Q(\hat{v})$. It follows from (3.1) that

$$h^{Q'}(\hat{v}) = h^{\mu_{\hat{v}}(Q)}(\hat{v}) + 1$$

but this is not possible, because in $\mu_{\hat{v}}(Q)$ every edge incident to $\hat{v}$ is oriented towards it. A simple calculation shows that if u is a neighbour of $\hat{v}$ in $\mu_{\hat{v}}(Q)$ then

$h^{\mu_{\hat{v}}(Q)}(u) = h^{\mu_{\hat{v}}(Q)}(\hat{v})+1$, but we saw that $h^{\mu_{\hat{v}}(Q)}(u) = h^{Q'}(u)$. Therefore in Q', $\hat{v}$ and u are neighbouring vertices with the same height, contradicting Lemma 3.24. $\square$

LEMMA 3.26 (Elevation): *If $Q \neq \check{Q}$, there is a quiver Q_{elev} which is mutation equivalent to Q by a sequence of mutations at prominent vertices such that $H(Q) < H(Q_{\mathrm{elev}})$.*

Proof. We say that a vertex $v \in V$ is **visible** from v_0 on Q if $[v_0, v]$ is oriented, and directed towards v_0. Define Q_{vis} to be the the sub-quiver of Q induced by the set of visible vertices. If Q_{vis} is the single point v_0, then v_0 is prominent and write $Q' = Q$. Otherwise every leaf in Q_{vis} (ie every vertex of valence 1) is prominent.

Let v be such a leaf in Q_{vis} and consider $\mu_v(Q)$; v is no longer visible, but every other vertex which was visible in Q is still visible in $\mu_v(Q)$, and there are no new visible vertices. Therefore, $(\mu_v(Q))_{\mathrm{vis}} \subset Q_{\mathrm{vis}}$ is the subgraph obtained by removing v and the single edge incident to it. We can therefore inductively reduce the number of visible vertices by mutating at the leaves of the sub-quiver of visible vertices. We end up with a quiver Q', which is mutation equivalent to Q, and in which $Q'_{\mathrm{vis}} = \{v_0\}$. Note that in doing this, we perform a mutation at each vertex in $VQ_{\mathrm{vis}} - \{v_0\}$ exactly once, and at no other vertex. Now, v_0 is prominent in Q'. We can compute $H(Q')$ in terms of $H(Q)$ by repeated applications of (3.1):

$$H(Q') = H(Q) - 2\#(VQ_{\mathrm{vis}} - \{v_0\}) \tag{3.2}$$

Since v_0 is prominent in Q', define $Q_{\mathrm{elev}} := \mu_{v_0}(Q')$. The height of v_0 is 0 by definition, so 'eroding' v_0 increases the height of every other vertex by 2:

$$h^{Q_{\mathrm{elev}}}(v) = h^{Q'}(v) + 2 \qquad \forall v \in V - \{v_0\} \tag{3.3}$$

Combining (3.2) and (3.3) we can compare $H(Q)$ and $H(Q_{\mathrm{elev}})$:

$$\begin{aligned} H(Q_{\mathrm{elev}}) &= \sum_{v \in V} h^{Q_{\mathrm{elev}}}(v) \\ &= H(Q') + 2\#(V - \{v_0\}) \\ &= H(Q) + 2\#(V - VQ_{\mathrm{vis}}) \end{aligned}$$

So $H(Q_{\mathrm{elev}}) \geqslant H(Q)$, with equality if and only if $V = VQ_{\mathrm{vis}}$, but this is the case only if $Q = \check{Q}$. □

Proof of Proposition 3.18. Let Q_1 and Q_2 be two quivers with the same underlying tree, we perform a series of mutations at prominent vertices starting with Q_1 to reach Q_2. If $Q_1 \succ Q_2$ then we can apply Lemma 3.25 repeatedly, necessarily ending at Q_2 after finitely many mutations by the final part of Lemma 3.22.

Suppose therefore that $Q_1 \not\succeq Q_2$, then in particular $Q_1 \neq \check{Q}$ since $\check{Q} \succeq Q$ for all quivers Q with the same underlying graph. This also means that $H(\check{Q}) \geqslant H(Q)$ for all quivers Q. It follows that we can apply Lemma 3.26 a finite number of times to produce a quiver Q_1' which dominates Q_2, and then apply the argument from the first paragraph. □

Although we make no further use of the following, we remark that Proposition 3.18 can be partially generalised to non-treelike presentation quivers. There is no hope of a complete generalisation as the following example demonstrates. Figure 3.5 shows two mutation-Dynkin quivers with the same underlying weighted graph which are mutation equivalent, however every mutation of either quiver changes that underlying weighted graph.
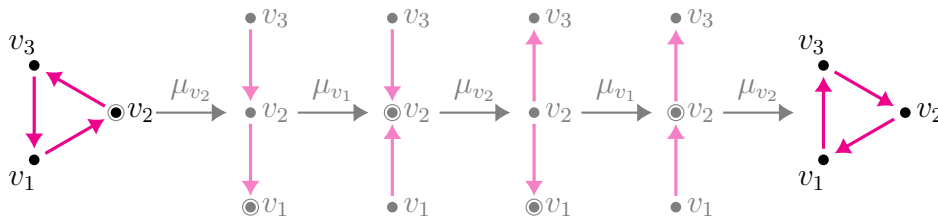
Figure 3.5: Two orientations of the same weighted graph, which are inequivalent under shape preserving mutations.

Definition 3.27. Given a quiver Q, call a vertex $v \in V$ **core in** Q if there is some chordless cycle in Q which passes through v. If we consider the sub-quiver obtained by deleting every edges which is contained in any chordless cycle, we get a quiver whose underlying graph is a forest which we call the **non-core sub-quiver of** Q.

We can now prove the following Corollary of Proposition 3.18.

Corollary 3.28: *Let (Q, ϕ) be a presentation quiver, and consider a component Q' of the non-core sub-quiver of Q. This is a tree in which we choose the distinguished vertex v_0 to be the unique core vertex in Q'. Let Q'' be another quiver with the same underlying graph as Q'. If Q' dominates Q'' then we can perform a sequence of shape preserving mutations at vertices away from v_0 which turn Q' into Q''. Moreover, this sequence can be chosen so that performing it on (Q, ϕ) leaves ϕ unchanged as well.*

The proof is the same as that for Proposition 3.18, except that we cannot perform mutations at v_0 since these are necessarily not shape preserving. This restriction means that we cannot perform elevations, only erosions, and hence Q' must dominate Q''.

3.4   Mutation versus reflection equivalence

In this Section we use the freedom to change the orientation on a treelike presentation quiver to prove that every presentation quiver is mutation equivalent to a standard presentation quiver.

3.4.1   Mutations of presentation Dynkin diagrams

By definition, Q is mutation equivalent to some orientation of $\overline{\mathcal{V}}$ and so, after applying a suitable sequence of mutations to (Q, ϕ), we reduce to the case that the underlying weighted graph of Q is $\overline{\mathcal{V}}$. By Proposition 3.18 we can change the orientation on Q freely without changing ϕ, and so we can forget these orientations and replace (Q, ϕ) by the presentation Dynkin diagram $(\overline{\mathcal{V}}, \phi)$, see Definition 3.17.

We want to define mutations of a presentation Dynkin diagram $(\overline{\mathcal{V}}, \phi)$. Such mutations must necessarily be induced by shape preserving mutations of some orientation (Q, ϕ) of $(\overline{\mathcal{V}}, \phi)$ so that we can continue to apply Proposition 3.18 afterwards. From the discussion following Definition 3.23, a mutation of Q at the vertex v is shape preserving if and only if all edges incident to v are pointing either towards v or all pointing away. If the latter, then the mutation does not change ϕ and so has no effect on $(\overline{\mathcal{V}}, \phi)$.

Pick a vertex v of $\overline{\mathcal{V}}$. We define $\mu_v(\overline{\mathcal{V}}, \phi)$ as follows. Choose an orientation of $\overline{\mathcal{V}}$, Q, such that every edge incident to v is pointing towards v. Now, mutating at v gives a new presentation quiver $\mu_v(Q, \phi) = (\mu_v(Q), \phi \circ \psi_v^{-1})$, and forgetting the orientation on $\mu_v(Q)$ we get a new presentation Dynkin diagram by defining $\mu_v(\overline{\mathcal{V}}, \phi) := (\overline{\mathcal{V}}, \phi \circ \psi_v^{-1})$. Note that this definition does not depend on the choice of the orientations on the edges of Q which do not meet v.

Definition 3.29. Let $(\overline{\mathcal{V}}, \phi)$ be a presentation Dynkin diagram, and $v \in V$, the **mutation** of $(\overline{\mathcal{V}}, \phi)$ at v is the presentation Dynkin diagram $(\overline{\mathcal{V}}, \phi \circ \psi_v^{-1})$ where $\psi_v \in \mathrm{Aut}(G_{\overline{\mathcal{V}}})$ is the automorphism of $G_{\overline{\mathcal{V}}}$ defined on V by

$$\psi_v(u) := \begin{cases} vuv^{-1} & \text{if } u \text{ is adjacent to } v \\ u & \text{otherwise} \end{cases}$$

We define **mutation equivalence** in the same way as Definition 3.12.

Remark 3.30 (Mutations and inner automorphisms). Note that ψ_v, as defined above, agrees with the isomorphism $G_Q \to G_{\mu_v(Q)}$ coming from Theorem 3.9, when all edges incident to v are directed towards v in Q. We call it an automorphism because there is a canonical isomorphism $G_{\mu_v(Q)} \to G_Q$ induced by μ_v, since μ_v is shape preserving (see Lemma 3.8). There is a natural identification of $G_{\overline{\mathcal{V}}}$ with G_Q since Q is an orientation of $\overline{\mathcal{V}}$. Therefore, we can view the composition of ψ_v with this canonical isomorphism as an automorphism of $G_{\overline{\mathcal{V}}}$.

If u is not adjacent to v in $\mathcal{V}$, then that means that u and v commute as elements of $W(\mathcal{V})$, so $u = vuv^{-1}$. We can therefore express ψ_v more concisely by

$$\psi_v : u \mapsto vuv^{-1} \in \mathrm{Inn}(G_{\overline{\mathcal{V}}}).$$

If two presentation Dynkin diagrams are mutation equivalent, then we can 'lift' the sequence of mutations to a sequence of mutations of presentation quivers by choosing some orientation on each of the presentation Dynkin diagrams. Between each lift of presentation Dynkin diagram, perform a suitable sequence of shape preserving mutations coming from Proposition 3.18. In this way, two presentation quivers which have $\overline{\mathcal{V}}$ as their underlying weighted graph are mu-

tation equivalent by a sequence of shape preserving mutations if and only if the corresponding presentation Dynkin diagrams are mutation equivalent.

Therefore, we need to show that all presentation Dynkin diagrams are mutation equivalent to a standard one.

Definition 3.31. Let $(\overline{\mathcal{V}}, \phi)$ be a presentation Dynkin diagram and $w \in G_{\overline{\mathcal{V}}}$ with $w = v_1 \cdots v_k$ for $v_i \in V$, then define the sequence of mutations

$$\mu_w(\overline{\mathcal{V}}, \phi) := (\mu_{v_k} \circ \cdots \circ \mu_{v_1})(\overline{\mathcal{V}}, \phi) = (\overline{\mathcal{V}}, \phi \circ \psi_{v_1}^{-1} \circ \cdots \circ \psi_{v_k}^{-1})$$

One can easily check that this is well-defined using Remark 3.30, since for any $w' \in G_{\overline{\mathcal{V}}}$

$$(\phi \circ \psi_{v_1}^{-1} \circ \cdots \circ \psi_{v_k}^{-1})(w') = \phi\left(w w' w^{-1}\right)$$

which depends only on the element w, and not on the word representing it. If we write $(\alpha_w : g \mapsto wgw^{-1}) \in \text{Inn}(G_{\overline{\mathcal{V}}})$ for the inner automorphism associated to w, then $\mu_w(\overline{\mathcal{V}}, \phi) = (\overline{\mathcal{V}}, \phi \circ \alpha_w)$.

3.4.2 Automorphisms of finite Coxeter groups

In order to understand when a presentation Dynkin diagram is mutation equivalent to a standard one, and when standard presentation Dynkin diagrams are equivalent, we need to understand the automorphism groups of Weyl groups. Here we summarise what happens for all irreducible finite Coxeter systems. In particular, we look at those automorphisms which preserve the set of reflections in a given Coxeter system for a Weyl group. The automorphism groups for irreducible finite Coxeter groups were computed by William Franzsen [50, 51].

Definition 3.32. Any graph automorphism of the Coxeter diagram of a Coxeter system (W, S) induces an automorphism of W, and we call such automorphisms **graph automorphisms** of W. An automorphism is called **inner by graph** if it lies in the subgroup $\text{Inn}(W) \rtimes \text{Gr}(W, S) \leqslant \text{Aut}(W)$ generated by the subgroup of inner automorphisms, together with the group of graph automorphisms of W.

Note that every such automorphism can be written as a product $\gamma \circ \alpha_w$ where $\gamma \in \text{Gr}(W, S)$ and $\alpha_w \in \text{Inn}(W)$ is the inner automorphism associated to some $w \in W$.

PROPOSITION 3.33 (Corollary 19 in [51]): *If the Coxeter diagram of a (possibly infinite) Coxeter system (W, S) is a forest, and all edge labels lie in the set $\{3, 4, 6\}$, then all automorphisms of W which preserve the set of reflections are inner by graph.*

Note that Weyl groups are precisely the finite irreducible Coxeter groups which have all edge labels in the set $\{3, 4, 6\}$. The Proposition below follows from the classification of automorphisms of irreducible finite Coxeter groups.

PROPOSITION 3.34 (Theorem 31 in [51]): *Let (W, S) be an irreducible finite Coxeter system and let $\alpha \in \mathrm{Aut}(W)$ preserve the set of reflections of (W, S).*

- *If (W, S) has type A_n, B_n, D_{2k+1}, or E_n then $\alpha \in \mathrm{Inn}(W)$.*

- *If (W, S) has type D_{2k} for $k \geqslant 3$, F_4, or G_2 then $\alpha \in \mathrm{Inn}(W) \rtimes \langle \gamma \rangle$ where γ is the non-trivial graph automorphism.*

- *If (W, S) has type D_4 then $\alpha \in \mathrm{Inn}(W) \rtimes S_3$ where S_3 is the symmetric group which is the group of graph automorphisms of $W(D_4)$.*

- *If (W, S) has type H_n then $\alpha \in \mathrm{Inn}(W) \rtimes \langle \rho \rangle$ where ρ is an exceptional automorphism coming from the fact that $W(H_n)$ does not satisfy Proposition 3.33.*

- *If (W, S) has type $I_2(m)$ then every automorphism preserves the set of reflections, and we can write $\mathrm{Aut}(W) = \mathrm{Aut}_1(W)\langle \beta \rangle$, where β cyclically permutes the reflections in the usual action of $W(I_2(m))$ on $\mathbb{S}^1$, and*

$$\mathrm{Aut}_1(W) = \{\alpha_h \mid \gcd(h, m) = 1\} \cong \mathbb{Z}_m^\times,$$

  where $\alpha_h(s_1) = s_1$ and $\alpha_h(s_2) = \beta^h(s_1)$ (taking $S = \{s_1, s_2\}$ and $s_2 := \beta(s_1)$).

Geometrically, $W(H_n)$ acts on an $(n-1)$-sphere and the generators are reflections in great circles. These bound a simplex with dihedral angles $\{\pi/2, \pi/3, \pi/5\}$. The exceptional automorphism ρ in the case of H_3 comes from choosing different generating reflections whose corresponding great circles bound a simplex with dihedral angles $\{\pi/2, \pi/3, 2\pi/5\}$ (this generating triple is shown in Figure 4.17d). The automorphism ρ is defined explicitly in the proof of Proposition 32 in [51].

3.4.3   The main Theorem

We can now state and prove the main result of this Chapter.

Proposition 3.35: *Let $(\overline{\mathcal{V}}, \phi)$ be a presentation Dynkin diagram, then it is mutation equivalent to a standard presentation Dynkin diagram.*

Proof. By Theorem 2.2, ϕ maps the set of reflections of the Coxeter system $(G_{\overline{\mathcal{V}}}, V)$ to the set of reflections of $W(\mathcal{V})$. It follows that there exists some automorphism $\alpha \in \mathrm{Aut}(W(\mathcal{V}))$ which maps $\phi(V)$ to $V\mathcal{V}$ bijectively. Proposition 3.33 guarantees that this automorphism can be decomposed as $\alpha = \gamma \circ \alpha_w$ for some graph automorphism $\gamma \in \mathrm{Gr}(W)$ and $\alpha_w \in \mathrm{Inn}(W(\mathcal{V}))$ for some $w \in W(\mathcal{V})$. Let $w' = \phi^{-1}(w)$, then we can apply the mutation

$$\mu_{w'}(\overline{\mathcal{V}}, \phi) = (\overline{\mathcal{V}}, \phi \circ \alpha_{w'}).$$

Now for any $v \in V$, $\phi(\alpha_{w'}(v)) = \phi(w'vw'^{-1}) = \alpha_w(\phi(v)) \in V\mathcal{V}$, and hence $\mu_{w'}(\overline{\mathcal{V}}, \phi)$ is standard.   □

This Proposition now immediately implies the following Theorem.

Theorem 3.36: *All presentation quivers of type $\mathcal{V}$ are mutation equivalent to a standard one.*

Corollary 3.37: *If $\mathcal{V}$ is A_n, B_n, D_{2k+1}, or E_n then all presentation quivers of type $\mathcal{V}$ are mutation equivalent.*

Proof. It suffices to show that any two standard presentation Dynkin diagrams are mutation equivalent in these cases. If $(\overline{\mathcal{V}}, \phi)$ and $(\overline{\mathcal{V}}, \phi')$ are two standard presentation Dynkin diagrams, then $\phi^{-1} \circ \phi'$ is a reflection preserving automorphism of $G_{\overline{\mathcal{V}}}$. By Proposition 3.34, this automorphism is inner, so we can find $w \in G_{\overline{\mathcal{V}}}$ such that $\phi' = \phi \circ \alpha_w$. It follows that $\mu_w(\overline{\mathcal{V}}, \phi) = (\overline{\mathcal{V}}, \phi')$ as required.   □

3.5   The type A case

We have shown that all presentation quivers of type A_n for fixed n are mutation equivalent, which answers Question 3.15 for this type. That still leaves Ques-

tion 3.14, which we answer here. Fix an identification of $W(A_n)$ with the symmetric group S_{n+1}; for concreteness we map the Coxeter generator s_i to the transposition $(i\,j+1)$ for each $1 \leqslant i \leqslant n$. Instead of starting with a presentation quiver (Q, ϕ) where we assume *a priori* that Q is mutation-Dynkin, we want to associate a pair (Q, ϕ) to a minimal reflection generating tuple T where Q is some (not necessarily mutation-Dynkin) quiver, and ϕ in some sense represents the marking associated to T.

3.5.1 The graph of a reflection generating tuple

The reflections in $W(A_n)$ correspond to the transpositions in S_{n+1}. Consider a tuple of n transpositions, $T = (t_1, \cdots, t_n)$, then we can associate a weighted graph $\overline{\mathcal{V}}_T$ to T as follows. Start with the same vertex set V of size n we have used throughout (recall V has been given some order), and define $\phi\colon V \to W(A_n)$ by mapping the ith vertex to t_i. $\overline{\mathcal{V}}_T$ will have an edge between vertices v and u whenever $\phi(v)$ and $\phi(u)$ do not commute. In that case $\phi(v)\phi(u) = (a\,b)(b\,c) = (a\,b\,c)$ has order 3. We therefore give this edge weight 1 (compare with the definition of $\overline{\mathcal{V}}$ in Definition 3.5. If $(W(A_n), T)$ is a Coxeter system, $\overline{\mathcal{V}}_T = \overline{\mathcal{V}}$).

LEMMA 3.38: *Let T be a tuple of reflections, and $\overline{\mathcal{V}}_T$ the associated weighted graph. Then T generates $W(A_n)$ if and only if $\overline{\mathcal{V}}_T$ is connected, and for each $1 \leqslant j \leqslant n+1$ there is some $t_i \in T$ such that $t_i(j) \neq j$.*

Proof. The only if direction is straightforward. For the if direction, we show that for each $1 \leqslant i \leqslant n$, T is reflection equivalent to a tuple which contains $s_i = (i\,i+1)$. It follow that the group generated by T contains $\{s_1, \ldots, s_n\}$ as a subset, and hence the whole of S_{n+1}.

Fix an index i, and let d be the minimum combinatorial distance in $\overline{\mathcal{V}}_T$ between vertices u and v such that $\phi(u)(i) \neq i$ and $\phi(v)(i+1) \neq i+1$. If $d = 0$, then $u = v$ and $\phi(u) = (i\,i+1)$, so T already contains s_i.

Assume therefore that $d > 0$, and consider a path between u and v of length d, and label the vertices $u = u_0, u_1, \ldots, u_d = v$. The sequence of transpositions given

by $\phi(u_0), \phi(u_1), \ldots, \phi(u_{d-1}), \phi(u_d)$ has the form

$$(i\, a_0), (a_0\, a_1), \ldots, (a_{d-2}\, a_{d-1}), (a_{d-1}\, i+1).$$

Since these are all elements of T and the path has minimal length, we know that $i, a_1, \ldots, a_{d-1}, i+1$ are all pairwise distinct. Indeed, if $a_j = a_k$ for some $j \neq k$, say, then $(a_j\, a_{j+1})$ and (a_{k-1}, a_k) do not commute, and the edge between u_{j+1} and u_k shortcuts the path. Denote the elements of this sequence by $t_{i_0}, t_{i_1}, \ldots, t_{i_d}$.

Now we transform T by setting $t'_{i_d} = t_{i_d} = (a_{d-1}\, i+1)$, and iteratively replacing t_{i_j} by its conjugate conjugate $t'_{i_j} = t'_{i_{j+1}} t_{i_j} t'^{-1}_{i_{j+1}}$, for $d-1 \geqslant j \geqslant 1$. Call the resulting tuple T'; one can check that the sequence of elements $t'_{i_0}, t'_{i_1}, \ldots, t'_{i_d}$ it contains has the form

$$(i\, i+1), (a_0\, i+1), \ldots, (a_{d-2}\, i+1), (a_{d-1}\, i+1).$$

Thus, T' is a tuple obtained from T by a sequence of partial conjugations, and contains $s_i = (i\, i+1)$, as required. □

The statement of this Lemma implicitly assumes that T is minimal since we defined the graph $\overline{\mathcal{V}}_T$ to have a vertex set of size n, however dropping this hypothesis, the proof still goes through unaltered.

3.5.2 Generalised presentation quivers and mutation modulo 2

Given a minimal reflection generating tuple, we want to know whether it is possible to orient $\overline{\mathcal{V}}_T$ to produce a mutation-Dynkin quiver of type A_n. For $n \leqslant 3$ one can check explicitly that this is always possible—in general it is not. As an example, let $n = 4$, and consider $T = ((1\, 2), (1\, 3), (1\, 4), (1\, 5))$. The associated weighted graph is complete, so by Lemma 3.38, T generates S_5, however any orientation of $\overline{\mathcal{V}}_T$ must contain an unoriented 3-cycle, see Figure 3.6. Thus, by Theorem 3.6, any orientation of $\overline{\mathcal{V}}_T$ is a quiver which is not mutation-Dynkin.

This problem can be repaired, however, by reducing edge weights modulo 2. We sketch this only in the case of type A.

Definition 3.39. Let T be a minimal reflection generating tuple for $W(A_n)$, then a **generalised presentation quiver** associated to T is a pair (Q, ϕ) where the underlying weighted graph of Q is $\overline{\mathcal{V}}_T$, and ϕ is the map from $V \to W(A_n)$ which maps
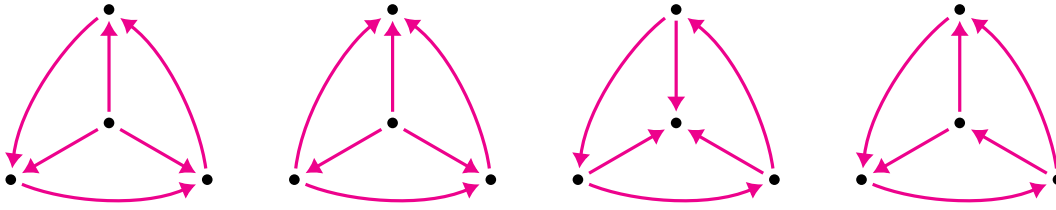
Figure 3.6: The possible orientations of the complete graph on four vertices up to graph automorphisms.

each vertex of Q to the corresponding element of T. For a vertex $v \in V$, define the **mutation of Q modulo** 2, $\mu_v^{(2)}(Q)$ to be the quiver obtained by first mutating Q at v as usual, and then deleting any edges of weight 2 which are produced. We extend this to a mutation of (Q, ϕ) modulo 2 by setting $\mu_v^{(2)}(Q, \phi) = (\mu_v^{(2)}(Q), \mu_v(\phi))$, where $\mu_v(\phi)$ is the map defined in Definition 3.12 restricted to V.

LEMMA 3.40: *Let (Q, ϕ) be a generalised presentation quiver associated to a generating tuple T, and for $v \in V$, let T_v be the image of $\mu_v(\phi)$ in $W(A_n)$. Then $\mu_v^{(2)}(Q, \phi)$ is athe generalised presentation quiver associated to T_v.*

Proof. We need to prove that the underlying weighted graph of $\mu_v^{(2)}(Q)$ is $\overline{\mathcal{V}}_{T_v}$, and $\mu_v(\phi)$ is the map associated to $\overline{\mathcal{V}}_{T_v}$. Recall that all edges in Q have weight 1.

First consider a neighbour u of v. If $\phi(v) = (a\,b)$ then without loss of generality we can assume that $\phi(u) = (b\,c)$ for some $c \neq a$. In $\mu_v^{(2)}(Q)$ there is still an edge between v and u with the same weight. It follows from the definition that $\mu_v(\phi)(v) = (a\,b)$, and $\mu_v(\phi)(u)$ is either $(b\,c)$ or $(a\,c)$. In either case there is still an edges between v and u with weight 1.

Now let w be another neighbour of v, and write $\phi(w) = (a\,d)$ or $(b\,d)$ for some d. The only case which we need to consider which is not covered by Theorem 3.9 is when v, u, and w form an unoriented cycle in Q. The minimality of T ensures that $\phi(w) = (b\,d)$ for $d \notin \{a, c\}$. This case is shown in Figure 3.7. All other vertices and edges of Q are not changed by $\mu_v^{(2)}$, so this completes the proof. $\qquad\square$

3.5.3 All presentation quivers are mutation-Dynkin modulo 2

Mutation modulo 2 does not, as stated, yield an equivalence relation on the space of generalised presentation quivers of type A_n. This is because we have not de-
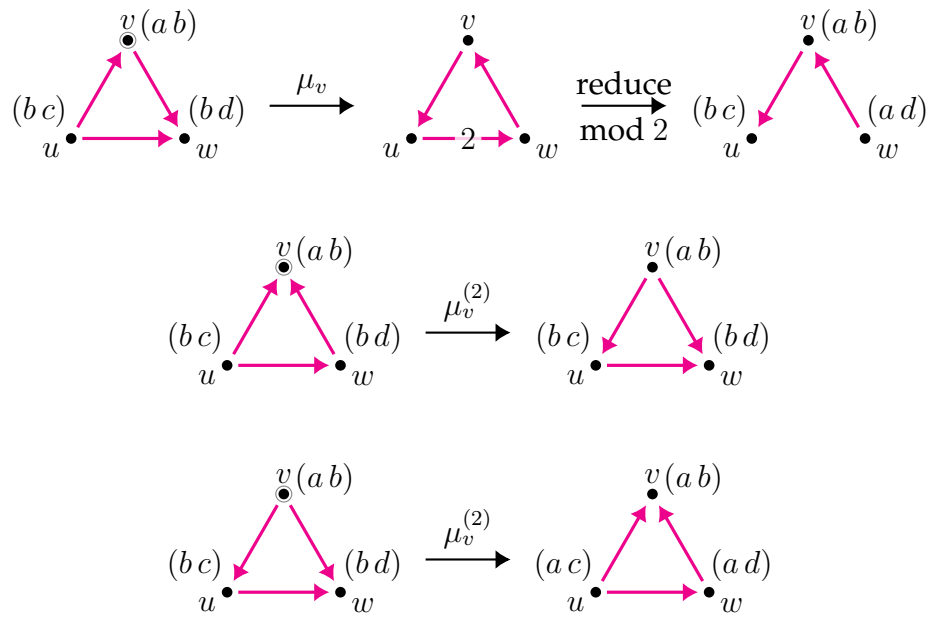
Figure 3.7: Mutation modulo 2 of an unoriented 3-cycle. The maps ϕ and $\mu_v(\phi)$ are represented by labelling each vertex by its image in $W(A_n)$.

fined an inverse to the mutation $\mu_v^{(2)}$ shown in the top row of Figure 3.7. It is possible to reinterpret all edge weights in a quiver as living in $\mathbb{Z}_2$, and define mutation with respect to this. Making this definition consistent, however, becomes quite *ad hoc* and messy, even when just restricting to quivers associated to minimal generating tuples of $W(A_n)$. We omit a detailed discussion of this as it is not clear how it might generalise outside of type A. We can however prove the following.

THEOREM 3.41: *Let (Q, ϕ) be a generalised presentation quiver of type A, then it is possible to perform a sequence of mutations modulo 2 to produce a generalised presentation quiver (Q', ϕ') such that Q' is mutation-Dynkin.*

In order to prove this we need to consider more closely the local geometry of weighted graphs associated to reflection generating tuples of $W(A_n)$.

Definition 3.42. Let T be a (minimal) reflection generating tuple, $\overline{\mathcal{V}}_T$ the associated weighted graph, and v a vertex of $\overline{\mathcal{V}}_T$. Then the **link** of v, denoted $\mathrm{lk}(v)$, is the subgraph of $\overline{\mathcal{V}}_T$ induced by the set of vertices which neighbour v.

LEMMA 3.43: *Let T be a minimal reflection generating tuple, and v a vertex of $\overline{\mathcal{V}}_T$, then $lk(v)$ has at most two connected components, each of which is a complete graph*

Proof. Let ϕ be the map $V \to W(A_n)$ associated to T, and write $\phi(v) = (a\,b)$. If u is a neighbour of v, then $\phi(u) = (a\,c)$ or $(b\,c)$ for some $c \notin \{a, b\}$. If $\phi(u) = (a\,c)$ and w is any other neighbour of v such that $\phi(w) = (a\,d)$ for some d, then u and w are connected in $\text{lk}(v)$. It follows that the connected component of $\text{lk}(v)$ containing u is a complete graph. Moreover, we can label each connected component either a or b depending on which appears in the ϕ-images of the vertices of that component. Consequently, there are at most two such components. $\qquad\square$

Proof of Theorem 3.41. The idea of the proof is to perform a sequence of mutations modulo 2 so that the resulting quiver is tree-like. Then we can argue that this tree must be a path graph, and hence $\overline{A}_n$. First we show that all chordless cycles have length 3. Removing unoriented cycles of length 3 is straightforward: it uses the top row of Figure 3.7 and Lemma 3.43. Call the result Q'. In order to remove the remaining oriented 3-cycles Q', we consider the *graph of oriented 3-cycles* in Q'. This is a graph whose vertex set is the set of 3-cycles in Q'. We show that this graph must be a forest. We give a procedure to prune this forest one leaf at a time until all 3-cycles are removed.

Bounding the length of cycles First we claim that Q contains no chordless cycles of length greater than 3. Assume such a cycle does exist, and cyclically label its vertices $v_0, v_1, \ldots, v_k$. Write $\phi(v_i) = (a_i, a_{i+1})$, where the indices are read modulo k. Since the cycle is chordless, the a_i's are pairwise distinct. But then the subgroup generated by $\phi(v_0), \ldots, \phi(v_k - 1)$ contains $\phi(v_k)$, contradicting the assumption that $\phi(V) = T$ is a minimal reflection generating tuple.

Removing unoriented 3-cycles Let $u \bullet\!\!-\!\!\overset{v}{\bullet}\!\!-\!\!\bullet w$ be a path of length 2 in Q, carrying some orientation. If this path is not part of a 3-cycle or is part of an oriented 3-cycle, then it follows from the definition that mutation modulo 2 at v does not create any unoriented cycles. If it is part of an unoriented 3-cycle, then the mutation is one of those shown in Figure 3.7, which does not increase the number of unoriented 3-cycles. By performing a sequence of mutations of the first type shown in that Figure, therefore, we can inductively remove all unoriented 3-cycles.

Now assume that Q does not contain any unoriented 3-cycles. Fix a vertex v, and consider its link. By Lemma 3.43, this link has at most two components, both of which are complete graphs. If a component contains a triangle, then Q contains a complete graph with four vertices. It then follows from Figure 3.6 that Q contains an unoriented cycle, which is a contradiction. Therefore, each component of $\mathrm{lk}(v)$ is a vertex or an edge. We define a sequence of mutations modulo 2 to remove the oriented cycles. These mutations do not introduce any edges of weight 2, and so are simply mutations.

The graph of 3**-cycles** Define a new graph whose vertices are the oriented 3-cycles in Q, and two vertices are joined by an edge if the corresponding 3-cycles share a vertex (by the Lemma two 3-cycles cannot share an edge). If this graph itself contains a chordless k-cycle, then this corresponds to a k-cycle of connected 3-cycles in Q. Since Q contains no edge loops or double edges, k cannot be 1 or 2. If k is 3, then Q contains two 3-cycles which share an edge, which is not possible. Finally if $k > 3$, assume that k is minimal. Then Q contains a k-cycle, c, made up of the vertices where adjacent 3-cycles meet. Since Q contains no chordless cycles of length greater than 3, this k-cycle is short-cut by an edge which forms a 3-cycle with two of the edges from c. This new 3-cycle short-cuts the k-cycle of 3-cycles, contradicting the minimality of k. These cases are illustrated in Figure 3.8.
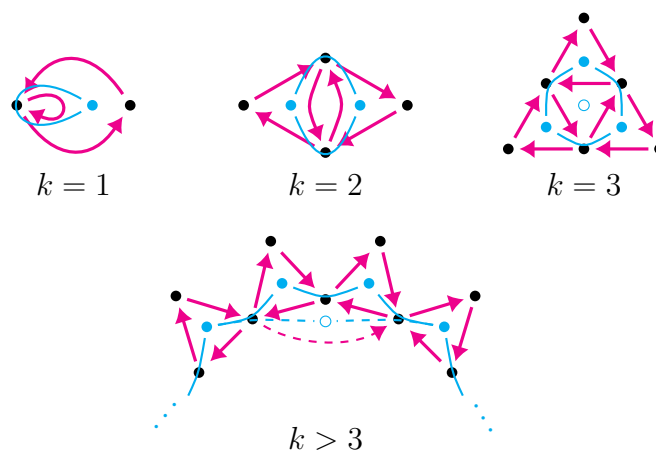
Figure 3.8: Proof that a quiver of type A cannot contain a cycle of 3-cycles. The quiver is shown in red, and overlaid in blue is the graph of 3-cycles.

This shows that this graph formed using the 3-cycles in Q is acyclic and hence a

finite forest. Augment this graph by adding an edge between two vertices if there
is some path in Q connecting the corresponding 3-cycles, and none of the edges in
that path lies in a 3-cycle. We claim that this new graph is still a forest. Indeed, a
similar argument which ruled out k-cycles for $k > 3$ above works here. A k-cycle
in this new graph descends to an ℓ-cycle in Q for some $\ell > k$ by taking the union of
the paths connecting adjacent 3-cycles, together with an edge from each 3-cycle.
This ℓ-cycle must be short-cut by an edge which forms a 3-cycle with at least one
of the edges of one of the connecting paths, contradicting the definition of these
paths. This is illustrated in Figure 3.9.
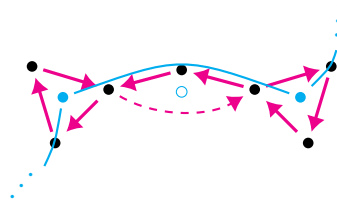
Figure 3.9: Part of a cycle in the augmented graph of 3-cycles. Again, the quiver is shown
in red, and overlaid in blue is the graph of 3-cycles.

Pruning the forest   Pick a vertex with valence 0, 1, or 2 in this forest, and we
show that the corresponding 3-cycle can be removed to decrease the total number
of 3-cycles. Indeed, this 3-cycle must have a vertex v which is not connected to any
other 3-cycles by a path not containing an edge from a 3-cycle. Thus, either v has
valence 2 or, following any path away from v, we eventually reach a vertex u of Q
with valence 1. In fact, by applying Lemma 3.43, if v has valence greater than 2,
it must have valence 3 and there is a unique path leading away from v. If v has
valence 2, then the mutation μ_v removes the 3-cycle without creating a new one,
see the first row in Figure 3.10.

Otherwise, let $d \geqslant 1$ be the length of the path connecting v to the valence 1
vertex u. The mutation at v replaces the 3-cycle with a new one which is closer
to u. Hence, the total number of 3-cycles remains constant, but d decreases. By
induction on d, we can decrease the number of 3-cycles by 1 after $d + 1$ mutations,
see the second row of Figure 3.10. Repeating this for each oriented 3-cycle we are

able to remove each one to leave a quiver Q' which is a tree.
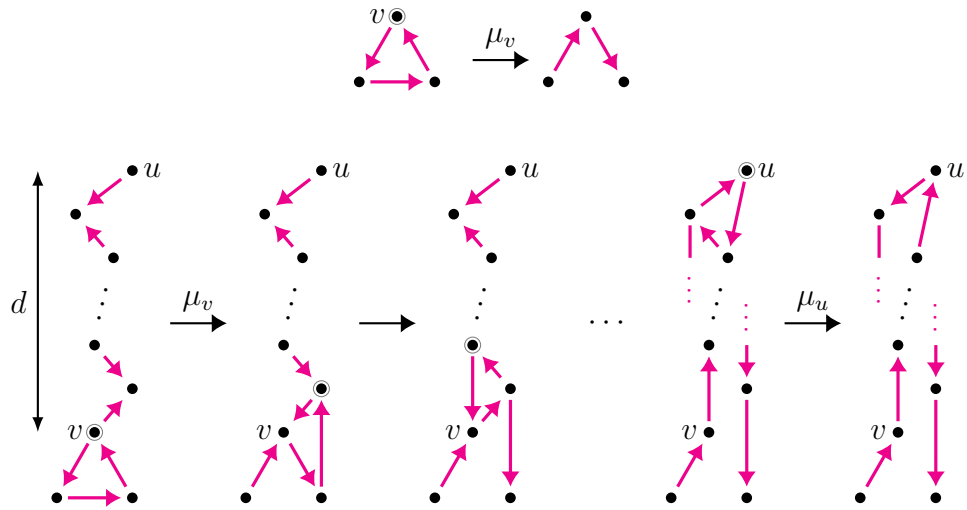
Figure 3.10: Removing an oriented cycle from Q.

The underlying tree must be a path   Since Q' contains no cycles, every component of the link of a vertex is a point. By applying Lemma 3.43, we can conclude that every vertex in Q' has valence 1 or 2 (assuming $n > 1$). Therefore, the underlying graph of Q' is a path of length n—in other words, the underlying weighted graph of Q' is $\overline{A}_n$, and Q' is mutation-Dynkin.   □

We can immediately conclude from this Proposition, Proposition 3.16, and Corollary 3.37 the following result.

COROLLARY 3.44: *Any generalised presentation quiver of type A can be transformed into a standard presentation quiver by a sequence of mutations modulo 2, and hence every minimal reflection generating tuple of A_n is reflection equivalent to any other.*

We have shown that, while the answer to Question 3.14 is no, we can generalise the notion of mutation somewhat to make it yes in the type A case. Moreover, the answer to Question 3.15 remains yes in type A in this generalised setting.

It is not immediately clear whether this generalisation can be extended to other types, but that would be an interesting question to try to answer. Another question which should be considered is whether this generalisation can be reinterpreted in the original context of cluster algebras.

Chapter 4

Reflection equivalence for arbitrary Coxeter systems

As we remarked in Section 2.1, there are two types of question one can ask about Nielsen equivalence in Coxeter groups. In this Chapter we discuss the second: given a Coxeter system (W, S) with reflections R, classify all generating tuples for W with entries in R up to reflection equivalence. Our main tool is the *Davis complex* associated to a Coxeter system (W, S). This is a CAT(0) CW complex on which W acts discretely by isometries such that the elements of S act by the natural generalisation of the notion of reflections in this context.

The fixed sets of the reflections in W are *hyperplanes*, so to any tuple of reflections we can associate a hyperplane arrangement. We use the geometry of this hyperplane arrangement to give a new proof of a criterion for when a tuple of reflections form a Coxeter generating tuple for (a reflection subgroup of) W. This gives us a terminating condition for an algorithm which takes in a tuple of reflections and outputs a Coxeter system for the Coxeter (sub)group they generate. This gives an effective method of testing whether a tuple of reflections in (W, S) generate W, and if not, what the index of the subgroup they generate is.

Modifying this algorithm slightly allows us to study reflection equivalence. Given a tuple of reflections this new algorithm returns another tuple of reflections which is reflection equivalent to the original one, but which is in some sense geometrically and algebraically simplest (Theorem 4.44). This gives, at least in prin-

ciple, a method to classify all generating tuples of reflections for a given Coxeter system. It also proves that for many Coxeter systems (including all Weyl groups and RACGs) there is only one reflection equivalence class of generating tuples of reflections (Corollary 4.45). Moreover, it allows us to prove that in any Coxeter system, after performing a single stabilisation, every generating tuple of reflections is reflection equivalent to some stabilisation of S, see Theorem 4.47.

4.1 The basic construction and the Davis complex

First we review the so-called *basic construction* and its application to defining the Davis complex of a Coxeter system. We follow the exposition in Chapters 5 and 7 of [32] in part. The purpose of the Davis complex is to generalise the natural geometric action of a *geometric reflection group* on a space of constant curvature to arbitrary Coxeter systems. We make heavy use of (abstract) simplicial complexes, so we summarise the relevant definitions and notation in Section 4.A at the end of this Chapter. Note Assumption 4.62: we do not distinguish between an abstract simplicial complex and the corresponding geometric simplicial complex.

4.1.1 Motivation: geometric reflection groups

A Coxeter group W which acts discretely, co-compactly, and by isometries on a space of constant curvature $\mathbb{X}^n$ such that a tuple of Coxeter generators S act by reflections is called a *geometric reflection group*. The reflection hyperplanes from S bound an acute-angled polytope which is a *strict fundamental domain*.

Definition 4.1. Let a group G act on a topological space X by homeomorphisms. A subset $\mathcal{F} \subset X$ is called a **fundamental domain** for G if

- It is open and connected;

- Every G-orbit intersects $\overline{\mathcal{F}}$, the closure of $\mathcal{F}$, in at least one point; and

- Whenever a G-orbit intersects $\overline{\mathcal{F}}$ at a point in $\mathcal{F}$, then this is the unique point of intersection with $\overline{\mathcal{F}}$.

A fundamental domain is **strict** if the third point holds whenever a G-orbit intersects the boundary $\partial \mathcal{F}$ as well.

We now explore an extended example of a geometric reflection group which exhibits all of the essential features the Davis complex is designed to capture.

Example 4.2. Consider the example of the Coxeter system (W, S) with presentation diagram:
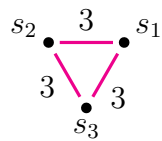
The group W acts on $\mathbb{E}^2$, the Euclidean plane, with strict fundamental domain an equilateral triangle, where the generators s_1, s_2, and s_3 act by reflections in the sides of this triangle. We call this fundamental domain the fundamental chamber C.
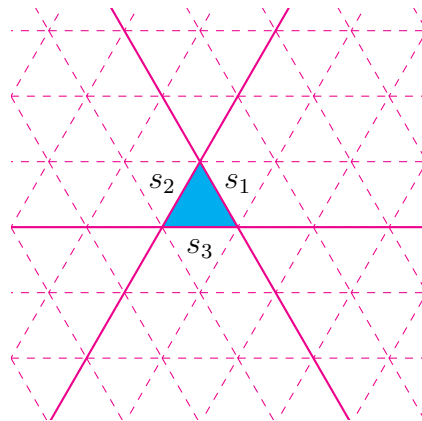
Figure 4.1: The action of a geometric reflection group on $\mathbb{E}^2$. The dashed lines represent the reflection lines of the reflections, the solid lines represent the reflection lines for the Coxeter generators, and the blue triangle represents the strict fundamental domain bounded by these lines.

The fundamental chamber C can be given a simplicial structure by identifying it with a 2-simplex. Since the W-translates of the fundamental domain tile $\mathbb{E}^2$, we can equivariantly extend this to a simplicial structure on the whole plane.

Since the fundamental domain is strict, the stabiliser of each simplex fixes that simplex point-wise. Labelling each simplex by the special subgroup which stabilises it, we get Figure 4.2. If we include the empty simplex which is labelled by

$W_{\{s_1,s_2,s_3\}} = W_S = W$ then the labelling corresponds to a bijection between the simplices in c and the power set of S. Moreover, this bijection has the following additional property: given two simplices σ and τ with labels $\lambda(\sigma)$ and $\lambda(\tau)$, $\sigma \subseteq \tau$ if and only if $\lambda(\sigma) \supseteq \lambda(\tau)$.
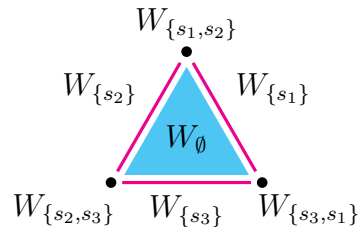
Figure 4.2: The fundamental chamber with simplices labelled by the Coxeter generators which fix them.

Just as we extended the simplicial structure of C to the whole plane, we can extend the labelling to the whole simplicial structure. Every simplex is a W translate of some simplex in C. Let $w\sigma$ be such a simplex, where $w \in W$ and $\sigma \subset C$. Then the stabiliser of $w\sigma$ is the parabolic subgroup $w\lambda(\sigma)w^{-1}$, and we can label $w\sigma$ by the coset $\lambda(w\sigma) := w\lambda(\sigma) \in \lambda(\sigma)\backslash W$. This labelling does not depend on the choice of representation $w\sigma$ for the simplex and still has the property that, for any two simplices σ and τ in $\mathbb{E}^2$ with labels $\lambda(\sigma)$ and $\lambda(\tau)$, $\sigma \subseteq \tau$ if and only if $\lambda(\sigma) \supseteq \lambda(\tau)$.

We have two posets (see Definition 4.54): the poset of simplices in the simplicial structure on $\mathbb{E}^2$ with simplices ordered by inclusion,

$$(\text{simplices in } \mathbb{E}^2, \subseteq);$$

and the poset of cosets of special subgroups of (W, S) again ordered by inclusion,

$$(\text{left cosets of } W_T \text{ for } T \subseteq S, \subseteq).$$

The labelling λ can be viewed as a poset isomorphism from the first poset to the second with the opposite order relation to inclusion, see Definition 4.56.

One natural way to attempt to generalise this to an arbitrary Coxeter system (W, S) is to start with the set of all left cosets of special subgroups of W, turn this

into a poset by ordering cosets with the opposite ordering to that given by inclusion (see Definition 4.56), and then replace $\mathbb{X}^n$ by the simplicial complex which has this poset as its poset of simplices. The resulting simplicial complex has a natural W-action defined by $w \cdot uW_T = (wu)W_T$. Furthermore, the subcomplex whose simplices are special subgroups forms a strict fundamental domain for the action. This simplicial complex is called the *Coxeter complex* of (W, S) and plays an important role in the theory of *Tits buildings*, see [16].

For our purpose, the Coxeter complex is in some sense *too big*—including cosets of all special subgroups causes problems with the infinite special subgroups. Instead, the Davis complex essentially deletes the simplices which are cosets of infinite special subgroups. We shall give two definitions, one based on the so-called *basic construction*, and one based on cosets. The basic construction starts with a simplicial complex K which ends up being a strict fundamental domain, and 'unfolds' or 'develops' a complex with a W-action according to a *mirror structure* on K, and a family of subgroups (the finite special subgroups). The coset construction models the Coxeter complex definition above with the infinite special subgroups excluded.

4.1.2 The basic construction

Definition 4.3. Let K be a simplicial complex, and S a finite set. A **mirror structure** on K is a collection of subcomplexes called **mirrors** $\{K_s \mid s \in S\}$ of K indexed by S. For a point $x \in K$, write $S(x) := \{s \in S \mid x \in K_s\}$.

A **family of groups** over S consists of a discrete group G and a collection of subgroups $\{G_s\}_{s \in S}$, again indexed by S. For any $T \subset S$, write G_T for the subgroup generated by $\{G_t \mid t \in T\}$.

In particular, if $T \subset T'$ then $G_T \leqslant G_{T'}$. The basic construction takes as input a mirrored space and family of groups over a set S and outputs a simplicial complex on which G acts with strict fundamental domain K.

Definition 4.4. Let G be a group with the discrete topology and let S be a finite set. Let K be a simplicial complex with a mirror structure over S, and $\{G_s\}_{s \in S}$ a

family of groups for G. Consider the product $G \times K$ with the product topology. Let $\sim$ denote the equivalence relation given by

$$(g, x) \sim (g', x') \text{ if and only if } x = x', \text{ and } g' \in gG_{S(x)}.$$

Then define the quotient space

$$\mathcal{U}(G, K) := (G \times K)/\sim.$$

Points in $\mathcal{U}(G, K)$ are denoted by $[g, x]_{\mathcal{U}(G,K)}$; we drop the subscript if the space $\mathcal{U}(G, K)$ is clear from context. The group G acts on $\mathcal{U}(G, K)$ by $h[g, x] = [hg, x]$, and there is a natural inclusion $i : K \to \mathcal{U}(G, K) : x \mapsto [1, x]$. Identifying K with its image, its G-translates are called **chambers**, and K itself is the **fundamental chamber**.

Before using the basic construction to define the Davis complex, we illustrate it in a simpler setting by using it to reconstruct the simplicial structure on $\mathbb{E}^2$ we found in Example 4.2. Note that the complex we construct in this example is *not* the Davis complex.

Example 4.5. Let (W, S) be the Coxeter system in Example 4.2, and define the family of groups over S to be $W_{s_i} = W_{\{s_i\}}$, ie the special subgroup generated by s_i. Let K be a 2-simplex, and we give it a mirror structure over S by letting the three edges be K_{s_i} for $i = 1, 2, 3$.
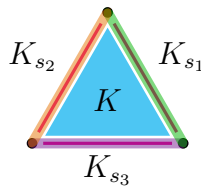
Figure 4.3: The fundamental chamber with its mirror structure.

Now we form the product $W \times K$ which consists of an infinite collection of disjoint copies of K. Finitely many of these have been arranged suggestively in Figure 4.4 with the mirror structure highlighted.

Consider one of the copies of K in this product: $(s_1 s_2, K)$. Taking a point x in the interior of K_{s_1}, $S(x) = \{s_1\}$ so $G_{S(x)} = \{e, s_1\}$. We can ask which points
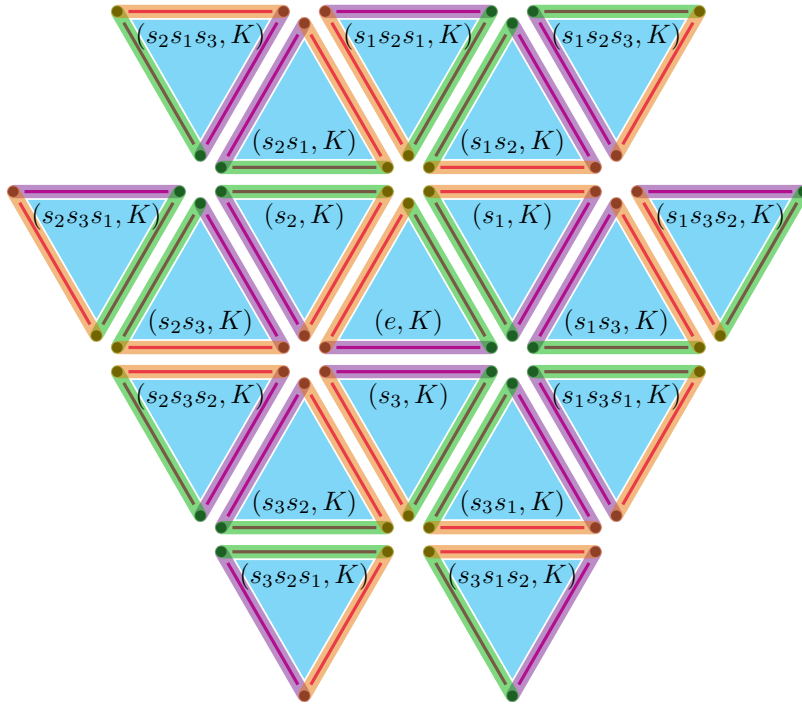
Figure 4.4: A finite portion of the product $W \times K$.

(s_1s_2, x) is identified with. They are points (w, x) such that $w \in s_1s_2G_{S(x)}$ which
equals $\{s_1s_2, s_1s_2s_1\}$. Hence (s_1s_2, x) is only identified with itself and $(s_1s_2s_1, x)$.
Therefore, (s_1s_2, K) is glued to $(s_1s_2s_1, K)$ along their respective copies of K_{s_1}
(shown in green in the Figure).

Now take the point $y \in K_{s_1} \cap K_{s_2}$. Then $S(y) = \{s_1, s_2\}$ so

$$G_{S(y)} = \{e, s_1, s_1s_2, s_1s_2s_1, s_2s_1, s_2\} = s_1s_2G_{S(y)}.$$

The point $(s_1s_2, y) \in (s_1s_2, K)$ is identified with each of the points (e, y), (s_1, y),
(s_1s_2, y), $(s_1s_2s_1, y)$, (s_2s_1, y), and (s_2, y).

In this way we can form the quotient $\mathcal{U}(W, K)$ by gluing together edges and
vertices to recover the simplicial structure on $\mathbb{E}^2$, see Figure 4.5.

4.1.3   First definition of Σ: the basic construction

We can now use the basic construction to define the *Davis complex* of a Coxeter
system (W, S). First we must define the complex K which we do via the *nerve* of
(W, S). This is done using the set of *spherical* subsets of S
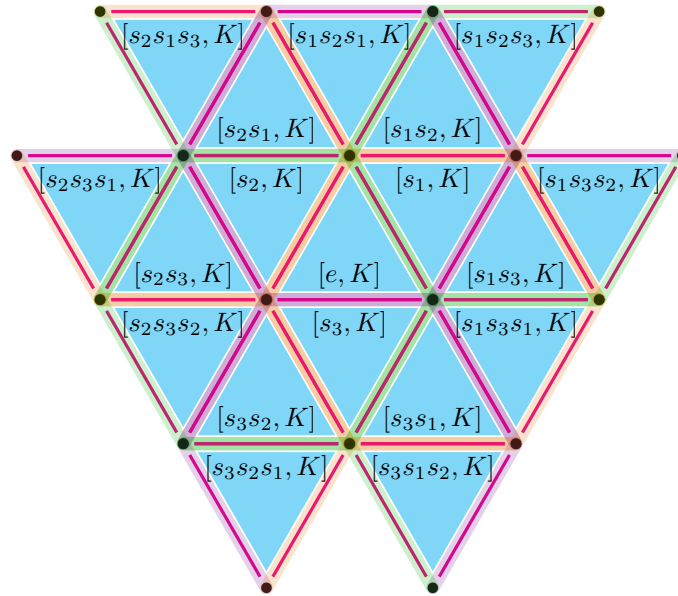
Figure 4.5: A finite portion of the quotient space $\mathcal{U}(W, K)$, in particular the chambers corresponding to elements of W with length at most 3.

Definition 4.6. A subset $T \subset S$ is **spherical** if the special subgroup W_T is finite (see Definition 1.3). Denote by $\mathcal{S} = \mathcal{S}(W, S)$ the set of spherical subsets of S, which is a poset ordered by inclusion.

Since any subset of a spherical subset is again a spherical subset, the poset $\mathcal{S}_{\supset\emptyset}$ of non-empty spherical subsets of S is an abstract simplicial complex, see Definition 4.60.

Definition 4.7. The **nerve** of a Coxeter (W, S) is the complex $L = L(W, S)$ whose poset of simplices is $\mathcal{S}_{\supset\emptyset}$. Define the **fundamental chamber** $K = K(W, S)$ for (W, S) to be the geometric realisation of the poset $\mathcal{S}$, see Definition 4.63.

The nerve can be thought of as a combinatorial encoding of how the non-trivial spherical subgroups of (W, S) intersect with one another. It is possible to construct K more concretely from the presentation diagram Γ (recall Definition 1.2) via the nerve L. Starting with Γ, attach a k-simplex wherever the 1-skeleton of a k-simplex, ie a complete $(k+1)$-subgraph, appears as a subgraph of Γ and the special subgroup corresponding to this complete graph is spherical. This complex is isomorphic to L. Then K is formed by taking the barycentric subdivision of L and then taking the cone over this complex, see Section 4.A.4, $K = \text{Cone}(\text{Bs}(L))$.

We can illustrate this with our running example.

Example 4.8. The presentation diagram Γ is shown at the start of Example 4.2. We will compute $K(W, S)$ both using the definition, and via L. The poset $\mathcal{S}$ is shown in Figure 4.6. Its geometric realisation is shown on the right of the Figure.
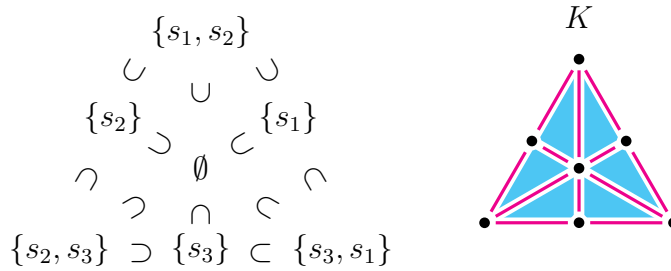
Figure 4.6: The poset $\mathcal{S}$ and the corresponding complex K.

Alternatively, starting with Γ, it contains the 1-skeleton of a 2-simplex, however the corresponding special subgroup is the whole of W which is infinite. Therefore in this case $L = \Gamma$. To form K we take the cone over the barycentric subdivision of L. This yields the same complex as above.

The relation $\subseteq$ on the simplices of L leads to an ordering of the vertices of the barycentric subdivision (see Remark 4.66) which we use to give K a mirror structure over S. In particular, let K_s be union of the closed simplices in the barycentric subdivision of L (thought of as a subcomplex of K) which have $\{s\}$ as their minimum vertex. We also choose a family of groups over S by setting $G = W$ and $G_s = W_{\{s\}}$.

Definition 4.9. The **Davis complex** of the Coxeter system (W, S) is

$$\Sigma = \Sigma(W, S) := \mathcal{U}(W, K(W, S)).$$

Example 4.10. Finally, let us construct the Davis complex for the running example. The mirror structure on K defines K_{s_i} to be the union of simplices of K—which are chains in the poset $\mathcal{S}$—which have $\{s_i\}$ as their minimum vertex. This gives the mirror structure in Figure 4.7.

Compare this with Figure 4.3. It is straightforward to see that applying the basic construction will give the barycentric subdivision of Figure 4.5 in this case.
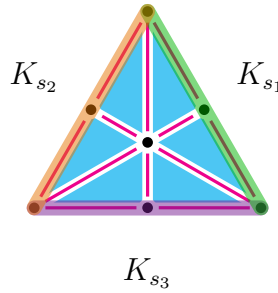
Figure 4.7: The mirror structure on K.

In general there is not such a simple relationship between the Coxeter complex of (W, S) and the Davis complex.

Example 4.11. As a simple example, consider the Coxeter system (W, S) with Coxeter presentation

$$\langle\, s_1, s_2, s_3 \mid s_1^2, s_2^2, s_3^2 \,\rangle.$$

This group acts on the hyperbolic plane, and its Coxeter complex can be visualised as the tiling of the hyperbolic plane by ideal triangles. At the ideal vertices of these triangles, the Coxeter complex fails to be locally finite owing to the fact that the special subgroups generated by (s_i, s_j) for $i \neq j$ are infinite (in particular, infinite dihedral groups). The construction of the Davis complex removes the ideal vertices and leaves behind a graph: the barycentric subdivision of the regular trivalent tree, see Figure 4.8.

4.1.4 Second definition of Σ: cosets

We already defined the fundamental chamber as the geometric realisation of the poset $\mathcal{S}$, in fact we can define the Davis complex in this way as well.

Definition 4.12. Denote by $W\mathcal{S}$ the set of leftcosets of spherical subgroups of (W, S). In other words

$$W\mathcal{S} \coloneqq \bigcup_{T \in \mathcal{S}} W/W_T.$$

This again is a poset ordered by inclusion. We define the **Davis complex** of (W, S), $\Sigma(W, S)$, to be the geometric realisation of this poset.
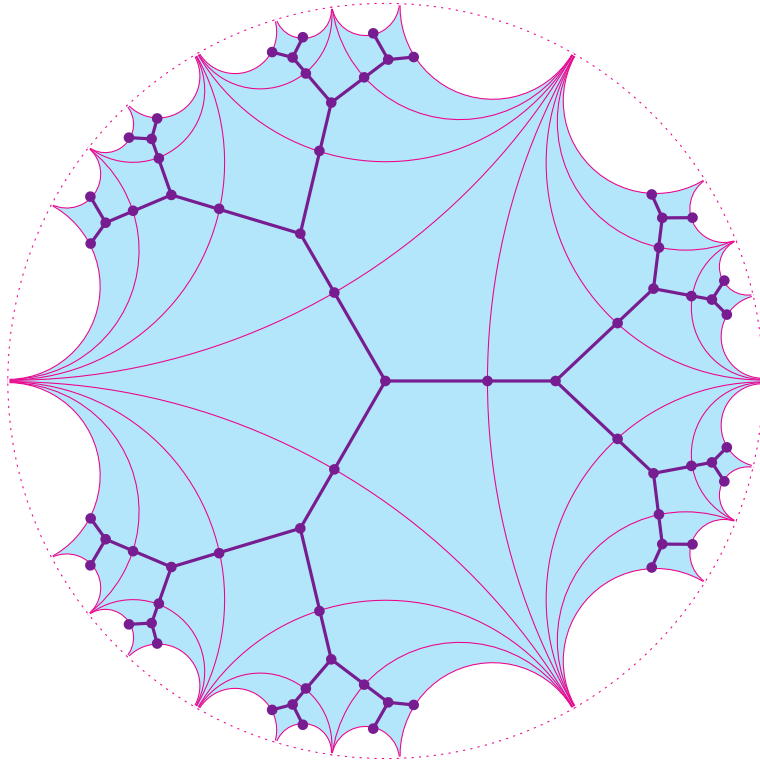
Figure 4.8: A representation of a finite part of the Coxeter complex of (W, S) in the hyperbolic plane. The corresponding finite part of the Davis complex is overlaid in purple, in particular it shows all chambers corresponding to elements of W of length at most 3.

There is an injective map $\mathcal{S} \hookrightarrow W\mathcal{S} \colon T \mapsto W_T$ which preserves the order relation. This induces an injective map $K(W, S)$ into the geometric realisation of $W\mathcal{S}$ whose image is again called the *fundamental chamber*. The proof that these two definitions are equivalent can be found as Theorem 7.2.4 in [32], see also Remark 4.32.

The action of W on this formulation of the Davis complex is straightforward to define. W acts on $W\mathcal{S}$ by left multiplication: $w \cdot uW_T = (wu)W_T$, and this induces an action in the geometric realisation.

Example 4.13. We give one more example, this time with a Coxeter system which does not act nicely on a space of constant curvature. Consider the Coxeter system with presentation diagram as shown in Figure 4.9. The special subgroup generated by (s_1, s_2, s_3) is the symmetry group of the 3-cube, while the special subgroup generated by (s_3, s_4, s_5) is the group we considered in the running example above. Overall, the group W is the amalgamated product of these two groups over $\langle s_3 \rangle$.

In the Figure we have also shown the fundamental chamber K with its mir-

rors. Visualising $\Sigma(W, S)$ is harder, but it can be viewed as a tree of spaces, where the underlying tree is 2-coloured, say black and white. Each black vertex corresponds to a copy of $\mathbb{E}^2$ tiled by equilateral triangles and has infinite valence. Each white vertex corresponds to a 3-cube and has valence 48. Each edge in the tree corresponds to where a cube and a plane intersect in an edge.
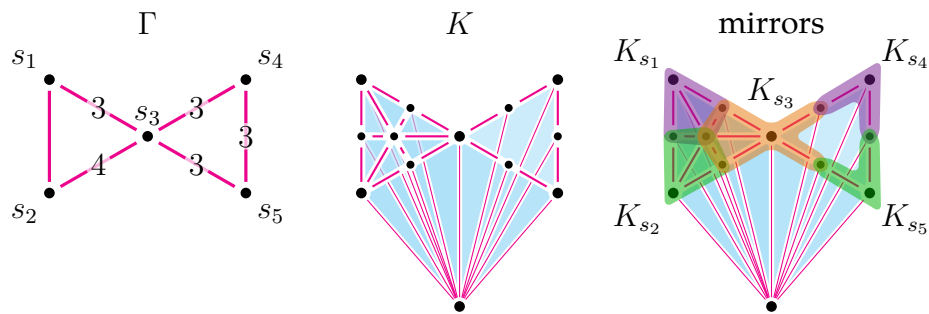
Figure 4.9: A presentation diagram Γ and the corresponding fundamental chamber K. On the right, the mirror structure is overlaid on K.

4.1.5 Basic properties of the Davis complex

We summarise some of the useful properties of Σ.

Theorem 4.14 (Lemma 5.3.3, Proposition 7.3.4, and Theorem 12.3.3 in [32])**:** *For any Coxeter system, Σ is contractible, and each element of $r \in R(W, S)$ fixes a subcomplex Σ_r which separates Σ into two components. Moreover Σ admits a 'dual' CW structure (the cells of which are called* Coxeter polytopes*) such that if each cell is given its natural Euclidean metric*

1. *The piece-wise Euclidean metric on Σ is CAT(0)*

2. *The action of W on Σ is by isometries*

3. *The 1-skeleton of this CW structure is the Cayley graph of (W, S) (see Definition 4.67)*

In addition Σ satisfies a certain *universal property* with respect to spaces on which W acts, however we postpone the statement, where we give it in slightly greater generality in Theorem 4.31.

Definition 4.15. The complex Σ_r, for $r \in R$, is called the **hyperplane** associated to r. Similar to Theorem 1.8, write Σ_r^+ for the component of $\Sigma - \Sigma_r$ which contains the interior of K and write Σ_r^- for the other component.

A useful result we will use throughout this Chapter is the following.

LEMMA 4.16: *For two reflections r, r', $\Sigma_r \cap \Sigma_{r'} \neq \emptyset$ if and only if rr' has finite order.*

Proof. In one direction, if they meet at some point, this point is fixed by rr'. But by the construction of Σ, rr' must be conjugate into some spherical subgroup which is finite.

Conversely, assume that $\Sigma_r \cap \Sigma_{r'} = \emptyset$. There are exactly three connected components of $\Sigma - (\Sigma_r \cup \Sigma_{r'})$, only one of which meets both hyperplanes—after a suitable conjugation we can assume that this component is $\Sigma_r^+ \cap \Sigma_{r'}^+$. Then (using the notation for conjugation introduced in Notation 2.22)

$$rr'\Sigma_{r'} = \Sigma_{rr'r'} = \Sigma_{rr'} \subset \Sigma_r^-, \text{ and}$$

$$rr'\Sigma_r = \Sigma_{rr'r} = \Sigma_{(rr')r} \subset \Sigma_{rr'}^- \subset \Sigma_r^-.$$

Notice also that $^{(rr')}r^r r' = rr'$. Setting $r_1 = {}^{(rr')}r$ and $r_1' = {}^r r'$ we can repeat the argument to produce $r_2 = {}^{(r_1 r_1')}r_1$ and $r_2' = {}^{r_1} r_1'$ such that $r_2 r_2' = rr'$ and

$$\Sigma_{r'}^+ \supset \Sigma_r^- \supset \Sigma_{r_1'}^- \supset \Sigma_{r_1}^- \supset \Sigma_{r_2'}^- \supset \Sigma_{r_2}^-.$$

This is illustrated in Figure 4.10. Inductively we can produce an infinite sequence of reflections $\{r_i, r_i'\}_{i=1}^\infty$ which are all distinct since the half-spaces bounded by their fixed sets are properly nested. Moreover, $r_i = {}^{(rr')^i}r$, implying the orbit of the hyperplane Σ_r under $\langle rr' \rangle$ is infinite. It follows that rr' has infinite order.   □

PROPOSITION 4.17: *Let (W, S) be a Coxeter system with length function ℓ. Then for any $w \in W$*

$$\ell(w) = \#\{r \in R(W, S) \mid wK \subset \Sigma_r^-\}.$$

This result with Σ replaced by the Coxeter complex (which we introduced briefly in Section 4.1.1) is proved as Theorem II.1 in [16]. The same proof works
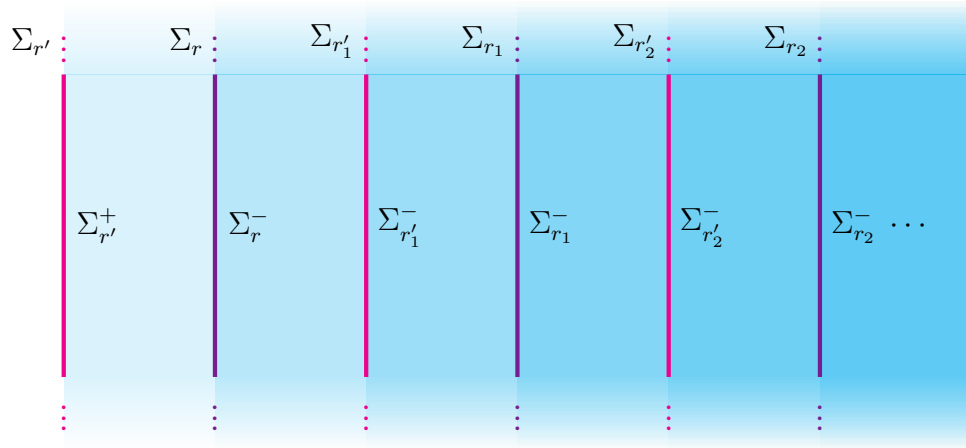
Figure 4.10: The hyperplanes produced by translation Σ_r and $\Sigma_{r'}$ by powers of rr', together with the nested structure of their half-spaces.

mutatis mutandis in the Davis complex. Alternatively, the Lemma can be reduced directly from Theorem II.1 by viewing the Σ as a subcomplex of the cone on the barycentric subdivision of the Coxeter complex—see the discussion at the end of Section D.2 in [32].

Remark 4.18 (Complexes of groups)*.* Although we will not make use of the following in this thesis, we will remark that in some settings, including in the construction of the Davis complex, the basic construction can be thought of in terms of covering maps of *complex of groups* (see [56]). For the interested reader we sketch this now, for a related discussion see Chapter II.12 in [14].

Let K be a simplicial complex, S a finite set, and G a group. Give K a mirror structure over S and choose a family of groups over S. From this data we can define a complex of groups by associating to each simplex the subgroup $G_{S(x)}$ of G, where x lies in the interior of the simplex unless the simplex is a vertex, in which case x is that vertex. The edge maps are simply inclusion maps.

The group G acts on $\mathcal{U}(G, K)$ with strict fundamental domain K. Thus the quotient $\mathcal{U}(G, K)/G$ can be identified with K. We can label the simplices of K by the stabilisers of the their images in $\mathcal{U}(G, K)$ under $K \hookrightarrow \mathcal{U}(G, K)\colon x \mapsto [1, x]$. This labelling recovers the complex of groups structure defined above, so K is *developable*.

The quotient map $\mathcal{U}(G, K) \to K$ is a complex of groups covering map. We

might ask when $\mathcal{U}(G, K)$ is the *universal* cover of K.

THEOREM 4.19 (Theorem 9.1.3 in [32]): $\mathcal{U}(G, K)$ *is simply connected if and only if*

1. K *is simply connected*

2. *For each $s \in S$, the mirror K_s is non-empty and path connected*

3. *For each $\{s, t\} \subset S$ such that $G_{s,t}$ is finite, $K_s \cap K_t \neq \emptyset$*

One can check that these conditions hold for the definition of $\Sigma(W, S)$ in terms of the basic construction, $\mathcal{U}(W, K(W, S))$, and hence $\mathcal{U}(W, K(W, S))$ is the universal cover of $K(W, S)$, thought of as a complex of groups.

4.2 When is a reflection system a Coxeter system?

Fix a Coxeter system (W, S) and write R for the set of reflections in (W, S).

Definition 4.20. Let $X \subset R$ a finite tuple of reflections. Write $W_X = \langle X \rangle$ for the subgroup of W generated by X. Note that *a priori* it is not clear that W_X is a Coxeter group. We call the pair (W_X, X) a **reflection system** for the subgroup W_X of W, which is called a **reflection subgroup**.

In [34], Vinay Deodhar shows that if (W_X, X) is a reflection system, W_X is a Coxeter group (although in general it is not a special subgroup of W), and it admits a canonical Coxeter system with respect to (W, S). The system is canonical in the sense that it it is minimal with respect to a certain pre-order relation on R. We will not give this definition since it plays no further role in our work. For our purposes we take the criterion stated in Theorem 4.26 as the definition of this canonical Coxeter system.

Such a Coxeter system is, *a fortiori*, a reflection system and in [42], Matthew Dyer gave a geometric criterion for when a reflection system is this canonical Coxeter system. Before we state this criterion we must first define a notion of *angles* between reflections in a Coxeter group.

4.2.1 Angles in Coxeter systems

To state this Recall the symmetric bilinear form B defined in Definition 1.7.

Definition 4.21. Let r be a reflection in (W, S). A **reduced palindromic expression** for r is an expression ${}^w s$ which equals r such that w a reduced word; $s \in S$; and $\ell(ws) = \ell(w) + 1$.

Fix a palindromic reduced expression $r = {}^w s$ for a reflection r and set $e_r := w e_s$ where e_s is the basis vector associated to $s \in S$ (in this thesis we do not discuss root systems at all, however the interested reader will note that the condition on w and s guarantees that e_r is what is called a *positive root*).

LEMMA 4.22 ([41] Lemma 3.11)**:** *Let $r, r' \in R$ be reflections. Then*

$$B(e_r, e_{r'}) \in (-\infty, -1] \cup \{\cos(k\pi/m) \mid k, m \in \mathbb{N}, m \neq 0\} \cup [1, \infty)$$

and the order of rr' is finite if and only if

$$B(e_r, e_{r'}) = -\cos\left(\frac{k\pi}{m}\right)$$

where $k < m \in \mathbb{Z}^+$, and $\gcd(k, m) = 1$; in this case the order is m.

Definition 4.23. Let $r, r' \in R$ be reflections, define the **angle** between r and r', written $\angle rr'$, to be 0 if rr' has infinite order, and otherwise to be

$$\angle rr' = \arccos(-B(e_r, e_{r'})).$$

Following [83], we call the pair (r, r') **sharp-angled** with respect to (W, S) if the angle $\angle rr' = 0$ or π/m for some $m \geqslant 2$. More generally a tuple of reflections X is **sharp-angled** with respect to (W, S) if for any $r \neq r' \in X$, the pair (r, r') is sharp-angled.

Remark 4.24 (Sharpness versus non-obtuseness). Being sharp-angled implies the angles are non-obtuse but is a strictly stronger condition since we do not allow angles such as $2\pi/5$.

The vector e_r can be thought of as the normal to the hyperplane fixed by r which points towards the fundamental chamber (in the sense of Theorem 1.8).
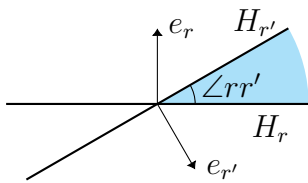
Figure 4.11: B measures the angle between hyperplanes.

Then B measures the angle between the normal vectors e_r and $e_{r'}$, and $\angle rr'$ measures the angle between the hyperplanes, as measured in the sector of their complement containing the fundamental chamber.

Example 4.25. In any Coxeter system, computing the angle between two reflections r and r' such that rr' has finite order reduces to a computation in a special dihedral subgroup. This is because the finite subgroup $\langle r, r' \rangle$ is conjugate into W_T for some $T = \{s, s'\} \subset S$.

One way to see this is to apply the Bruhat-Tits Fixed Point Theorem (see for example Theorem I.2.11 in [32]) to $\langle r, r' \rangle$ acting on $\Sigma(W, S)$ to find a fixed point. There is some $w \in W$ which sends this fixed point to the fundamental chamber K. Therefore, we can conclude that $\langle r, r' \rangle$ is conjugate by w into the stabiliser of a point in K, but by the construction of $\Sigma(W, S)$ this must be a spherical special subgroup.

If we choose $w \in W$ to be of minimal length such that $\langle r, r' \rangle \leqslant {}^w W_T$ (ie the parabolic subgroup obtained by conjugating W_T by w), then there are reflections $t, t' \in W_T$ such that $r = {}^w t$ and $r' = {}^w t'$. Using the minimality of w and the fact that B is W-invariant, we get

$$-\cos(\angle rr') = B(e_r, e_{r'}) = B(we_t, we_{t'}) = B(e_t, e'_t) = -\cos(\angle tt').$$

Now we want to compute $\angle tt'$. Any element of W_T can be expressed as a reduced alternating word of the form $ss's \cdots$ or $s'ss' \cdots$ of length at most $m_{ss'}$. We write $a_d(s, s') = ss's \cdots$ for the alternating word of length d. For $0 < d < m_{ss'}$, all words $a_d(s, s')$ or $a_d(s', s)$ represent distinct elements of W_T; $a_0(s, s') = a_0(s', s)$ is the identity; and $a_{m_{ss'}}(s, s') = a_{m_{ss'}}(s', s)$ is the unique element of longest length. An element of W_T is a reflection if and only if any (or equivalently every) alter-

nating word representing it has odd length, and hence is palindromic.

Now suppose $t \neq t'$ and $\ell(t) \leqslant \ell(t')$. Up to relabelling s and s' we can assume $t = a_{\ell(t)}(s, s')$, and express t' by a (possibly non-reduced) word $a_d(s', s)$ for $d = \ell(t')$ or $2m_{ss'} - \ell(t')$. Then it is straightforward to check that

$$\angle tt' = \frac{\ell(t) + d}{2} \frac{\pi}{m_{ss'}},$$

see Figure 4.12.
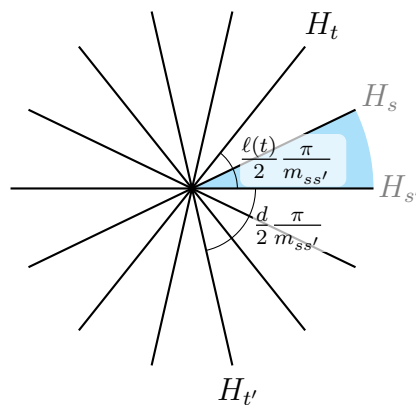
Figure 4.12: Measuring the angle between hyperplanes.

Theorem 4.26 (Equivalent to Theorem 4.4 in [42])**:** *Let $X \subset R$ be a tuple of reflections, then (W_X, X) is the canonical Coxeter system associated to W_X with respect to (W, S) if and only if all elements of X are distinct and X is sharp-angled.*

Remark 4.27 (Canonicity)*.* It follows from Proposition 3.5 and equation 3.10 in [42] that the canonical Coxeter system for a reflection system (W_X, X) is the unique reflection system (W, X') which minimises $\sum_{r \in X'} \ell_S(r)$, where ℓ_S is the length function on the host Coxeter system (W, S).

4.2.2 Geometric criterion on the Davis complex

We can reinterpret this Theorem in the context of the Davis complex of (W, S), Σ. Let $X \subset R$ be a tuple of reflections, and let $\mathcal{H}_X := \{\Sigma_r \mid r \in X\}$ be the hyperplane arrangement in Σ associated to X. If two of these hyperplanes intersect, then there is a point $x \in \Sigma_r \cap \Sigma_{r'}$ with an open neighbourhood contained in a Coxeter

polytope as mentioned in Theorem 4.14. The dihedral angle between Σ_r and $\Sigma_{r'}$ at x, as measured with respect to the Euclidean metric induced by the Euclidean metric on the Coxeter polytope, does not depend on the choice of x. Moreover if the angle as measured in $\Sigma_r^+ \cap \Sigma_{r'}^+$, it is $\angle rr'$.

Definition 4.28. A reflection $r' \in X$ an **outlier** if for any reduced palindromic expression $r' = {}^w s$ there is some $r \in X$ such that $\ell(rw) < \ell(w)$.

Define K^X to be the closed subcomplex of Σ which is the intersection $\bigcap_{r \in X} \overline{\Sigma_r^+}$. Note that $K \subset K^X$, so this complex is non-empty. If a tuple of reflections X does not contain any outliers, this implies that $K^X \cap \Sigma_r$ is non-empty for all $r \in X$. More generally an outlier might meet K^X, but that intersection locally has codimension greater than 1. As an example, in Figure 4.12, if $X = \{s, s', t\}$ then t is an outlier. We call K^X **sharp-angled** if X is.

Definition 4.29. The **tautological mirror structure** on K^X over X is given by declaring $K_r^X = K^X \cap \Sigma_r$. Define the **tautological family of groups** over X starting with $G = W_X$ and setting $G_r = \langle r \rangle \cong \mathbb{Z}_2$.

Given these definitions we can now apply the basic construction, yielding a space $\mathcal{U}(W_X, K^X)$ on which W_X acts. Notice that if $X = S$, then $W_X = W$, $K^X = K$, and the previous paragraph merely recovers the Davis complex Σ.

For $r, r' \in X$, let $m_{rr'}$ be the order of rr'. Write $(\overline{W}_X, S_X)$ for the Coxeter system with Coxeter presentation

$$\langle s_r \text{ such that } r \in X \mid s_r^2, (s_r s_{r'})^{m_{rr'}} \text{ for all} r, r' \in X \rangle,$$

where $S_X = (s_r)_{r \in X}$ is an abstract tuple of generators. Define the surjection $\phi : \overline{W}_X \to W_X$ by sending the generators $s_r \mapsto r$ for each $r \in X$.

Definition 4.30. Give K^X a second mirror structure, this time over the set S_X, by declaring $K_{s_r}^X = K^X \cap \Sigma_r$, and define a family of groups over S_X with $G = \overline{W}_X$, and $G_{s_r} = \langle s_r \rangle$.

With this second mirror structure and family of groups we can also construct the space $\mathcal{U}(\overline{W}_X, K^X)$. In this slightly more general setting it is useful to state a *universal property* for $\mathcal{U}(G, K^X)$.

THEOREM 4.31 ([109] Proposition 1): *Let G be either W_X or $\overline{W}_X$, and suppose it acts on a space Y, and for each $r \in X$ write Y_r for the fixed set of r or s_r respectively. Suppose $i : K^X \to Y$ is a map such that $i(K_r^X) \subset Y_r$ or $i(K_{s_r}^X) \subset Y_r$ respectively for all $r \in X$. Then there is a unique extension of i, $\tilde{\imath} : \mathcal{U}(G, K^X) \to Y : [w, x] \mapsto wi(x)$ which is G-equivariant.*

Remark 4.32 (Equivalence of the definitions of Σ). Take $X = S$, $G = W$, and Y the geometric realisation of WS (see Definition 4.12). Then we can apply the universal property to the inclusion of K into the geometric realisation of WS induced by $S \hookrightarrow WS$. Together with some simple arguments to show that the the extension is a homeomorphism, this proves that the two definitions of $\Sigma(W, S)$ given previously are equivalent, in the sense that there is a W-equivariant homeomorphism between $\mathcal{U}(W, K(W, S))$ and the geometric realisation of WS, see Theorem 7.2.4 in [32].

The simplicial structure on K induces a simplicial structure on K^X and on the spaces obtained by the basic construction as above. The extension $\tilde{\imath}$ is a simplicial map which is a bijection when restricted to the interior of any translate of K^X. We use the following observation in the proof of the Theorem below.

LEMMA 4.33: *Let G be W_X or $\overline{W}_X$, and let $\mathcal{U}(G, K^X)$ be the space constructed above. Then the map $\tilde{\imath}_G : \mathcal{U}(G, K^X) \to \Sigma$, extending the inclusion of K^X into Σ, is surjective.*

Proof. First consider the case $G = W_X$. For any $v \in W_X$ and $x \in K^X$ write $[v, x]_{\mathcal{U}}$ for the corresponding point in $\mathcal{U}(W_X, K^X)$ Similarly, for any $w \in W$ and $y \in K$ write $[w, y]_\Sigma$ for the corresponding point in Σ. Let $A := \{w \in W \mid wK \subset K^X\}$, so that any $x \in K^X$ can be written $[w, y]_\Sigma$ for some $w \in A$ and $y \in K$. Identifying K^X with its image in Σ, $\tilde{\imath}_{W_X}$ maps

$$[v, [w, y]_\Sigma]_{\mathcal{U}} \mapsto [vw, y]_\Sigma$$

so $\tilde{\imath}_{W_X}$ is surjective if and only if $\{vw \mid v \in W_X, w \in A\} = W$. This is the case if and only if and only if A contains a representative of every right coset of W_X in W.

Assume that A does not contain a representative of the coset $W_X w$ and assume $\ell(w) \leqslant \ell(u)$ for all $u \in W_X w$. By definition, in Σ the translate $wK \not\subset K^X$, so there is some hyperplane Σ_r for $r \in X$ which separates K from wK. Then $\ell(rw) < \ell(w)$, but $r \in W_X$, so $rw \in W_X w$ contradicting the choice of w. Thus A contains a right transversal (ie a set of unique right coset representatives for W_X in W), and $\tilde{\imath}_{W_X}$ is surjective.

Now consider the case $G = \overline{W}_X$. The surjective map $\phi : \overline{W}_X \to W_X$ induces a map $\tilde{\jmath} : \mathcal{U}(\overline{W}_X, K^X) \to \mathcal{U}(W_X, K^X) : [u, y]_{\mathcal{U}(\overline{W}_X, K^X)} \mapsto [\phi(u), y]_{\mathcal{U}(W_X, K^X)}$ which is in turn surjective. Hence the composition $\tilde{\imath}_{W_X} \circ \tilde{\jmath} : \mathcal{U}(\overline{W}_X, K^X) \to \Sigma$ is continuous $\overline{W}_X$-equivariant map which restricts to the identity on K^X, and so by the uniqueness part of the universal property is the map $\tilde{\imath}_{\overline{W}_X}$. Therefore $\tilde{\imath}_{\overline{W}_X}$ is surjective, as required. $\square$

We can now state part of then Davis complex version of Theorem 4.26. For completeness, we give a proof which does not depend on the results in [34] or [42].

THEOREM 4.34: *Let $X \subset R$ be a finite tuple of reflections which does not contain any outliers, and consider the corresponding reflection system (W_X, X). Let $i : K^X \hookrightarrow \Sigma$ be the inclusion map. The group $\overline{W}_X$ acts on Σ via the surjection ϕ onto $W_X \leqslant W$, so let $\tilde{\imath}_{\overline{W}_X} : \mathcal{U}(\overline{W}_X, K^X) \to \Sigma$ be the $\overline{W}_X$-equivariant extension coming from the universal property. Then $\tilde{\imath}_{\overline{W}_X}$ is a homeomorphism if and only if K^X is sharp-angled. In this case*

1. *the image $i(K^X)$ is a strict fundamental domain for $\overline{W}_X$ acting on Σ (see Definition 4.1),*

2. *the map ϕ is an isomorphism, and*

3. *the pair (W_X, X) is a Coxeter system.*

4. *X is the unique reflection generating tuple which minimises $\sum_{r \in X} \ell(S)(r)$.*

Remark 4.35 (Geometric reflection groups case)*.* This can be thought of as a generalisation of Theorem 6.4.3 in [32]. That Theorem applies to the special case of geometric reflection groups (see Section 4.1.1). It replaces Σ with a space of constant curvature $\mathbb{X}^n$, and K^X with a simple convex polytope P. Then assuming that all the dihedral angles in P are sharp, ie of the form π/m it states that the group W generated by the reflections in the co-dimension 1 faces of P is a Coxeter group and the tiling of $\mathbb{X}^n$ by W-translates of P is homeomorphic to $\mathcal{U}(W, P)$ where P has the tautological mirror structure and family of groups over the set of co-dimension 1 faces of P.

The proof of Theorem 6.4.3 in [32] is quite different from the proof of our Theorem below, and does not generalise to our setting. Nevertheless, recall the dual CW structure on Σ in which the cells are Coxeter polytopes (Theorem 4.14). We can apply Theorem 6.4.3 in the interior of these Coxeter polytopes which is a key step in our proof to show that $\tilde{i}_{\overline{W}_X}$ is a local homeomorphism.

Proof. Assume that $\tilde{i}_{\overline{W}_X}$ is a homeomorphism. First we prove the last four claims.

1. The first follows immediately from the definition of $\mathcal{U}(\overline{W}_X, K^X)$.

2. By the first claim, there is a bijection between the elements of $\overline{W}_X$ and the translates of K^X in $\mathcal{U}(\overline{W}_X, K^X)$, and so by hypothesis there is a bijection with the translates of $i(K^X)$ in Σ, which we call β. Let $j : K^X \hookrightarrow \mathcal{U}(W_X, K^X)$ be the inclusion, and $\tilde{j} : \mathcal{U}(\overline{W}_X, K^X) \to \mathcal{U}(W_X, K^X)$ the extension to $\mathcal{U}(\overline{W}_X, K^X)$. Finally let $\tilde{i}_{W_X} : \mathcal{U}(W_X, K^X) \to \Sigma$ be the extension to $\mathcal{U}(W_X, K^X)$ of the inclusion of K^X into Σ.

$$
\begin{array}{ccc}
\mathcal{U}(\overline{W}_X, K^X) & \xrightarrow{\ \tilde{i}_{\overline{W}_X}\ } & \\
\tilde{j} \downarrow & \searrow & \Sigma \\
\mathcal{U}(W_X, K^X) & \xrightarrow{\ \tilde{i}_{W_X}\ } &
\end{array}
$$

By Lemma 4.33 and the paragraph before it, the map $\tilde{i}_{W_X}$ is simplicial and surjective. This map induces a surjective map κ from W_X to the translates of the image of K^X in Σ. Notice also that $\tilde{j}$ induces a map between the $\overline{W}_X$

translates of K^X in $\mathcal{U}(\overline{W}_X, K^X)$ and the W_X translates of K^X in $\mathcal{U}(W_X, K^X)$, ie a map between $\overline{W}_X$ and W_X. This is exactly the surjection ϕ. Thus $\kappa \circ \phi$ is a surjection from $\overline{W}_X$ to the translates of K^X in Σ, but by the uniqueness part of Theorem 4.31, $\tilde{\imath}_{\overline{W}_X} = \tilde{\imath}_{W_X} \circ \tilde{\jmath}$, and this must be the bijection β. Since $\beta = \kappa \circ \phi$ is injective, ϕ must be injective, and so bijective.

3. The third follows immediately from the second.

4. Since (W_X, X) is a Coxeter system, we can consider the set of reflections $R(W_X, X) \subset R(W, S)$. The hyperplane arrangement $\mathcal{H}_{R(W_X, X)}$ divides up Σ so that $\mathcal{H}_X$ bounds K^X and no other hyperplanes in $\mathcal{H}_{R(W_X, X)}$ meet the interior of K^X. The claim now follows by applying Proposition 4.17.

Continue to assume that $\tilde{\imath}_{\overline{W}_X}$ is a homeomorphism, and we prove that K^X is sharp-angled. Let $r, r' \in X$, recall from Lemma 4.16 that $\Sigma_r \cap \Sigma_{r'} = \emptyset$ if and only if rr' has infinite order—in this case $\angle rr' = 0$. Assume then that rr' has finite order m, and let $K^{\{r,r'\}}$ be the component of $\Sigma - (\Sigma_r \cup \Sigma_{r'})$ which contains K; it also contains the whole of K^X. Since K^X is a fundamental domain for W_X acting on Σ, as we proved in points 1 and 2 above, the set $K^{\{r,r'\}}$ is a fundamental domain for the action of $\langle r, r' \rangle \leqslant W_X$ on Σ. But $\langle r, r' \rangle$ is isomorphic to the dihedral group of order $2m$, and by considering its action on a Coxeter polytope neighbourhood of a point in $\Sigma_r \cap \Sigma_{r'}$, as discussed under Theorem 4.26, we can conclude that $\angle rr' = \pi/m$.

Conversely, assume that K^X is sharp-angled, and we prove that $\tilde{\imath}_{\overline{W}_X}$ is a homeomorphism. Let $[u, y]$ be a point in $\mathcal{U}(\overline{W}_X, K^X)$, then its stabiliser in $\overline{W}_X$ is conjugate to the spherical subgroup of $\overline{W}_X$ generated by the $s_r \in S_X$ such that $y \in K^X_{s_r}$, ie the elements of $S_X(y)$ (recall the notation from Definition 4.3). If $S_X(y) = \emptyset$ then $\tilde{\imath}_{\overline{W}_X}$ is a homeomorphism on a small neighbourhood of y because y is not contained in a mirror of K^X.

Otherwise, the subgroup of W generated by $\{r \in R \mid s_r \in S_X(y)\}$ is conjugate to a finite reflection subgroup contained in a maximal spherical subgroup W_T of W, for some $T \subset S$. Moreover in the cell structure on Σ mentioned in Theorem 4.14, y is contained in a Coxeter polytope of type W_T. By Theorem 6.4.3 in

[32], the sharp-angled assumption implies that $\tilde{\imath}_{\overline{W}_X}$ is a homeomorphism when restricted to this polytope.

Thus we have shown that $\tilde{\imath}_{\overline{W}_X}$ is locally a homeomorphism, in particular it is locally injective. By Lemma 4.33, $\tilde{\imath}_{\overline{W}_X}$ is also surjective. Since Σ is simply connected (in fact it is CAT(0) by Theorem 4.14), $\tilde{\imath}_{\overline{W}_X}$ is a homeomorphism. $\qquad\square$

In general, even if X is sharp-angled, (W_X, X) will not be equal to (W, S) since X may not generate W. If X does generate W, is sharp-angled, has no outliers, and contains no repeated entries then it is necessarily that case that X is a permutation of S.

Note that the converse of the Theorem does not hold, there are examples of *Coxeter* systems (W_X, X) such that X is *not* sharp-angled, the simplest examples coming from dihedral groups. Let $W = \langle s, s' \mid s^2, s'^2, (ss')^m \rangle$ be the dihedral group of order $2m$ with Coxeter system $(W, S) = (s, s')$. Suppose $m > 3$ is odd for the sake of example, and let $X = (r = s, r' = s'ss')$. Then $W_X \cong W$, (W_X, X) is a Coxeter system, but $\angle rr' = 2\pi/m$. Another example comes from $W(H_3)$ and the generating tuple coming from the exceptional automorphism in Proposition 3.34 which is illustrated in Figure 4.17d.

Remark 4.36 (Local injectivity and complexes of groups). Along the same lines as in Remark 4.18, given a tuple of reflections X which contains no outliers we can construct a complex of groups with underlying complex K^X using the tautological mirror structure and family of groups over X.

Then the space $\mathcal{U}(W_X, K^X)$ is simply connected using Theorem 4.19. Point 1 follows since K^X is a convex subset of Σ which is simply connected. Point 2 follows because X contains no outliers, and point 3 follows from Lemma 4.16.

The map $\tilde{\imath}_{W_X}: \mathcal{U}(W_X, K^X) \to \Sigma$ can be thought of as a *developing map* for $\mathcal{U}(W_X, K^X)$. The key to Theorem 4.34 is that this developing map is locally injective if and only if X is sharp-angled.

Suppose X fails to be sharp-angled and $r, r' \in X$ such that $\angle rr' = k\pi/m$ for some natural numbers k and m with $\gcd(k, m) = 1$. Then $\langle r, r' \rangle$ is isomorphic to the dihedral group of order $2m$. When building $\mathcal{U}(W_X, K^X)$ there are $2m$ copies of K^X corresponding to the elements of $\langle r, r' \rangle$. These are glued together such that

they all meet in $K_r^X \cap K_{r'}^X$. The result is that the total dihedral angle around this intersection is $2\pi k$.

Under the map $\tilde{\imath}_{W_X}$, this intersection is mapped to $\Sigma_r \cap \Sigma_{r'}$, around which the total dihedral angle is 2π. Thus the developing map looks locally like a k-fold branched cover near $K_r^X \cap K_{r'}^X$. This phenomenon is illustrated in Figure 4.13.
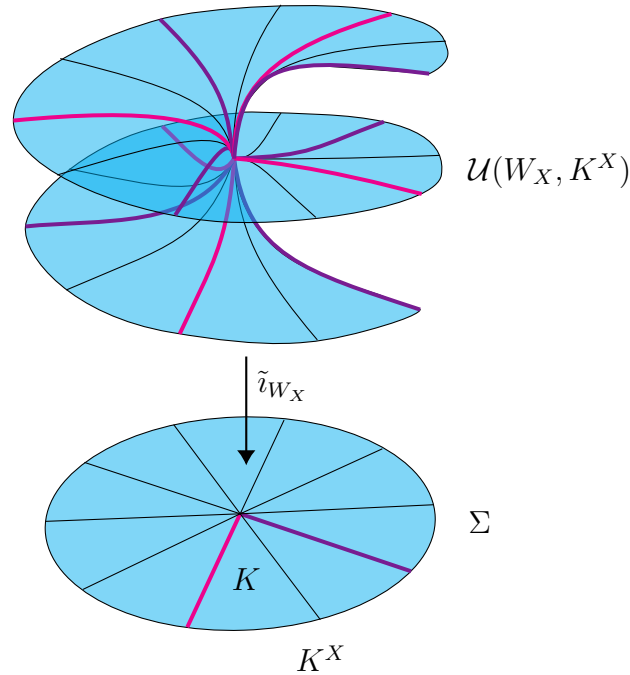
Figure 4.13: The developing map fails to be locally injective if X is not sharp-angled. In $\mathcal{U}(W_X, K^X)$, the two free purple edges are identified; they are drawn separate here for clarity of the illustration.

4.3   Computing a Coxeter system from a reflection system

The purpose of this Section is to use the criterion discussed in the last Section to give an algorithm which turns an arbitrary reflection system (W_X, X) with respect to some Coxeter system (W, S) into a Coxeter system for W_X. This gives a proof of the well-known fact that any reflection subgroup of a Coxeter group is a Coxeter group, and in fact outputs the *canonical* Coxeter system for W_X.

Using the formulation of Theorem 4.34 gives a geometrically more intuitive definition of this canonicity than is given by Deodhar's original definition in terms

of a pre-order on R [34]. This algorithm is alluded to in [42], although it is not studied at all, and again the formulation there differs from the approach here. In particular, we modify it in the next Section to work with reflection equivalence.

4.3.1   Transforming hyperplane arrangements

We describe three transformations, or moves, which can be applied to a tuple of reflections X such that the result still generates the same subgroup of W.

Assumption 4.37. From now on, we allow X to contain elements in $R \cup \{1\}$.

Before describing these moves, we usually do the following house-keeping operations without comment.

Type 0   First we can permute the entries of X into any order, and second, if two entries contain the same reflection, we can replace one of these entries with the identity. In particular we can assume X has the form $(r_1, \ldots, r_k, 1, \ldots, 1)$ where the r_i's are pairwise distinct reflections.

In the following, assume r and r' are distinct reflection entries of X. We give an algebraic description of each move, then its geometric interpretation in Σ.

Type I   Suppose $r' \in X$ is an outlier (see Definition 4.28), and choose a reduced palindromic expression $r' = {}^w s$. Fix an $r \in X$ such that $\ell(rw) < \ell(w)$, then replace r' in X by ${}^r r'$.

For the sake of illustration, consider the case that the hyperplanes Σ_r and $\Sigma_{r'}$ do not intersect. Each divides Σ into two half-spaces, and the condition $\ell(rw) < \ell(w)$ is equivalent to saying that Σ_r separates $\Sigma_{r'}$ from K. The hyperplane $\Sigma_{r r'}$ is the reflected image of $\Sigma_{r'}$ in Σ_r.
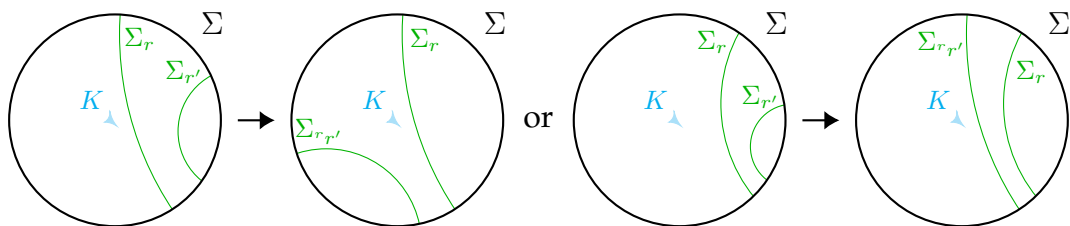
Type II   Suppose that $\angle rr' > \pi/2$; there is some pair $T = \{s, s'\} \subset S$ such that $m_{ss'} < \infty$ and so that $\langle r, r' \rangle$ is conjugate into the special dihedral subgroup W_T. We can write $\angle rr' = k\pi/m_{ss'}$ for some $m_{ss'}/2 < k < m_{ss'}$. Let $w \in W$ be the unique minimal length element such that $\langle r, r' \rangle \leqslant {}^{w}W_T$. Let $t, t' \in R(W_T, T)$ be such that $r = {}^{w}t$ and $r' = {}^{w}t'$; because w is minimal length, $\ell(r) = 2\ell(w) + \ell(t)$ and $\ell(r') = 2\ell(w) + \ell(t')$. Suppose further that $\ell(t) \leqslant \ell(t')$, then define $t'' = {}^{t}t'$, and replace r' in X with ${}^{w}t'' = {}^{r}r'$.

The hyperplanes Σ_r and $\Sigma_{r'}$ intersect, and the sector they cut out of Σ which contains K has obtuse dihedral angle $k\pi/m_{ss'}$. The move reflects $\Sigma_{r'}$ in Σ_r to bring a non-obtuse angled sector closer to containing K.
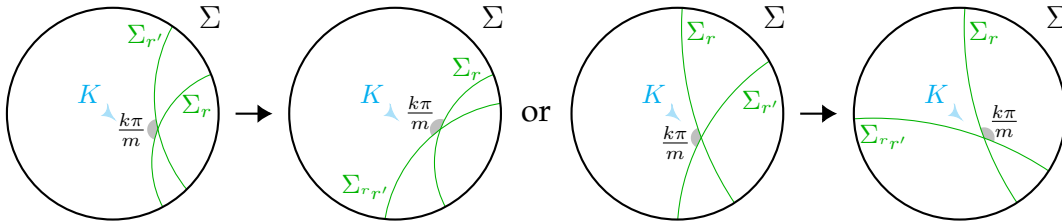
Type III   Suppose $\angle rr' = k\pi/m$ for some $1 < k < m - 1$ such that $\gcd(k, m) = 1$. Then $H = \langle r, r' \rangle$ is a dihedral group of order $2m$. Define T, w, t, and t' as above. Then there is some integer p such that $pm = m_{ss'}$ and $\angle rr' = pk\pi/m_{ss'}$. Define a new reflection

$$t'' = \begin{cases} a_{2p-\ell(t)}(s', s) & \text{if } \ell(t) < 2p \\ a_{\ell(t)-2p}(s, s') & \text{if } \ell(t) > 2p \end{cases}$$

and replace r' in X with $r'' = {}^{w}t''$. In [83] this type of move is called an *angle-deformation*, and is used to study generating sets of Coxeter groups up to automorphisms of the Coxeter group itself.

The hyperplanes Σ_r and $\Sigma_{r'}$ intersect at an angle which is a proper multiple of π/m, such that the supplementary angle is also a proper multiple of π/m. We replace $\Sigma_{r'}$ with another hyperplane $\Sigma_{r''}$ such that either $\angle rr''$ or $\pi - \angle rr'$ is π/m.
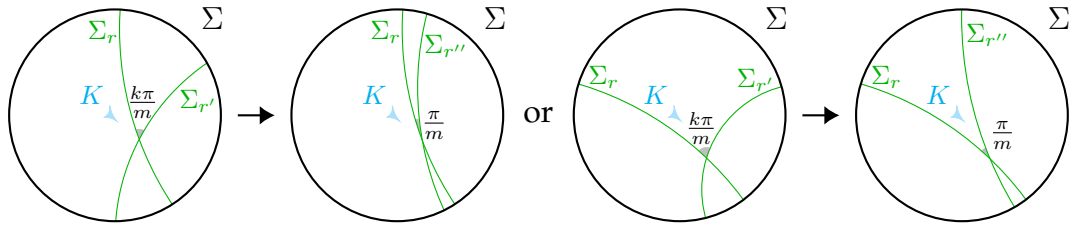
4.3.2   Decreasing complexity

Definition 4.38. Given a tuple of reflections X, we define its **complexity** to be

$$c(X) = \sum_{r \in X} \ell(r).$$

LEMMA 4.39: *Let X be a tuple or reflections, and let X' be obtained from X by one of the moves described above. Then $\langle X \rangle = \langle X' \rangle$, and except when X' is merely a permutation of X, $c(X') < c(X)$.*

Proof. This is true for the type 0 moves. For a move of type I and II, the reflection $r' \in X$ is replaced by ${}^r r'$, where r is another reflection in X. So X and X' generate the same subgroup. For type III moves, it is straightforward to check that t'' lies in the group generated by t and t', and so the dihedral group generated by (r, r'') is equal (as a subgroup of W) to $\langle r, r' \rangle$. Again, X and X' generate the same subgroup. Now we compute complexities

For type I   Since $\ell(rw) < \ell(w)$, we compute

$$\ell({}^r r') = \ell({}^{rw} s) \leqslant 2\ell(rw) + 1 < 2\ell(w) + 1 = \ell(r').$$

For type II   Recall the notation from Example 4.25. We assumed that $\ell(t) \leqslant \ell(t')$, and since they are elements of the finite dihedral special subgroup $W_{\{s,s'\}}$ we can write t as a reduced expression over $T = \{s, s'\}$ of length $\ell(t)$

$$t = a_{\ell(t)}(s, s') \text{ (possibly after swapping the roles of } s \text{ and } s').$$

Also write t' as a reduced word of length $\ell(t')$; there are two cases: (1) $t' = a_{\ell(t')}(s, s')$, or (2) $t' = a_{\ell(t')}(s', s)$.

   We want to compute $\ell({}^t t')$—assume we are in case (1) and $2\ell(t) < \ell(t')$, then we can cancel all occurrences of ss and $s's'$ to get

$$^t t' = a_{\ell(t)}(s, s') \cdot a_{\ell(t')}(s, s') \cdot a_{\ell(t)}(s, s') = a_{\ell(t') - 2\ell(t)}(s', s).$$

Since $0 < \ell(t') - 2\ell(t) < \ell(t') \leqslant m_{ss'}$ this is a reduced expression and $\ell(^tt') < \ell(t')$.
If we are in case (1) but $2\ell(t) > \ell(t')$ then

$$^tt' = a_{\ell(t)}(s, s') \cdot a_{\ell(t')}(s, s') \cdot a_{\ell(t)}(s, s') = a_{2\ell(t)-\ell(t')}(s, s').$$

In this case $0 < 2\ell(t) - \ell(t') < 2\ell(t') - \ell(t') = \ell(t') < m_{ss'}$ so again the expression
above is reduced and $\ell(^tt') < \ell(t')$.

Note that in case (1) we did not have to explicitly use the assumption that $\angle rr'$
is obtuse because this is guaranteed by the fact that the reduced words for t and t'
start with the same letter. On the other hand, if we are in case (2) we need to use
this assumption explicitly. From Example 4.25, using $\angle rr' = k\pi/m_{ss'}$, we can see
that $2k = \ell(t) + \ell(t')$. But by assumption $k > m_{ss'}/2$ because the angle is obtuse,
so

$$\ell(t) + \ell(t') > m_{ss'}. \tag{4.1}$$

We can write

$$^tt' = a_{\ell(t)}(s, s') \cdot a_{\ell(t')}(s', s) \cdot a_{\ell(t)}(s, s') = a_{2\ell(t)+\ell(t')}(s, s').$$

If $2\ell(t) + \ell(t') < 2m_{ss'}$ we can apply the relation $(ss')^{m_{ss'}} = 1$ in W_T to conclude
that $a_{2m_{ss'}-(2\ell(t)+\ell(t'))}(s', s)$ is a reduced expression for $^tt'$, so

$$\ell(^tt') = 2m_{ss'} - (2\ell(t) + \ell(t')) \overset{(4.1)}{<} 2(\ell(t) + \ell(t')) - (2\ell(t) + \ell(t')) = \ell(t').$$

Otherwise, $2m_{ss'} < 2\ell(t) + \ell(t') < 3\ell(t') \leqslant 3m_{ss'}$, in which case we can again
apply the relation to see that $^tt'$ has a reduced expression $a_{2\ell(t)+\ell(t')-2m_{ss'}}(s, s')$. In
Example 4.25 we observed that there is a unique maximum length element in W_T,
with length $m_{ss'}$. By assumption $t \neq t'$ and $\ell(t) < \ell(t')$, so $\ell(t) < m_{ss'}$, and hence

$$\ell(^tt') = 2\ell(t) + \ell(t') - 2m_{ss'} < 2\ell(t) + \ell(t') - 2\ell(t) = \ell(t').$$

In all cases we have shown that $\ell(^tt') < \ell(t')$, and hence

$$\ell(^rr') = \ell(^wt'') = 2\ell(w) + \ell(t'') = 2\ell(w) + \ell(^tt') < 2\ell(w) + \ell(t') = \ell(r').$$

For type III   If $t'' = a_{\ell(t)-2p}(s, s')$, then $\ell(t'') < \ell(t) \leqslant \ell(t')$. For the other case
$t'' = a_{2p-\ell(t)}(s', s)$ and $\ell(t) < 2p$. If $t' = a_{\ell(t')}(s', s)$ then from Example 4.25 we
know

$$\angle tt' = \frac{\ell(t) + \ell(t')}{2} \frac{\pi}{m_{ss'}} = \frac{pk\pi}{m_{ss'}},$$

so $\ell(t') = 2pk + \ell(t) \overset{k\geqslant 2}{\geqslant} 4p + \ell(t) > 2p - \ell(t) = \ell(t'')$. Else if $t' = a_{\ell(t')}(s, s')$ then

$$\angle tt' = \frac{\ell(t) + 2m_{ss'} - \ell(t')}{2} \frac{\pi}{m_{ss'}} = \frac{pk\pi}{m_{ss'}},$$

so

$$\ell(t') = \ell(t) + 2m_{ss'} - 2pk \overset{k\leqslant m-2}{\geqslant} \ell(t) + 2m_{ss'} - 2p(m-2)$$
$$= \ell(t) + 2m_{ss'} - 2m_{ss'} + 4p = \ell(t) + 4p > 2p - \ell(t) = \ell(t'').$$

In both cases we have shown that $\ell(t'') < \ell(t')$, and hence

$$\ell(r'') = \ell(^w t'') = 2\ell(w) + \ell(t'') < 2\ell(w) + \ell(t') = \ell(r'). \qquad \square$$

4.3.3   Algorithms for studying reflection systems

The first main result of this Chapter is the following Theorem, which allows us to study properties for reflection subgroups of arbitrary Coxeter systems.

THEOREM 4.40: *Let (W, S) be a Coxeter system, and let X be a finite tuple of reflections in W. Then there is an algorithm which produces a Coxeter system $(W_X, \widetilde{X})$ for the reflection subgroup generated by X.*

Proof. Apply moves of type I–III to X as many times as is possible. Only a finite number of these moves can be performed since $c(x)$ is finite, and each move decreases the complexity. Delete all entries which are the identity and call the resulting tuple $\widetilde{X}$.

Since we cannot apply any type I moves to $\widetilde{X}$ it contains no outliers, and hence we can define $K^{\widetilde{X}}$ with its tautological mirror structure and family of groups. Since we cannot apply any type II moves, all angles in $K^{\widetilde{X}}$ are non-obtuse, and since no type III moves are possible they must be sharp-angled. Applying Theorem 4.34 we conclude that $(W_X, \widetilde{X})$ is a Coxeter system. $\qquad \square$

Compare this with the proof of Proposition 3.7 in [42] and the exposition of that result in Section 3 of [111]. There a finite procedure to find $(W_X, \widetilde{X})$ from

(W_X, X) is given along the following lines. For a tuple of reflections X' define

$$\chi(W_{X'}) := \{r \in W_{X'} \cap R(W, S) \mid \ell_S(rr') > \ell_S(r)$$

$$\text{for all } r' \in W_{X'} \cap R(W, S)$$

$$\text{such that } r \neq r'\}.$$

This gives the canonical set of Coxeter generators for $W_{X'}$ mentioned in Theorem 4.26. Then the procedure produces a sequence of reflection generating tuples $X = X_0, X_1, X_2, \ldots$ where X_i is obtained from X_{i-1} by replacing a pair of generators (r_1, r_2) with $\chi(\langle r_1, r_2 \rangle)$ (as long as $(r_1, r_2) \neq \chi(\langle r_1, r_2 \rangle)$). Proposition 3.7 in [42] states that this sequence terminates in $\widetilde{X} = \chi(W_X)$.

The key differences between this approach and our proof of Theorem 4.40 are two-fold. Firstly, the moves I–III give a way to compute $\chi(\langle r_1, r_2 \rangle)$ which can be implemented as a practical algorithm—in particular applying all possible type I moves if $\langle r_1, r_2 \rangle$ is infinite, and all possible type II and III moves if $\langle r_1, r_2 \rangle$ is finite. Second we can leverage these moves to study reflection equivalence, as we do in the next Section.

Corollary 4.41: *Let (W, S) be a Coxeter system, and let X be a finite tuple of reflections in W. If W_X is finite index in W, then there is an algorithm to compute this index $[W : W_X]$.*

Proof. Applying the algorithm above yields, by Theorem 4.34, a strict fundamental domain $K^{\widetilde{X}}$ for the action of W_X on Σ. This contains the fundamental chamber K, and is tiled by translates of K. The index $[W, W_X]$ equals the number of these translates.

Using the linearity of W we can solve the word problem, and hence enumerate the elements of W by their length. The translate wK belongs to $K^{\widetilde{X}}$ if and only if $\ell(rw) > \ell(w)$ for all $r \in \widetilde{X}$. There is some uniform bound on the length of the $w \in W$ such that $wK \subset K^{\widetilde{X}}$, since this set is finite; and once we reach a point where all translates of K by elements of W of some fixed length do not lie in $K^{\widetilde{X}}$, then we have found all W-translates of K in $K^{\widetilde{X}}$, and hence computed the index. $\quad\square$

Corollary 4.42: *Let (W, S) be a Coxeter system, and let X be a finite tuple of reflections in W. There is an algorithm to determine whether X generates W.*

Proof. Since W has index 1 in itself, we can apply the proof of the previous result to conclude that $K^{\widetilde{X}} = K$, and hence X generates W if and only if $\widetilde{X}$ is the standard generating tuple, S, for W. □

This algorithm gives a relatively straightforward way to test whether a tuple X of reflections generates W, but we can also extract a necessary condition which can be checked against X directly and is useful later.

Corollary 4.43 (Compare with Lemma 6.4 in [111])**:** *Let (W, S) be a Coxeter system of rank n. If $X = (r_1, \ldots, r_\ell)$ is a finite tuple of reflections which generates W, then there is some permutation $\sigma \in S_\ell$ such that for $1 \leqslant i \leqslant n$, $r_{\sigma(i)}$ is conjugate to s_i.*

4.4   Equivalence of reflection generating tuples

The algorithm we developed in the previous Section is useful for studying reflection subgroups of Coxeter systems, but not all of the moves described are reflection equivalences. Moves of type 0 are elementary Nielsen transformations of type (T1) and (T3*). Notice also that moves of type I and II are transformations of type (T4). On the other hand, Theorem 2.1 illustrates that in general moves of type III cannot be achieved even by combinations of elementary Nielsen transformations, never mind by reflection equivalences.

Theorem 4.44: *Let (W, S) be a Coxeter system, and X a finite tuple of reflections which generate W. Then X is reflection equivalent to a tuple $\widetilde{X}$ containing no outliers and such that for any distinct reflections $r, r' \in \widetilde{X}$, $\angle rr' \leqslant \pi/2$.*

Proof. Similar to the proof of Theorem 4.40 we can apply moves of type 0–II to X a finite number of times until no more moves are possible. As remarked above, each move is a reflection transformation. Call the result $\widetilde{X}$. Since no type I moves are possible there are no outliers, and since no type II moves are possible, all angles are non-obtuse. □

Note that $\widetilde{X}$ is not unique in general, and nor are the angles $\angle rr'$. We demonstrate this in the case of the dodecahedral group $W(H_3)$ in Example 4.51.

We can view part of Theorem 2.1 as a special case of this Theorem. A pair of reflections generate $W(I_2(k))$ if and only if the angle between their corresponding hyperplanes is $\pi\ell/k$ for some $1 \leqslant \ell < k$ such that $\gcd(\ell, k) = 1$, otherwise we could apply Theorem 4.40 to produce a sharp-angled pair of generators which meet at angle $\pi/(k/\gcd(\ell, k))$ and hence generate a proper subgroup isomorphic to $W(I_2(k/\gcd(\ell, k)))$. Now applying Theorem 4.44 to this generating pair of reflections we can replace them by a pair which meet at some non-obtuse angle $\pi\ell/k$ for some $1 \leqslant \ell \leqslant k/2$, so that their product is a rotation by $2\pi\ell/k$. Compare also this dihedral case to Theorem 1.22.

In some special cases, not being able to use moves of type III has no effect on the algorithm, giving the strongest possible reflection equivalence classification of reflection generating tuples. This applies to RACGs, Weyl groups, and affine Coxeter systems, as well as direct and free products of groups of these types.

COROLLARY 4.45: *Let (W, S) be a Coxeter system such that $m_{ss'} \in \{2, 3, 4, 6, \infty\}$ for all $s \neq s' \in S$. Then if X is a reflection generating tuple, then it is reflection equivalent to (a stabilisation of) the standard generating tuple S.*

Remark 4.46 (Reflection equivalence and rigidity). Note that this result *cannot* be deduced simply by considering rigidity. While strong reflection rigidity is enough to guarantee that all Coxeter generating tuples S' for a Coxeter system (W, S) such that $S' \subset R(W, S)$ are reflection equivalent (see Definition 2.9), not all Coxeter systems with $m_{ss'} \in \{2, 3, 4, 6, \infty\}$ are strongly reflection rigid. Indeed not even all finite (Theorem 2.11) or RACGs (Theorem 2.10) are strongly reflection rigid.

Proof. If $\angle rr' = k\pi/m_{ss'}$ is non-obtuse for $m_{ss'} \in \{2, 3, 4, 6\}$ and $\gcd(k, m_{ss'}) = 1$, then it is sharp—or in other words it is never possible to apply a move of type III to X. Therefore the conclusion follows from the proof of Corollary 4.42. □

Recall Lemma 2.8 and Remark 1.18: for reflection equivalence, and more generally Nielsen equivalence, all generating tuples become equivalent after performing n stabilisations where n is the Coxeter rank or algebraic rank respectively. We

can extend this Corollary to all Coxeter systems at the cost of performing only a single stabilisation—in other words, if we perform a single stabilisation all reflection generating tuples become equivalent to a stabilisation of the standard generating tuple no matter how large the Coxeter rank is.

THEOREM 4.47: *Let (W, S) be a Coxeter system and suppose X is a reflection generating tuple. If X is reducible (see Definition 2.6), then it is reflection equivalent to some stabilisation of the standard generating tuple S. In particular, after performing a single stabilisation, every reflection generating tuple is equivalent to some stabilisation of S.*

Proof. We show that if $X = (\ldots, r, \ldots, r', \ldots, 1)$ and a type III move is possible with r and r', then this move can be achieved by a sequence of transformations of type (T1), (T3*), and (T4). The conclusion then follows by the proof of Corollary 4.42.

A type III move has the effect of changing

$$(\ldots, r, \ldots, r', \ldots, 1) \mapsto (\ldots, r, \ldots, r'', \ldots, 1)$$

where r'' is some reflection in $\langle r, r' \rangle$. But this means that r'' is expressible as a palindromic word over $\{r, r'\}$ which can be built by a (T3*) transformation followed by a sequence of (T4) transformations. For the sake of illustration suppose $r'' = {}^{a_d(r', r)^{-1}}r$, then:

$$
\begin{aligned}
(\ldots, r, \ldots, r', \ldots, 1) &\overset{(T3^*)}{\mapsto} (\ldots, r, \ldots, r', \ldots, r) \\
&\overset{(T4)}{\mapsto} (\ldots, r, \ldots, r', \ldots, {}^{r'}r) \\
&\overset{(T4)}{\mapsto} (\ldots, r, \ldots, r', \ldots, {}^{rr'}r) \\
&\quad\vdots \\
&\overset{(T4)}{\mapsto} (\ldots, r, \ldots, r', \ldots, {}^{a_d(r', r)^{-1}}r) \\
&\overset{(T1)}{\mapsto} (\ldots, r, \ldots, r'', \ldots, r')
\end{aligned}
$$

Now $\langle r, r'' \rangle = \langle r, r' \rangle$, so r' can be expressed as a word over $\{r, r''\}$, and so by a similar reverse process we show

$$(\ldots, r, \ldots, r'', \ldots, r') \mapsto (\ldots, r, \ldots, r'', \ldots, 1)$$

is a reflection equivalence, completing the proof. $\qquad\square$

Example 4.48. Consider the 555 triangle group which acts cocompactly on the hyperbolic plane and which has Coxeter-Dynkin diagram
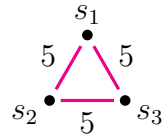
We will study this group rigorously in Example 4.53 below, however for now consider the reflection generating tuple $X = (s_1, {}^{s_2}s_1, {}^{s_3}s_1)$. This tuple satisfies the conclusion of Theorem 4.44 but is not sharp-angled. The corresponding hyperplane arrangement is shown in Figure 4.14.
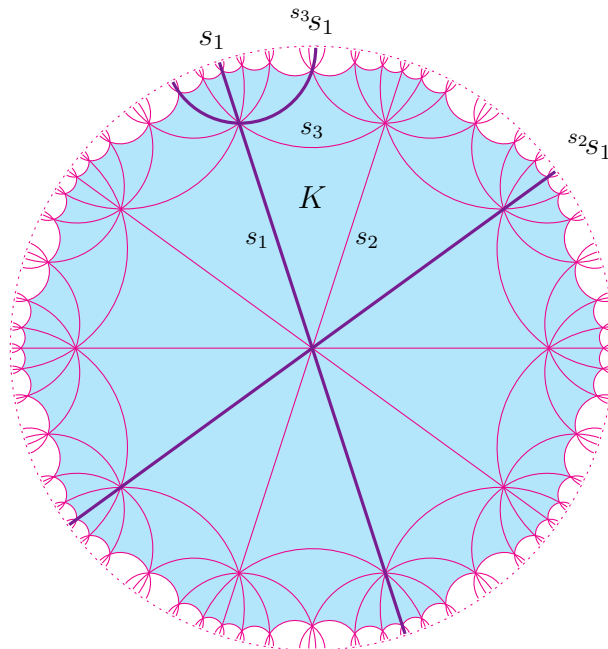
Figure 4.14: A finite portion of the hyperplane arrangement of the 555 triangle group viewed as a subset of the hyperbolic plane, together with the hyperplane arrangement associated to X in purple. The Davis complex is homeomorphic to the barycentric subdivision of this picture (compare with Remark 4.35).

We illustrate that after performing a single stabilisation, X becomes reflection equivalent to $(s_1, s_2, s_3, 1)$, a stabilisation of S. First, the reflection s_3 lies in the dihedral subgroup generated by the sub-tuple $(s_1, {}^{s_3}s_1)$ of $(s_1, {}^{s_2}s_1, {}^{s_3}s_1, 1)$, so we can apply a sequence of reflection equivalences to transform

$$(s_1, {}^{s_2}s_1, {}^{s_3}s_1, 1) \mapsto (s_1, {}^{s_2}s_1, {}^{s_3}s_1, s_3); \ s_3 = {}^{({}^{s_3}s_1)s_1}({}^{s_3}s_1).$$

Next, $^{s_3}s_1$ lies in the dihedral group generated by (s_1, s_3), so we can apply a sequence of reflection equivalences to transform

$$\left(s_1, {}^{s_2}s_1, {}^{s_3}s_1, s_3\right) \mapsto \left(s_1, {}^{s_2}s_1, 1, s_3\right).$$

This frees up an new slot in the generating tuple containing the identity and which we can use to apply the argument again—this is why only ever performing a single stabilisation is required. The reflection s_2 lies in the dihedral group generated by $(s_1, {}^{s_2}s_1)$:

$$\left(s_1, {}^{s_2}s_1, 1, s_3\right) \mapsto \left(s_1, {}^{s_2}s_1, s_2, s_3\right),$$

and then $^{s_2}s_1$ lies in the dihedral group generated by (s_1, s_2):

$$\left(s_1, {}^{s_2}s_1, s_2, s_3\right) \mapsto \left(s_1, 1, s_2, s_3\right).$$

After a final permutation we are left with $(s_1, s_2, s_3, 1)$, a stabilisation of (s_1, s_2, s_3). This process is illustrated in Figure 4.15.

Theorem 4.47 does not, of course, imply that all reflection generating tuples for a Coxeter system (W, S) of size greater than $\#S$ are equivalent to a stabilisation of the standard generating tuple. We know of no examples of non-minimal reflection generating tuples which are not reflection equivalent to a stabilisation of the standard one.

Question 4.49. Let (W, S) be a Coxeter system, and X a reflection generating tuple of W which is strictly bigger than S. Is it necessarily the case that X is reflection equivalent to a stabilisation of S?

4.5   Triangle groups

Having completely understood the rank two case in Theorem 2.1, the next class of groups to look at are the rank three Coxeter groups. It turns out almost all of these act co-finitely on a two dimensional space of constant curvature, and if all

$$\left(s_1, {}^{s_2}s_1, {}^{s_3}s_1, 1\right)$$

$$\left(s_1, {}^{s_2}s_1, {}^{s_3}s_1, s_3\right)$$

$$\left(s_1, {}^{s_2}s_1, 1, s_3\right)$$

$$\left(s_1, {}^{s_2}s_1, s_2, s_3\right)$$

$$\left(s_1, 1, s_2, s_3\right)$$
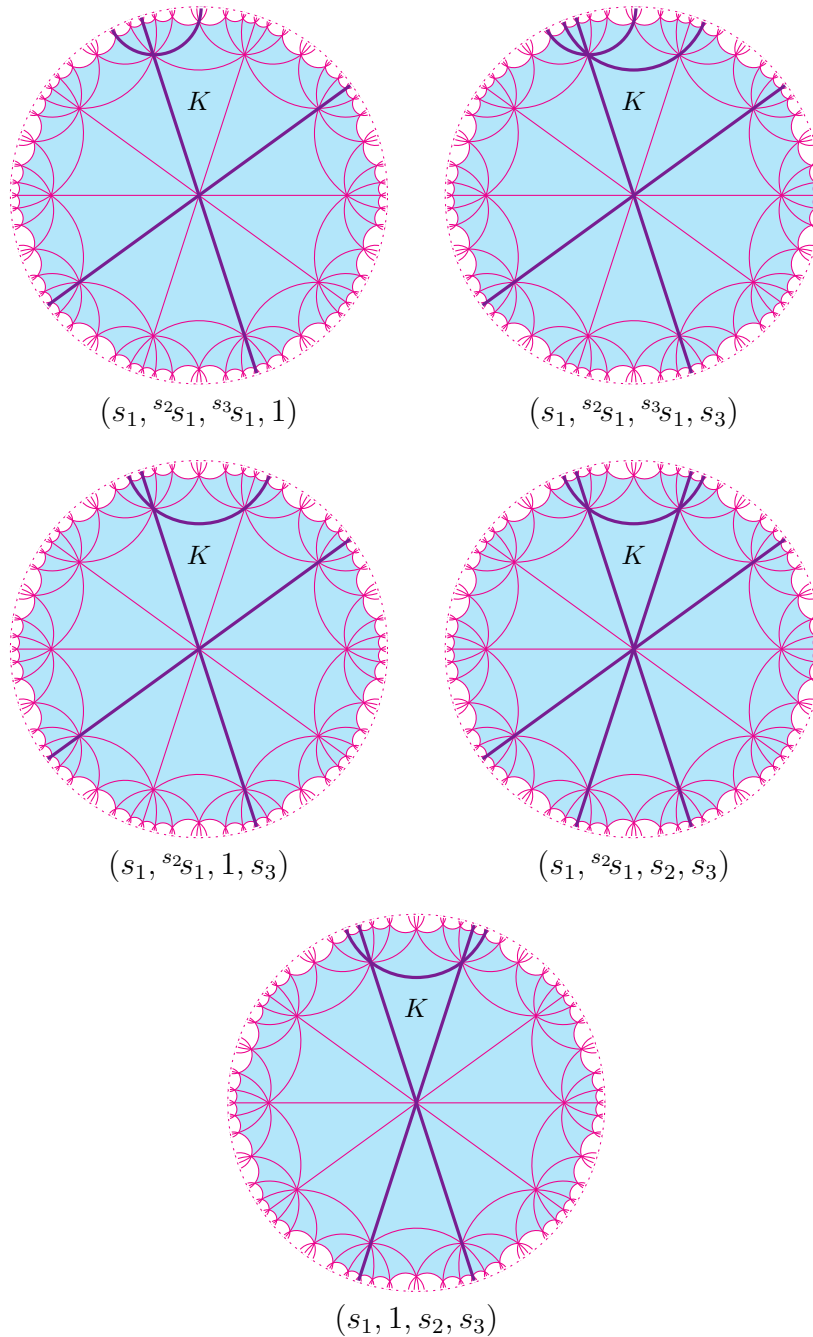
Figure 4.15: Transforming a stabilisation of one reflection generating tuple into a stabilisation of the standard generating tuple. Trivial generators are not shown.

m_{ij}'s are finite the action is co-compact. We summarise this well-known fact in the following Proposition.

PROPOSITION 4.50: *Let (W, S) be a rank three Coxeter system, and define*

$$\Delta = \frac{1}{m_{12}} + \frac{1}{m_{23}} + \frac{1}{m_{31}}.$$

Then (W, S) acts by isometries on $\mathbb{X}^2$ where

$$\mathbb{X}^2 = \begin{cases} \mathbb{S}^2 & \text{if } \Delta > 1 \\ \mathbb{E}^2 & \text{if } \Delta = 1 \\ \mathbb{H}^2 & \text{if } \Delta < 1. \end{cases}$$

If $\{m_{12}, m_{23}, m_{31}\} = \{2, 2, \infty\}$ then $\mathbb{X}^2 = \mathbb{E}^2$ and the action has fundamental domain an half-infinite strip; otherwise the action has fundamental domain a finite area triangle with interior angles π/m_{12}, π/m_{23}, and π/m_{31}. If all m_{ij}'s are finite, then the fundamental domain is compact.

We now want to use the characterisation of reflection equivalence classes to classify all reflection generating triples in triangle groups.

4.5.1   Reducible Coxeter systems

If (W, S) is reducible of rank 3 then it is the direct product $W(\overset{s_1}{\bullet}) \times W(s_2 \bullet\!\!\overset{k}{\rule{1.2em}{0.4pt}}\!\!\bullet s_3)$ for some $k \geqslant 2$, and if $k < \infty$ then it acts on $\mathbb{S}^2$. If $X = (r_1, r_2, r_3)$ is a reflection generating triple, then by Corollary 4.43, after applying a permutation we can assume that r_1 is conjugate to s_1.

But the only such reflection in W is s_1 itself, so we can assume that $X = (s_1, r_2, r_3)$ where (r_2, r_3) generates $W(s_2 \bullet\!\!\overset{k}{\rule{1.2em}{0.4pt}}\!\!\bullet s_3)$. The reflection equivalence classification for dihedral groups is the same as the Nielsen equivalence classification, see Theorem 2.1. The presence of s_1 does not identify any otherwise inequivalent generating pairs since s_1 lies in the centre of W.

4.5.2 Irreducible spherical and affine Coxeter systems

If (W, S) is a Weyl group or affine then its Coxeter-Dynkin diagram is one of the following:
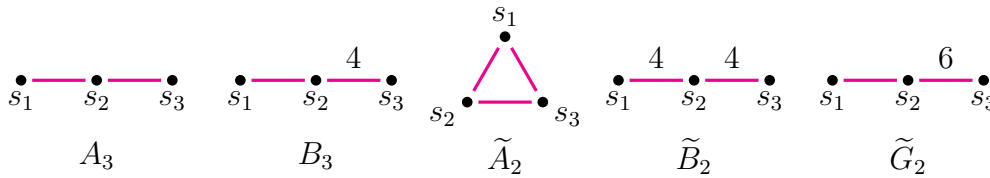
Figure 4.16: Irreducible rank 3 Weyl group or affine Coxeter system Coxeter-Dynkin diagrams.

In particular, Corollary 4.45 guarantees that all reflection generating triples are reflection equivalent to (s_1, s_2, s_3). The only remaining case is the dodecahedral group.

Example 4.51. Consider $W(H_3)$ which has Coxeter-Dynkin diagram
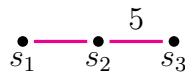
Given a reflection generating triple $X = (r_1, r_2, r_3)$ we can consider the triple of angles $(\angle r_1 r_2, \angle r_2 r_3, \angle r_3 r_1)$. By Theorem 4.44, we can assume that X is a triple where these angles are non-obtuse, ie they lie in the set $\{\pi/2, \pi/3, \pi/5, 2\pi/5\}$. Thinking geometrically, K^X is a spherical triangle with internal angles from that set and hence is uniquely defined up to congruence. Considering all finitely many possible triples of angles, there are only five which are realised by reflection triples X up to permutations:

$$\left(\frac{\pi}{2}, \frac{\pi}{3}, \frac{\pi}{5}\right), \left(\frac{\pi}{2}, \frac{\pi}{5}, \frac{2\pi}{5}\right), \left(\frac{\pi}{3}, \frac{\pi}{3}, \frac{2\pi}{5}\right), \left(\frac{\pi}{2}, \frac{\pi}{3}, \frac{2\pi}{5}\right), \left(\frac{2\pi}{5}, \frac{2\pi}{5}, \frac{2\pi}{5}\right).$$

Up to overall conjugation and permutation, there is just one reflection triple corresponding to each of these triples of angles. Respectively, these are:

$$(s_1, s_2, s_3), \ (s_1, {}^{s_2}s_3, s_3), \ (s_1, s_2, {}^{s_3}s_2), \ (s_1, {}^{s_2 s_3 s_1}s_2, s_3), \ ({}^{s_1 s_2}s_3, {}^{s_2}s_3, s_3).$$

The corresponding hyperplane arrangement for each is shown in Figure 4.17. They do not all represent distinct reflection equivalence classes. First:

$$\left(s_1, {}^{s_2}s_3, s_3\right) \rightarrow \left(s_1, {}^{s_2}s_3, {}^{({}^{s_2}s_3)}s_3\right) = \left(s_1, {}^{s_2}s_3, {}^{s_3}s_2\right)$$

$$\rightarrow \left(s_1, {}^{({}^{s_3}s_2)}({}^{s_2}s_3), {}^{s_3}s_2\right) = \left(s_1, s_2, {}^{s_3}s_2\right)$$

and second:

$$\left(s_1, {}^{s_2 s_3 s_1}s_2, s_3\right) \rightarrow \left(({}^{s_2 s_3 s_1 s_2})s_1, {}^{s_2 s_3 s_1}s_2, s_3\right) = \left({}^{s_1 s_2}s_3, {}^{s_2 s_3 s_1}s_2, s_3\right)$$

$$\rightarrow \left({}^{s_1 s_2}s_3, {}^{({}^{s_1 s_2}s_3)}({}^{s_2 s_3 s_1}s_2), s_3\right) = \left({}^{s_1 s_2}s_3, {}^{s_2}s_3, s_3\right).$$

These equivalences can be seen in Figure 4.17. In both cases, two of the generators r and r' generate a parabolic subgroup isomorphic to Dih_5. Each pair of equivalent tuples is related by conjugating r and r' by rr' or its inverse, a rotation through $4\pi/5$. Since there are only finitely many reflections in $W(H_3)$ one can check exhaustively that there are no other reflection equivalences and hence there are three reflection equivalence classes overall.

4.5.3 Hyperbolic Coxeter systems

There are a finite number of hyperbolic triangle groups to which we can apply Corollary 4.45. For a hyperbolic triangle group which do not fall into this category, we can proceed in a similar way to how we tackled the H_3 case. Given m_{12}, m_{23}, and m_{31} there are only finitely many possibilities for the triple of angles $(\angle r_1 r_2, \angle r_2 r_3, \angle r_3 r_1)$, and many of these are not realised by any reflection generating triple (r_1, r_2, r_3). Often many triples can be excluded for simple reasons, see the example below.

LEMMA 4.52: *Let (W, S) be a freely indecomposable hyperbolic triangle group. Suppose that $(0, 0, \pi k/m)$ is a triple of non-obtuse angles for some $m \in \{m_{12}, m_{23}, m_{31}\}$. If there is a reflection generating triple $X = (r_1, r_2, r_3)$ which realises these angles, then X is reflection equivalent to a triple of reflections whose triple of angles is $(0, \pi k/m, \theta)$ where $0 < \theta \leqslant \pi/2$.*

(a) (s_1, s_2, s_3)

(b) $(s_1, {}^{s_2}s_3, s_3)$

(c) $(s_1, s_2, {}^{s_3}s_2)$

(d) $(s_1, {}^{s_2 s_3 s_1}s_2, s_3)$

(e) $({}^{s_1 s_2}s_3, {}^{s_2}s_3, s_3)$
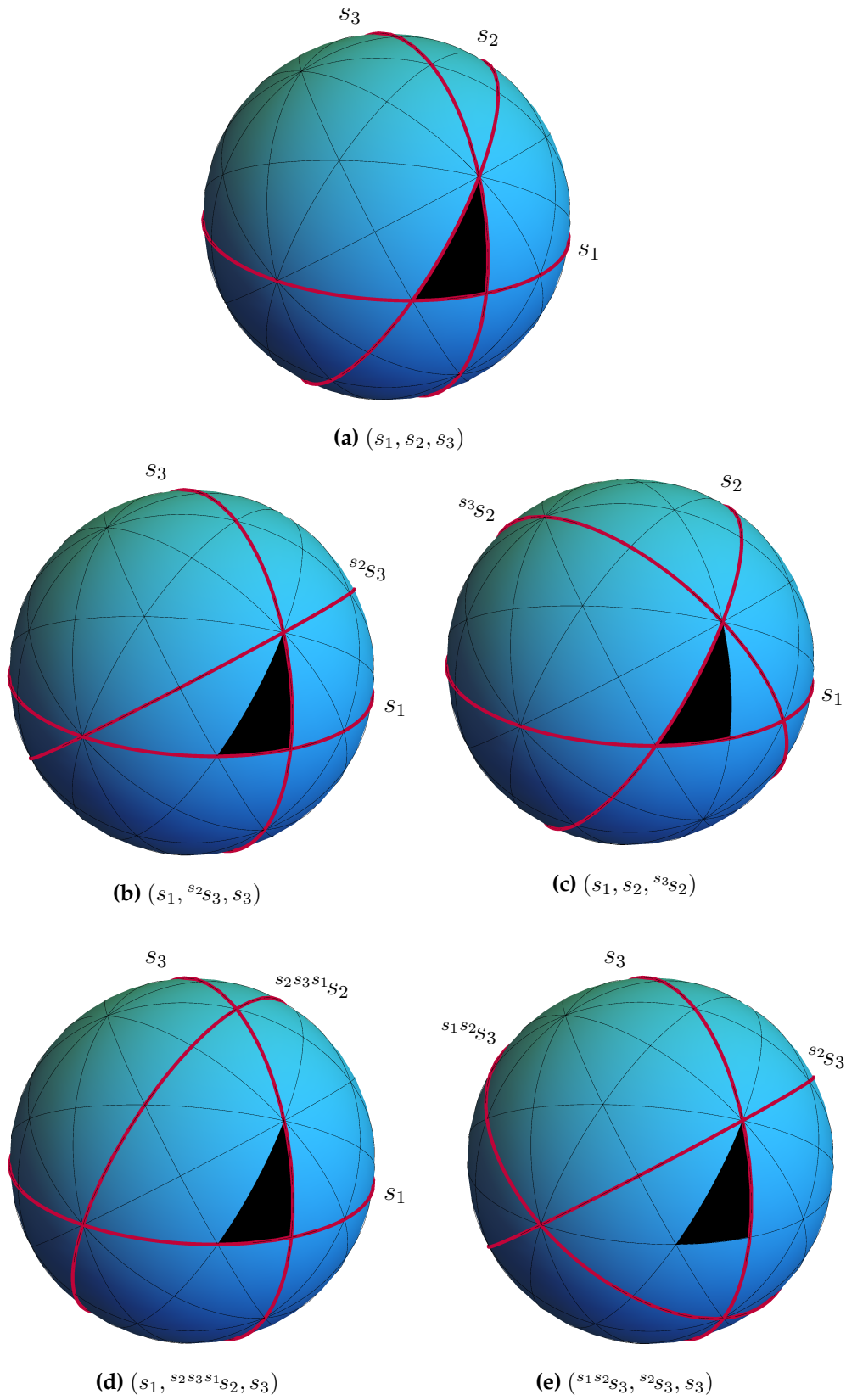
Figure 4.17: Five generating tuples of reflections for $W(H_3)$. In each case the fundamental chamber is shown in black, and the hyperplanes fixed by the generators are shown in red. The pairs of tuples (b) and (c), and (d) and (e) are reflection equivalent.

Proof. We have

$$\angle r_1 r_2 = 0 = \angle r_2 r_3, \text{ and } \angle r_3 r_1 = k\pi/m.$$

First note that $k > 0$, because otherwise X generates a freely decomposable Coxeter group isomorphic to $\mathbb{Z}_2 * \mathbb{Z}_2 * \mathbb{Z}_2$. Applying Theorem 4.40 to X, we must be able to apply a sequence of type II moves to show that there is a reflection $r \in \langle r_1, r_3 \rangle$ such that $\angle r r_2 = \theta > 0$, because otherwise the Coxeter generating tuple $\widetilde{X}$ this Theorem 4.44 produces corresponds to a freely decomposable Coxeter system.

Now, r is conjugate in $\langle r_1, r_3 \rangle$ to one of r_1 or r_3, say $r = {}^w r_3$. Applying a suitable sequence of partial conjugations it follows that X is reflection equivalent to $({}^w r_1, r_2, {}^w r_3) = ({}^w r_1, r_2, r)$, which has the corresponding triple of angles $(0, \theta, \pi k/m)$. That $\angle {}^w r_3 r_1 = 0$ can be seen from Figure 4.18. If $\theta > \pi/2$ we can replace ${}^w r_1$ by ${}^{rw} r_1$ which replaces θ with $\pi - \theta$. Thus we have got th triple of angles we needed, up to a permutation. $\qquad\square$
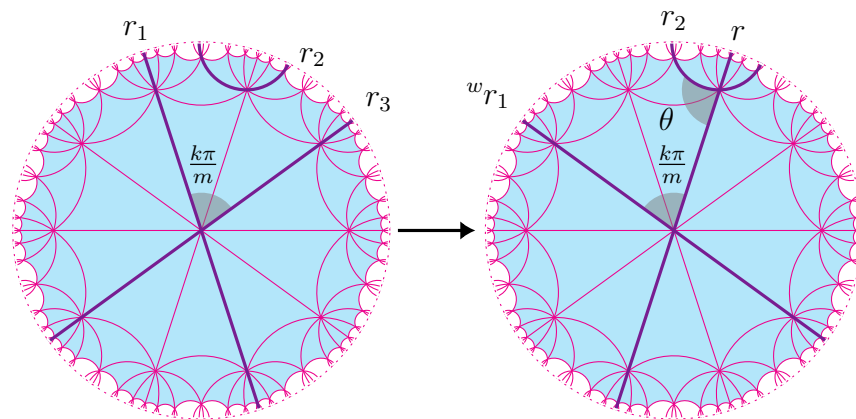
Figure 4.18: Transforming a reflection generating tuple with angles $(0, 0, k\pi/m)$ into one with angles $(0, k\pi/m, \theta)$.

Up to an overall conjugation there are only finitely many reflection generating triples with a given triple of angles (at most one of which is 0), and one can enumerate these. We omit the details, but the idea is as follows. Fix a discrete co-finite action of W by isometries on $\mathbb{H}^2$ such that the generators act by reflections, and let $\rho\colon W \to \mathrm{Isom}(\mathbb{H}^2)$ be the corresponding representation. In particular, this means that if $m_{ij} = \infty$ for some pair (i, j) then the product $s_i s_j$ acts by a parabolic isometry. The geometry of the hyperbolic plane and of the hyperplane arrangement

coming from the set of all reflections in W allow one to find all finitely many reflection triples (up to conjugation) which realise a certain triple of angles. One can then apply Corollary 4.42 to check whether this triple generates W. Lemma 4.52 reduces the space of reflection triples which must be checked.

Example 4.53. We illustrate what happens in the 555 triangle group, which has Coxeter-Dynkin diagram
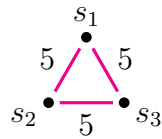
Up to permutations, the possible triples of angles are

$$\left(\tfrac{\pi}{5}, \tfrac{\pi}{5}, \tfrac{\pi}{5}\right), \qquad \left(\tfrac{\pi}{5}, \tfrac{\pi}{5}, \tfrac{2\pi}{5}\right), \qquad \left(\tfrac{\pi}{5}, \tfrac{2\pi}{5}, \tfrac{2\pi}{5}\right),$$
$$\left(\tfrac{\pi}{5}, \tfrac{\pi}{5}, 0\right), \qquad \left(\tfrac{\pi}{5}, \tfrac{2\pi}{5}, 0\right), \qquad \left(\tfrac{\pi}{5}, 0, 0\right),$$
$$\left(\tfrac{2\pi}{5}, \tfrac{2\pi}{5}, \tfrac{2\pi}{5}\right), \qquad \left(\tfrac{2\pi}{5}, \tfrac{2\pi}{5}, 0\right), \qquad \left(\tfrac{2\pi}{5}, 0, 0\right).$$

We can immediately exclude $\left(\tfrac{\pi}{5}, \tfrac{2\pi}{5}, 0\right)$ and $\left(\tfrac{\pi}{5}, 0, 0\right)$ as these are sharp-angled and so any corresponding reflection triple cannot generate W. We can also ignore $\left(\tfrac{2\pi}{5}, 0, 0\right)$ by Lemma 4.52. The angle sums for $\left(\tfrac{\pi}{5}, \tfrac{2\pi}{5}, \tfrac{2\pi}{5}\right)$ and $\left(\tfrac{2\pi}{5}, \tfrac{2\pi}{5}, \tfrac{2\pi}{5}\right)$ are greater than or equal to π, and so there are no hyperbolic triangles with these internal angles. Finally we can exclude $\left(\tfrac{\pi}{5}, \tfrac{\pi}{5}, \tfrac{2\pi}{5}\right)$. This is because a hyperbolic triangle with these angles has area $\pi/5$, but it is tiled by copies of the fundamental domain for W acting on $\mathbb{H}^2$, which is a hyperbolic triangle with angles $\left(\tfrac{\pi}{5}, \tfrac{\pi}{5}, \tfrac{\pi}{5}\right)$, and hence area $2\pi/5$, which is not possible. Thus we are left with three viable triples:

$$\left(\frac{\pi}{5}, \frac{\pi}{5}, \frac{\pi}{5}\right), \ \left(\frac{\pi}{5}, \frac{2\pi}{5}, 0\right), \ \left(\frac{2\pi}{5}, \frac{2\pi}{5}, 0\right).$$

It turns out that each of these can be realised, and in a unique way up to an overall conjugation and a diagram automorphism. Examples of reflection triples which do this are

$$X_1 = (s_1, s_2, s_3), \ X_2 = (s_1, s_2, {}^{s_3}s_1), \ X_3 = (s_1, {}^{s_2}s_1, {}^{s_3}s_1),$$

respectively. The hyperplane arrangements corresponding to these generating tuples are shown in Figure 4.19.

$$X_1 = (s_1, s_2, s_3) \qquad X_2 = (s_1, s_2, {}^{s_3}s_1) \qquad X_3 = (s_1, {}^{s_2}s_1, {}^{s_3}s_1)$$
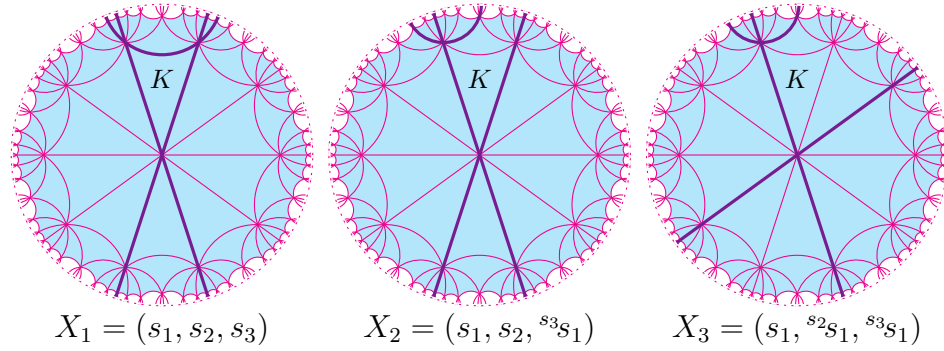
Figure 4.19: The hyperplane arrangements associated to three reflection generating tuples of the 555 triangle group. These are visualised in a finite portion of hyperbolic plane, compare with Figure 4.14.

It is not obvious that any of these are reflectionequivalent to each other but, unlike the case of $W(H_3)$, W is infinite, so it is not clear how to verify this by an exhaustive check. Instead, we need to use an invariant. The 555 triangle group was chosen for this example since it is the simplest example for which we have a chance of usefully applying the invariant we developed in Section 2.2.4.

Note that $X_1 = S$, so we compute χ_η where $\eta\colon \mathbb{Z}W \to \mathbb{Z}_5 = A$ is the composition $\xi \circ \eta^{\mathrm{ab}}$ mapping each generator s_i to -1. Thus the correction ideal is the trivial ideal, and $A_W = \{\pm 1\}$.

For X_2 we have

$$\phi(\partial_S(\mathbf{X_2})) = \phi \begin{pmatrix} \partial_{x_1} x_1 & \partial_{x_2} x_1 & \partial_{x_3} x_1 \\ \partial_{x_1} x_2 & \partial_{x_2} x_2 & \partial_{x_3} x_2 \\ \partial_{x_1} x_3 x_1 x_3 & \partial_{x_2} x_3 x_1 x_3 & \partial_{x_3} x_3 x_1 x_3 \end{pmatrix}$$

$$= \phi \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ x_3 & 0 & 1 + x_3 x_1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ s_3 & 0 & 1 + s_3 s_1 \end{pmatrix}.$$

Then we can compute

$$\chi_\eta(X_2) = \det(\eta(\phi((\partial_S(\mathbf{X_2})))) = \det \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -1 & 0 & 2 \end{pmatrix} = 2.$$

Since this does not lie in A_W, we can conclude that $S = X_1$ and X_2 are not Nielsen equivalent, and so definitely not reflection equivalent. On the other hand, for X_3 we have

$$\phi(\partial_S(\mathbf{X_3})) = \phi \begin{pmatrix} \partial_{x_1} x_1 & \partial_{x_2} x_1 & \partial_{x_3} x_1 \\ \partial_{x_1} x_2 x_1 x_2 & \partial_{x_2} x_2 x_1 x_2 & \partial_{x_3} x_2 x_1 x_2 \\ \partial_{x_1} x_3 x_1 x_3 & \partial_{x_2} x_3 x_1 x_3 & \partial_{x_3} x_3 x_1 x_3 \end{pmatrix}$$

$$= \phi \begin{pmatrix} 1 & 0 & 0 \\ x_2 & 1 + x_2 x_1 & 0 \\ x_3 & 0 & 1 + x_3 x_1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ s_2 & 1 + s_2 s_1 & 0 \\ s_3 & 0 & 1 + s_3 s_1 \end{pmatrix}.$$

Then we can compute

$$\chi_\eta(X_3) = \det(\eta(\phi((\partial_S(\mathbf{X_2})))) = \det \begin{pmatrix} 1 & 0 & 0 \\ -1 & 2 & 0 \\ -1 & 0 & 2 \end{pmatrix} = -1 \in A_W.$$

This invariant cannot distinguish X_1 and X_3, so we cannot tell whether or not they are equivalent.

4.A Appendix: simplicial complexes

In this Appendix we introduce the the relevant background on simplicial complexes which are used to define the *Davis complex*.

4.A.1 Posets

Before talking about complexes, we introduce posets which are extremely useful when defining and manipulating simplicial complexes.

Definition 4.54. A **poset** (short for **p**artially **o**rdered **set**) is a pair $(A, \preceq)$ consisting of a non-empty set A and a binary relation $\preceq$ on A which satisfies:

1. (reflexivity) $a \preceq a$ for all $a \in A$

2. (antisymmetry) if $a \preceq b$ and $a \succeq b$ then $a = b$

3. (transitivity) if $a \preceq b$ and $b \preceq c$ then $a \preceq c$

We write $a \prec b$ if $a \preceq b$ and $a \neq b$.

Given two posets $(A, \preceq)$ and $(A', \sqsubseteq)$, a **poset isomorphism** between them is a bijective map $f \colon A \to A'$ such that for any $a, b \in A$, $a \preceq b$ if and only if $f(a) \sqsubseteq f(b)$.

Example 4.55. Let P be any set and A a non-empty collection of subsets of P. Then $(A, \subseteq)$ is a poset ordered by set inclusion.

Every poset naturally gives rise to a *dual* poset.

Definition 4.56. Let $(A, \preceq)$ be a poset, then the **opposite** relation on A to $\preceq$ is the binary relation $\overset{\mathrm{op}}{\preceq}$ defined by

$$a \overset{\mathrm{op}}{\preceq} b \text{ if and only if } a \succeq b,$$

for all $a, b \in A$.

It is straightforward to check that $(A, \preceq)^{\mathrm{op}} := (A, \overset{\mathrm{op}}{\preceq})$ is another poset structure on A. A second way to create a new poset from an old one is to look at the *chains* it contains.

Definition 4.57. Let $(A, \preceq)$ be a poset. A **chain** is a subset $\{a_0, a_1, \dots, a_k\}$ of A such that $a_0 \prec a_1 \prec \cdots \prec a_k$. Such a chain has **length** k. Denote the set of chains in $(A, \preceq)$ by $\mathrm{Ch}(A, \preceq)$.

The set of chains in a poset is naturally ordered by inclusion, giving rise to the poset $(\mathrm{Ch}(A, \preceq), \subseteq)$.

Definition 4.58. The **height** of a poset is the maximum length of a chain in that poset. If no maximum exists, the poset will have infinite height.

Assumption 4.59. From now on we shall only consider posets whose underlying set is countable and which have finite height.

4.A.2 Abstract and geometric simplicial complexes

We can now define *abstract simplicial complexes*.

Definition 4.60. Let V be a countable set which will be called the **set of vertices**. An **abstract simplicial complex** is the pair $(V, \mathcal{A})$, where $\mathcal{A}$ is a collection of subsets of V such that

1. Each subset is finite

2. For each $v \in V$, $\{v\} \in \mathcal{A}$

3. $\mathcal{A}$ is closed under taking subsets

An element a of $\mathcal{A}$ is called a k-**simplex** where $k = \#a - 1$, and k is the **dimension** of a. If a is a k-simplex then any subset $b \subset a$ again belongs to $\mathcal{A}$ and is called a **face** of a. The collection $\mathcal{A}$ is called the **set of simplices**.

For an integer $k \geqslant 0$, the k-**skeleton** of an abstract simplicial complex is

$$(V, \{a \in \mathcal{A} \mid \text{dimension of } a \text{ is at most } k).$$

$\mathcal{A}$ contains the empty set $\emptyset$ which is the unique simplex of dimension -1 called the **empty simplex**. An abstract simplicial complex $\mathcal{A}$ has a natural poset structure under set inclusion; the poset $(\mathcal{A}, \subseteq)$ is called the **poset of simplices** of $\mathcal{A}$.

To any abstract simplicial complex we can associate a topological space which is simply called a *simplicial complex*.

Definition 4.61. Let $\mathcal{A}$ be an abstract simplicial complex with vertex set V and fix some order on V. Let $\mathbb{R}^V := \bigoplus_{v \in V} \mathbb{R}$ be the real vector space which is the direct sum of copies of $\mathbb{R}$ indexed by V. In particular, all but finitely many of the coordinates of any point in $\mathbb{R}^V$ are zero. For $a \in \mathcal{A}$, let

$$\sigma_a := \{(x_v)_{v \in V} \mid x_v \geqslant 0 \text{ with equality for all } v \notin a \text{ and, } \textstyle\sum_{v \in a} x_v = 1\}.$$

This set is the convex hull of the standard basis vectors in $\mathbb{R}^V$ corresponding to the vertices in a, it is called the **(geometric) simplex** associated to a. Consider

the set $\mathcal{C} = \bigcup_{a \in \mathcal{A}} \sigma_a$ with the subspace topology induced from $\mathbb{R}^V$. The **(geometric) simplicial complex** associated to $\mathcal{A}$ consists of the space $\mathcal{C}$ together with the collections of subsets $\mathcal{C}_k = \{\sigma_a \mid a \in \mathcal{A} \text{ is a } k\text{-simplex}\}$ for $k \in \{0, 1, \dots\}$.

Given a simplicial complex $(\mathcal{C}, \{\mathcal{C}_k\}_{k \geqslant 0})$ one can recover its original abstract simplicial complex as follows. Let V be the set of points in $x \in \mathbb{R}^V$ such that $\{x\} \in \mathcal{C}_0$. Then define $\mathcal{A}$ to be

$$\bigcup_{k=0}^{\infty} \bigcup_{\sigma \in \mathcal{C}_k} \{\{x \in V \mid x \in \sigma\}\}.$$

Assumption 4.62. Every abstract simplicial complex can be turned into a geometric simplicial complex, and the abstract simplicial complex can be recovered from this topological space together with its decomposition into geometric simplices. Therefore, we will make no distinction between abstract and geometric simplicial complexes hereafter.

4.A.3 The geometric realisation of a poset

We have seen that every abstract simplicial complex is a poset, however not every poset is isomorphic to the poset of simplices of a simplicial complex (recall Assumption 4.59, we only consider countable finite height posets). Nevertheless there is a way to associate a simplicial complex to any poset.

Definition 4.63. The **geometric realisation** of a poset $(A, \preceq)$ is the simplicial complex $(A, \mathrm{Ch}(A, \preceq))$.

This name is standard, albeit a little confusing since the definition is given as an abstract simplicial complex, not a geometric one. Moreover, if $(\mathcal{A}, \subseteq)$ is the poset of simplices of an abstract simplicial complex $(V, \mathcal{A})$, then its geometric realisation as a poset is *not* the corresponding geometric simplicial complex from Definition 4.61.

The poset of simplices of $(A, \mathrm{Ch}(A, \preceq))$ is exactly the poset $(\mathrm{Ch}(A, \preceq), \subseteq)$ mentioned after Definition 4.57.

4.A.4 Cones and subdivisions

We will define two operations which can be applied to modify a simplicial complex. The first is taking the *simplicial cone* of a complex.

Definition 4.64. Let $(V, \mathcal{A})$ be a simplicial complex. The **cone** over this complex is the simplicial complex

$$\mathrm{Cone}(V, \mathcal{A})) = (V \cup \{v_0\}, \mathcal{A} \cup \{a \cup \{v_0\} \mid a \in \mathcal{A}\}),$$

where $v_0 \notin V$ is a new vertex called the **cone point**.

Topologically, the cone over a complex $\mathcal{C}$ can be thought of as the the space $\mathcal{C} \times [0, 1]/ \sim$, where $(x, 1) \sim (y, 1)$ for all $x, y \in \mathcal{C}$. The image of $(\mathcal{C}, 1)$ in the quotient is a single point, the cone point. The second operation does not change the topology of the complex, only its combinatorial structure.

Definition 4.65. The **barycentric subdivision** of a simplicial complex $(V, \mathcal{A})$ is the geometric realisation of the poset of simplices $(\mathcal{A}, \subseteq)$, denoted $\mathrm{Bs}(V, \mathcal{A})$.

The 0-simplices of $\mathrm{Bs}(V, \mathcal{A})$ are the chains in $(\mathcal{A}, \subseteq)$ of length 0, ie chains of the form $\{a\}$ for $a \in \mathcal{A}$. The 0-simplex $\{a\}$ of $\mathrm{Bs}(V, \mathcal{A})$ is called the *barycentre* of the simplex a in $(V, \mathcal{A})$. Thus the vertex set of $\mathrm{Bs}(V, \mathcal{A})$ is exactly the set of simplices of $(V, \mathcal{A})$, $\mathcal{A}$.

Remark 4.66 (Ordering vertices in $\mathrm{Bs}(V, \mathcal{A})$). The poset of cells $(\mathcal{A}, \subseteq)$ therefore defines a partial ordering on the vertices of $\mathrm{Bs}(V, \mathcal{A})$. Given a simplex of $\mathrm{Bs}(V, \mathcal{A})$, ie a chain $\{a_0, a_1, \ldots, a_k\}$ where $a_0 \subset a_1 \subset \cdots \subset a_k$, we can say that a_0 is the *minimum vertex* in this simplex.

4.A.5 Cayley graph of a Coxeter system

A special case of a simplicial complex is a graph, ie a simplicial complex which contains no simplices with dimension greater than 1. Here we define the *Cayley graph* of a Coxeter system.

Definition 4.67. Let (W, S) be a Coxeter system. The **Cayley graph** of (W, S) is the simplicial complex $(W, \mathcal{A})$ where

$$\mathcal{A} = \{\emptyset\} \cup \{\{w\} \mid w \in W\} \cup \{\{w, ws\} \mid w \in W, s \in S\}.$$

Notice that since each $s \in S$ is an involution, the we have the equality of the 1-simplices

$$\{ws, wss\} = \{ws, w\} = \{w, ws\}.$$

It follows that a pair of vertices in the Cayley graph are contained in most one 1-simplex. This is contrary to some other standard definitions of the Cayley graph in which the vertices w and ws would be contained in two distinct edges which are oriented in different directions.

Chapter 5

Nielsen equivalence in right-angled Coxeter groups

RACGs ARE A VERY IMPORTANT class of Coxeter groups to study with respect to Nielsen equivalence. Notwithstanding their centrality in much of geometric group theory, there are many reasons to suspect that their generating tuples are well-behaved when it comes to Nielsen equivalence. As Γ varies over finite simplicial graphs, W_Γ interpolates between $\mathbb{Z}_2 * \cdots * \mathbb{Z}_2$ when Γ is totally disconnected, and $\mathbb{Z}_2 \times \cdots \times \mathbb{Z}_2$ when Γ is complete. In both of these cases, all generating tuples are either reducible, or Nielsen equivalent to the standard one (albeit for quite different reasons: see Theorem 1.20 and Theorem 1.19 respectively).

As discussed in Section 2.1, since RACGs are even, their algebraic rank equals their Coxeter rank, so any Coxeter generating tuple is minimal. Additionally they are rigid (see Definition 2.9) so any two Coxeter systems for the same RACG have isomorphic diagrams. As a consequence it is not unreasonable to hope that for many RACGs, all generating tuples are either reducible or Nielsen equivalent to the standard one.

Our tool for studying RACGs will be certain labelled *cube complexes* introduced by Palavi Dani and Ivan Levcovitz in [31]. On the face of it these are quite distinct from the tool we used in the previous Chapter: the Davis complex. In fact, the cube complexes we will use to study a RACG W_Γ are very closely related to certain quotients on subsets of the Davis complex Σ_Γ. In the first Section we will define

these complexes and summarise the relevant results from [31]

5.1 Cube complexes and completion sequences

The work of Dani and Levcovitz is closely inspired by Stallings' study of subgroups of free groups [105]. In turn, it was generalised to study subgroups of the fundamental groups of $\mathrm{CAT}(0)$ cube complexes in [7]. We begin by recalling the definition of a cube complex and introducing a class of labelled cube complexes. For a detailed introduction to the general theory of cube complexes, see [57].

5.1.1 Cube complexes and Γ-labellings

Definition 5.1. A **cube complex** Ω is a cell complex in which all cells are identified with Euclidean unit cubes $\left[-\frac{1}{2}, \frac{1}{2}\right]^k$ and attaching maps restrict to Euclidean isometries on the faces of the cubes.

A **mid-cube** of a cube c in a cube complex is the result of restricting one of the coordinates of c to be 0. The **cubical subdivision** $\mathrm{Sub}(\Omega)$ of Ω is the cube complex obtained by subdividing the interiors of each k-cube into 2^k cubes by cutting along all of the mid-cubes.
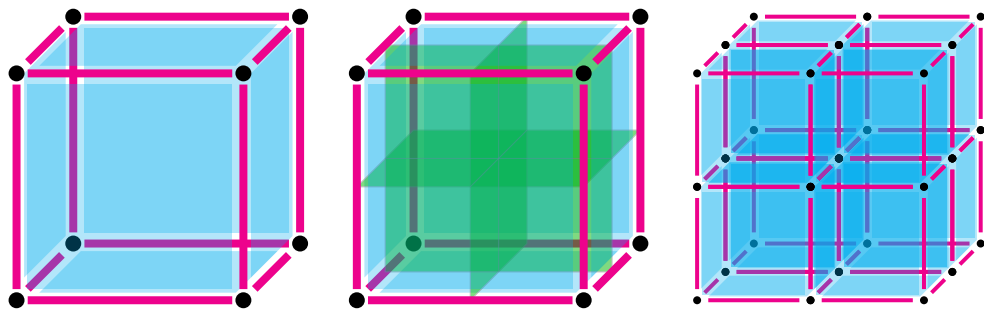
Figure 5.1: From left to right: a 3-cube, its three mid-cubes, and its cubical subdivision.

Two edges of a cube c are **opposite** each other if they intersect the same mid-cube. Consider the equivalence relation on the edges of Ω which is the transitive closure of the opposite relation. An equivalence class under this relation is called a **wall** of Ω, see Figure 5.2 for an example. For a fixed wall w in Ω, the **hyperplane dual to** w is the union of all mid-cubes which meet an edge in w. Note that hyper-

planes in a cube complex Ω can be identified with subcomplexes of $\mathrm{Sub}(\Omega)$ and inherit the structure of a cube complex.
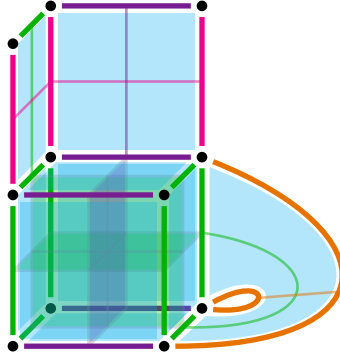
Figure 5.2: An example of a cube complex which contains four walls and hence four hyperplanes: red, orange, green, and purple. The green hyperplane self-intersects, showing that edges which are not opposite in a cube may nevertheless be in the same wall in a cube complex.

We use cube complexes to encode information about a subgroup of a RACG W_Γ with presentation diagram Γ. To this end we label the cube complexes.

Definition 5.2. Let Γ be a finite simplicial graph. A Γ-**complex** is a cube complex Ω where each hyperplane is labelled by a vertex of Γ such that, if two hyperplanes intersect, their labels are distinct and adjacent in Γ.

The **type** of a k-cube c in a Γ complex is the set of k distinct labels of the hyperplanes which intersect it.

Notice that a labelling of the hyperplanes induces a dual labelling on the walls of Ω, and hence on all of the edges. Thus, the type of a cube c is also the set of its edge labels.

5.1.2   The complex of groups $\mathcal{O}_\Gamma$

The labelling of a Γ-complex Ω encodes a map $\pi_1(\Omega) \to W_\Gamma$. To see this, we first define a complex of groups $\mathcal{O}_\Gamma$ whose fundamental group is W_Γ (see [56] for an account to complexes of groups in general, as well as Remark 4.18). One way to describe $\mathcal{O}_\Gamma$ is to start with the Davis complex Σ of W_Γ (recall the construction in Section 4.1, see Section 1.2 of [32] for a summary of this in the right-angled case).

Taking the dual cell structure from Theorem 4.14 yields a CAT(0) cube complex having the Cayley graph of (W_Γ, S) as its 1-skeleton.

This cube complex has k-cubes attached equivariantly wherever the 1-skeleton of a k-cube can be found. These correspond to complete subgraphs with k vertices in Γ. Taking the cubical subdivision of Σ, W_Γ acts by cellular isometries so that the stabiliser of every cell fixes the cell point-wise. Define $\mathcal{O}_\Gamma$ to be the quotient $\mathrm{Sub}(\Sigma)/W_\Gamma$. As a cell-complex, $\mathrm{Sub}(\Sigma)/W_\Gamma$ is contractible, so it lifts to a subset of $\mathrm{Sub}(\Sigma)$. We can label cells in the quotient by the stabilisers of their lifts, giving the quotient the desired complex of groups structure.

Thought of another way, the quotient $\mathrm{Sub}(\Sigma)/W_\Gamma$ essentially coincides with the strict fundamental domain $K(W_\Gamma, V\Gamma)$, just with a different cell structure. Giving $K(W_\Gamma, V\Gamma)$ a complex of groups structure according to the mirror structure and family of groups over $V\Gamma$ is another way to define $\mathcal{O}_\Gamma$.

Since Σ is CAT(0), it is simply connected, so we can identify the fundamental group of the complex of groups, $\pi_1(\mathcal{O}_\Gamma)$, with W_Γ. Notice that, since the 1-skeleton of Σ is the Cayley graph of (W_Γ, S) which comes with a natural labelling by $V\Gamma = S$, Σ is a Γ-complex. This induces a labelling of some edges of $\mathcal{O}_\Gamma$ by $V\Gamma$.

Now suppose we have a cellular map $\mathrm{Sub}(\Omega) \to \mathcal{O}_\Gamma$ for some Ω. This induces a map $\pi_1(\Omega) \to W_\Gamma$ across which we can pull back the labelling on $\mathcal{O}_\Gamma$ to a Γ-labelling on Ω. Since there is exactly one edge of $\mathcal{O}_\Gamma$ labelled by each vertex of Γ (as W_Γ acts transitively on edges with the same label in Σ), the Γ-labelling on Ω completely determines the map $\Omega \to \mathcal{O}_\Gamma$, and hence $\pi_1(\Omega) \to W_\Gamma$.

5.1.3   Completion sequences of Γ-complexes

In [31], Dani and Levcovitz consider three operations on Γ-complexes, where the result of each is a new Γ-complex. They are as follows.

Edge fold   If Ω contains two edges e and e' which have a common end point and share the same label, then pick orientations for e and e' such that the common vertex is the origin vertex. Then the map $\Omega \to \mathcal{O}_\Gamma$ factors through $\Omega' = \Omega/e \sim e'$, where the edges have been identified by an isometry which matches up the

orientations. The quotient map $\Omega \to \Omega'$ is an **edge fold** and Ω' is a Γ-complex with the labelling induced from Ω.

Compare this to *Stallings folds* shown in Figure 1.1. Edge folds fall into the same four types shown in Figure 5.3, the difference is that there are no fixed orientations on the edges, we are free to pick orientations each time we fold.

(a)   (b)   (c)   (d)
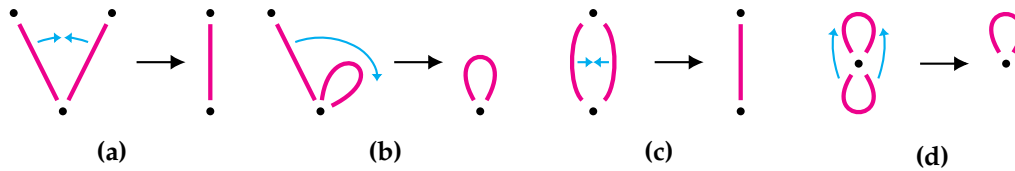
Figure 5.3: The four types of edge folds.

In the present setting, folds of type (a) and (b) are important because they induce a homotopy equivalence between the complex before and after it is folded. So, in particular, they always preserve the property that the fundamental group is free. On the other hand, folds of type (c) and (d) are important because they do not change the vertex set of the complex being folded, which gives us tight control on the connection between *standard completion sequences* (see below), and standard *free* completion sequences in Section 5.4.1.

Cube identification   If Ω contains two k-cubes c and c' (k at least 2) with the same attaching maps (up to a Euclidean isometry of c'), then the map $\Omega \to \mathcal{O}_\Gamma$ factors through $\Omega' = \Omega / c \sim c'$. Here we have identified c and c' with their images in Ω (unlike for edges, when $k \geqslant 2$ there is a unique way to continuously extend the identifications of the boundary attaching maps of c and c' to identifications of the interiors of the cubes). The quotient map $\Omega \to \Omega'$ is a **cube identification**.

Cube attachment   If Ω contains a vertex v with k of the edges incident to v labelled by distinct vertices of Γ which span a complete subgraph, then attach a new k-cube c to Ω to construct a new cube complex Ω' as follows. Give each of the k edges incident to v an orientation such that v is their origin vertex. Similarly orient each of the k edges in c which meet $(-1/2, \ldots, -1/2)$. The attachment map sends the vertex $(-1/2, \ldots, -1/2)$ to v and the k edges incident to that vertex in c

are identified with the k edges which meet at v such that the orientations match. Label all the other edges of c in Ω' by the label of the edge opposite them which is identified with an edge in Ω. The inclusion map $\Omega \to \Omega'$ is a **cube attachment**.

With these operations in mind we make the following definition.

Definition 5.3. A Γ-complex is **folded** if it is not possible to perform an edge fold or cube identification. It is **cube full** if, wherever a cube attachment is possible, there is already a cube present. Given a Γ-complex Ω, a **completion sequence** for Ω is a possibly infinite sequence of Γ-complexes

$$\Omega = \Omega_0 \to \Omega_1 \to \Omega_2 \to \cdots ,$$

which satisfy the following conditions: consecutive Ω_i's differ by one of the operations defined above and the direct limit of the sequence, $\widehat{\Omega}$, is folded and cube full. In this case, $\widehat{\Omega}$ is called a **completion** of Ω.

5.1.4   Standard completion sequences

Dani and Levcovitz show that a completion always exists by constructing a so-called *standard* completion sequence where alternately one performs all possible folds and identifications, followed by all possible cube attachments. More precisely, the construction is as follows.

Standard completion sequence: Suppose that some finite portion of a standard completion sequence of Ω has already been constructed

$$\Omega = \Omega_0 \to \Omega_1 \to \cdots \to \Omega_i,$$

it follows by induction that Ω_i is a finite complex. Suppose Ω_i is not folded. Because it is finite, it contains finitely many edges and cubes, so there are a finite number of folds and cube identifications possible, say j. Performing each of these in turn, the standard completion sequence extends as

$$\Omega_i \to \Omega_{i+1} \to \cdots \to \Omega_{i+j},$$

and Ω_{i+j} is still a finite cube complex.

On the other hand, if Ω_i is a folded complex, then for each vertex v in Ω_i attach all possible maximal cubes in turn which are not already present. Since Ω_i is finite, this involves adding only a finite number of cubes j, so the standard completion sequence extends as

$$\Omega_i \to \Omega_{i+1} \to \cdots \to \Omega_{i+j}.$$

Remark 5.4 (Uniqueness of completion sequences). A standard completion sequence is not unique as the folds, identifications and attachments can be performed in different orders. Additionally, if a fold or attachment involves edge loops (ie edges which have the same start and end vertex), there may be several essentially different ways of folding or attaching. It is also worth noting that, while the result of the folding and identifying step of the construction is necessarily a folded complex, after performing all possible cube attachments at vertices of Ω_i, the resulting complex may not be cube full. This is because the newly attached cubes introduce new vertices where more cubes could be attached, but are not in that step.

PROPOSITION 5.5 (Propositions 3.3 and 3.5 in [31]): *The direct limit of the standard completion sequence is both folded and cube full, and so is a completion. If this completion is a finite Γ-complex, then the standard completion sequence producing it has finite length.*

Completions are useful for studying finitely generated subgroups of W_Γ. Given a finite tuple $X \subseteq W_\Gamma$, let Ω_X be the Γ-complex constructed as follows. Start with a wedge of circles indexed by X, $\bigvee_{x \in X} \mathbb{S}^1$ and call the common vertex v_0, which we set to be the base-point. For each $x \in X$, choose a (reduced) word $t_1 \cdots t_k$ representing x and subdivide the corresponding circle into k edges which are labelled cyclically by $t_1, \ldots, t_k$. Now, if $\widehat{\Omega}_X$ is a completion of Ω_X, we also call it a **completion of** $G = \langle X \rangle$. Many algebraic and geometric properties of $G \leqslant W_\Gamma$ can be easily read off.

5.1.5 Core graphs

It is not the case that Ω_X is uniquely determined by $\langle X \rangle$, or even by X, however Dani and Levcovitz define the *core graph* of $\widehat{\Omega}_X$, which is uniquely defined.

Definition 5.6. Let Ω be a Γ-complex with base point v_0. The **core graph** of (Ω, v_0), denoted $C(\Omega, v_0)$ is the union of all loops based at v_0 in Ω (ie paths which start and end at v_0) which are labelled by reduced expressions (see Definition 1.9)in (W_Γ, S).

Example 5.7. Consider the RACG with presentation diagram:
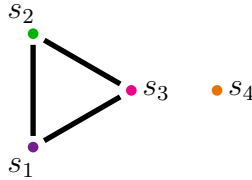
Given the pair $X = (s_2 s_4 s_3, s_3 s_4 s_3)$, we can form Ω_X and compute its completion.
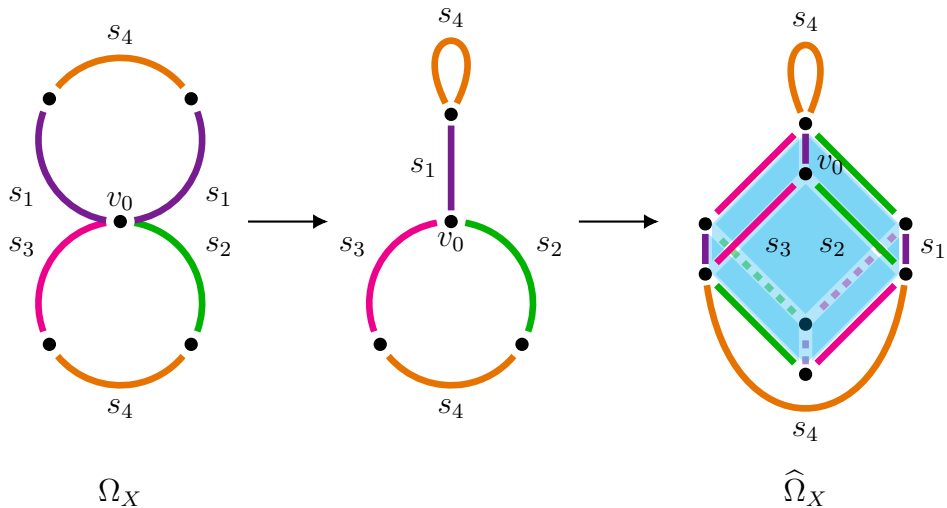
Figure 5.4: An example of a completion sequence.

In this particular case, the image of the intermediate graph in the final completion is the core graph $C(\widehat{\Omega}_X, v_0)$. This can be seen by observing that any based loop which contains any other edge of $\widehat{\Omega}_X$ must contain a sub-path in the 1-skeleton of the 3-cube which traverses two edges which are opposite. Then one can check that the label of such a sub-path can be reduced using Corollary 1.12.

To state the uniqueness of core graphs, we need the following definition which generalises the construction of $\widehat{\Omega}_X$.

Definition 5.8 (Definition 4.3 in [31]). Let Ω be a connected Γ-complex with base point v_0, and let $\mathbb{F}(y_1, \ldots, y_n)$ be the fundamental group of the 1-skeleton of Ω,

where each y_i is a loop in Ω based at v_0. Consider the map $\mathbb{F}(y_1, \ldots, y_n) \to W_\Gamma \colon y_i \mapsto x_i$, where x_i is the element represented by the word which labels y_i. Then we call the group $\langle x_1, \ldots, x_n \rangle$ the **subgroup of W_Γ associated to** Ω. In particular, if the map $\mathbb{F}(y_1, \ldots, y_n) \to W_\Gamma$ is surjective, then Ω represents W_Γ. We say that, if Ω is a completion with associated subgroup G, then Ω is a **completion of** G.

The subgroup of W_Γ associated to a Γ-complex is preserved by taking completion sequences. Thus the subgroup associated to $\widehat{\Omega}_X$ is $\langle X \rangle$. The core graphs of completions have the following property, which follows from Lemma 4.2 of [31].

PROPOSITION 5.9: *Let $\widehat{\Omega}$ be a completion with base point v_0, and G the subgroup of W_Γ associated to $(\widehat{\Omega}, v_0)$. If $w \in G$ and $t_1 \cdots t_k$ is a reduced word representing w, then there is a based loop in $C(\widehat{\Omega}, v_0)$ labelled $t_1 \cdots t_k$.*

The uniqueness of core graphs of completions can now be stated as follows.

THEOREM 5.10 (Proposition 5.3 in [31])**:** *Let (Ω, v_0) and (Ω', v_0') be two based completions with the same associated subgroup, then there is a based isomorphism $C(\Omega, v_0) \to C(\Omega', v_0')$.*

5.1.6 Finite completions and quasiconvexity

In the rest of this Chapter, we are most interested in finite completions. Dani and Levcovitz prove an equivalence between the existence of a finite completion and the associated subgroup being *quasiconvex* in W_Γ. Recall the Davis complex Σ_Γ of W_Γ. In particular, its dual CW structure contains the Cayley graph of W_Γ as its 1-skeleton by Theorem 4.14, and so we can view $G \subset W_\Gamma$ as a subset of Σ_Γ.

Definition 5.11. Let W_Γ be a RACG, and G a subgroup. G is **quasiconvex** in W_Γ if there exists some $M > 0$ such that any geodesic in the Davis complex of W_Γ, whose endpoints lie in G, lies wholly in an M-neighbourhood of G.

For more information of quasiconvexity in general, consult Definition III.Γ.3.4 in [14].

THEOREM 5.12 (Theorem 8.4 in [31])**:** *Let G be a subgroup of W_Γ, then the following are equivalent:*

1. *G is quasiconvex*

2. *Some completion for G is finite*

3. *Every standard completion for G is finite*

In the remainder of this Chapter, we cast our net somewhat wider than just Nielsen equivalence in RACGs and consider all quasiconvex subgroups of RACGs. The reason is that their finite completion sequences ensure that algorithmic processes discussed below terminate.

Assumption 5.13. For the remainder of this Chapter, fix a finite simple graph Γ, and we work in the category of Γ-complexes, ie all spaces are cube complexes with a Γ-labelling, and all maps between these spaces are combinatorial maps of the underlying cube complexes which preserve the Γ-labelling.

Compare completion sequences with the topological approach to Nielsen equivalence discussed in Section 1.2.2. When studying Coxeter groups, it makes sense to let $\mathcal{O}$ be a *complex of groups* whose complex of groups fundamental group is the corresponding Coxeter group. The roles of cellular maps and the notion of local injectivity have to be modified slightly to work in this setting. The connection to completion sequences should be clear, except instead of trying to build a folded and cube full complex, we need to stay in the category of cube complexes with free fundamental group. Notice that because the standard generators of a Coxeter group are involutions, when using the pulled back the labelling of $\mathcal{O}$, we do not need to fix and pull back an orientation on the edges, whence the definition of Γ-complexes above.

5.2   Generating tuples of quasiconvex subgroups of RACGs

As when looking at reflection generating tuples, it is useful to have a way to tell whether a finite tuple of elements of W_Γ generates a given quasiconvex subgroup G. Fortunately completion sequences give us an easy way to do this.

THEOREM 5.14: *(Follows from Theorem 13.1 in [31]) Let (W, S) be a RACG, and G a quasiconvex subgroup given by a finite tuple of generators Y. Then given a finite tuple of elements X from W, there is an algorithm which terminates if $\langle X \rangle$ is quasiconvex and in this case outputs whether X generates G.*

Proof. Consider the complexes (Ω_Y, v_0) and (Ω_X, v_1), the based Γ-labelled graphs representing Y and X respectively which are defined on page 168. Let $(\widehat{\Omega}_Y, v_0)$ and $(\widehat{\Omega}_X, v_1)$ be standard completions of Ω_Y and Ω_X (where the images of v_0 and v_1 in the completions are again denoted by v_0 and v_1 respectively). Since G is quasiconvex, by Proposition 5.5 and Theorem 5.12, $(\widehat{\Omega}_Y, v_0)$ can be computed in finitely many steps. If X generates a quasiconvex subgroup, then $\widehat{\Omega}_X$ is also computable in finitely many steps.

Assume that $\langle X \rangle$ is quasiconvex. One can solve the membership problem in G or $\langle X \rangle$ using their completions, see Theorem 13.1 in [31]. In particular, suppose w is a word over S and we wish to test whether w represents an element of G. After applying Theorem 1.11 or Corollary 1.12 we can assume that w is a reduced word. If w represents an element of G, then Proposition 5.9 implies there is a based loop in $C(\widehat{\Omega}_Y, v_0) \subset (\widehat{\Omega}_Y, v_0)$ labelled by w. Conversely, if w represents an element which does not lie in G, then no based loop in $(\widehat{\Omega}_Y, v_0)$ is labelled by w. This is because no based loop in (Ω_Y, v_0) is labelled by a word representing the same element as w, and the subgroup associated to a Γ-complex is preserved under taking completion sequences, see Definition 5.8.

To check whether $\langle X \rangle = G$ it suffices to check the membership problem for the elements of Y in $\langle X \rangle$, and for the elements of X in G.   ◻

Remark 5.15 (Decidability). If X generates a quasiconvex subgroup then this algorithm terminates and gives a negative answer, however if this is not the case, then it never terminates. One might hope that there is some other procedure which takes X as input and terminates if $\langle X \rangle$ is *not* quasiconvex. However, it seems very unlikely that this problem is decidable—for example it is undecidable for *hyperbolic group*, see [15].

In Section 5.5.3 we briefly discuss an implementation of this algorithm in Math-

ematica. This has allowed us to compute many examples of generating tuples of RACGs and use these to explore Nielsen equivalence.

5.3   Free completion sequences

We can also use completion sequences to study Nielsen equivalence directly as long as we impose the restriction that any folds result in a complex with free fundamental group (cube identifications and attachments automatically preserve the property of having a free fundamental group). Such a *free* completion sequence exactly models the diagram shown in (1.2) on page 62, where the maps g_i to $\mathcal{O}$ are encoded by the Γ-labelling on the complexes in the sequence.

5.3.1   Free folds and attachments

Definition 5.16. A Γ-complex Ω is **free** if $\pi_1(\Omega)$ is free. Call a sequence of free Γ-labelled complexes $(\Omega_i)_i$ where consecutive complexes differ by a fold, cube identification, or cube attachment a **free completion sequence**.

In a free completion sequence, cube identifications can always be performed without changing the fundamental group, but we need slightly modified versions of edge folds and cube attachments.

Free edge fold   A free edge fold of a free Γ-complex is any fold which results in a free Γ-complex.

Free cube attachment   A free cube attachment to a free Γ-complex is a cube attachment $\Omega \to \Omega'$ such that there is no sequence of folds which can be applied to Ω', after which the newly attached cube shares a vertex with another cube of the same type (see Definition 5.2).

Definition 5.17. Let Ω be a free Γ-complex. It is **freely folded** if no cube identifications or free edge folds are possible; it is **freely cube full** if there are no possible free cube attachments; and it is a **free completion** if it is freely folded and freely cube full.

5.3.2   Characterising free cube attachments

Checking whether a cube attachment to a cube complex is a *free* cube attachment appears difficult in general as one must consider all possible sequences of folds. It turns out however that it is possible to get away with checking just a single sequence of folds.

PROPOSITION 5.18: *Let Ω be a free cube complex, and $\Omega \to \widetilde{\Omega}$ a cube attachment. Suppose $\widetilde{\Omega}^1$ is a complex obtained from $\widetilde{\Omega}$ by some sequence of (not necessarily free) folds such that it is not possible to fold $\widetilde{\Omega}^1$ further. If the cube newly attached in $\widetilde{\Omega}$ does not share a vertex with any other cubes in $\widetilde{\Omega}^1$ of the same type, then $\Omega \to \widetilde{\Omega}$ is a free cube attachment.*

The following Lemma is useful in proving the Proposition, and can be thought of a generalisation of the comment at the end of Section 3.3 in [105].

LEMMA 5.19: *Let Ω be a finite Γ-labelled graph, and let*

$$\Omega = \Omega_0^1 \xrightarrow{f_0^1} \Omega_1^1 \xrightarrow{f_1^1} \cdots \xrightarrow{f_{l_1-1}^1} \Omega_{l_1}^1 = \Omega^1$$

$$\Omega = \Omega_0^2 \xrightarrow{f_0^2} \Omega_1^2 \xrightarrow{f_1^2} \cdots \xrightarrow{f_{l_2-1}^2} \Omega_{l_2}^2 = \Omega^2$$

be two sequences of folds such that Ω^1 and Ω^2 cannot be folded further. Write $f^1 \colon \Omega \to \Omega^1$ and $f^2 \colon \Omega \to \Omega^2$ for the corresponding maps. Then there is an isomorphism $g \colon \Omega^1 \to \Omega^2$ such that the following diagrams commute

$$
\begin{array}{ccc}
 & V\Omega & \\
{\scriptstyle f_V^1}\swarrow & & \searrow{\scriptstyle f_V^2} \\
V\Omega^1 & \xrightarrow{\;g_V\;} & V\Omega^2
\end{array}
\qquad\qquad
\begin{array}{ccc}
 & E\Omega & \\
{\scriptstyle f_E^1}\swarrow & & \searrow{\scriptstyle f_E^2} \\
E\Omega^1 & \xrightarrow{\;g_E\;} & E\Omega^2
\end{array}
$$

where f_V and f_E denote the maps induced on the sets of vertices and edges respectively by some map of graphs f.

Proof. Define an equivalence relation on $V\Omega$ by saying $v_1 \sim v_2$ if $f_V^1(v_1) = f_V^1(v_2)$, and similarly define an equivalence relation on $E\Omega$ by saying $e_1 \sim e_2$ if $f_E^1(e_1) = f_E^1(e_2)$. Fix a choice of equivalence class representatives for the vertices and edges

of Ω. For a vertex $\bar{v} \in V\Omega^1$ let v be its representative in $V\Omega$ and define $g \colon \Omega^1 \to \Omega^2$ on v to be $g(\bar{v}) = f^2(v)$. For an edge $\bar{e} \in E\Omega^1$, let e be the representative of the class of edges which are mapped to $\bar{e}$ by f_E^1. For a point $\bar{x}$ in the interior of $\bar{e}$, let $g(\bar{x}) = f^2(x)$, where x is the point in the preimage of $\bar{x}$ under the map f^1 which lies in e. To show that g is well-defined and satisfies the conclusion of the Lemma, we prove the following claim.

Claim: a pair of edges or vertices in Ω are identified in Ω^1 if and only if they are identified in Ω^2. To simplify notation, we refer to the images of any edge e or vertex v from Ω in any graph obtained by a sequence of folds by e or v respectively as well. We prove the claim by induction on the number of terms in the folding sequence which produces Ω^1 before it becomes possible to identify a given pair of edges or vertices.

Suppose that in Ω it is possible to fold two distinct edges e_1 and e_2, ie these edges carry the same label and share an endpoint. Then if f_0^2 does not fold these edges together, they can still be folded in Ω_1^2. This is because the fold f_0^2 does not change the edge labels, and cannot separate edges which already meet. It follows by induction that for each $j \geqslant 0$, in Ω_j^2 either e_1 and e_2 have already been identified, or it is still possible to fold them together. In particular, since no folds are possible in Ω^2, $f_E^2(e_1) = f_E^2(e_2)$. It follows that if v_1 and v_2 are two distinct vertices in Ω which can be identified by a single fold, then either for each $j \geqslant 0$, in Ω_j^2 either v_1 and v_2 have already been identified, or it is still possible to identify them by a single fold.

Now consider two vertices $v_1 \neq v_2 \in V\Omega$ which are identified by f^1, and let j be minimal such that it is possible to identify v_1 and v_2 in Ω_j^1 by a single fold. Then there are edges e_1 and e_2 which have v_1 and v_2 as one of their endpoints respectively, which can be folded in Ω_j^1, but are disjoint in Ω_{j-1}^1. Therefore, f_{j-1}^1 folds a pair of edges e_1' and e_2', this identifies a pair of distinct vertices v_1' and v_2' which are endpoints of e_1 and e_2. An example of such a fold is shown in Figure 5.5.

By induction on j we can assume that there is some k such that each pair of distinct edges and vertices in Ω which are identified in Ω_j^1 have already been iden-
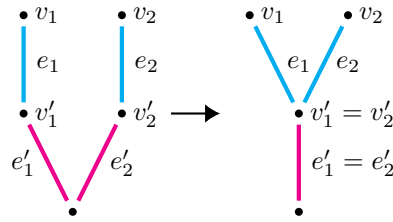
Figure 5.5: An example of the fold f^1_{j-1}.

tified in Ω^2_k. In particular v'_i either started off as, or has already been identified with, one of the endpoints of e_i and e'_i for $i = 1, 2$; v_i either started off as, or has already been identified with, one of the endpoints of e_i for $i = 1, 2$; and v'_1 has been identified with v'_2. Thus, in Ω^2_k, either e_1 and e_2 have already been folded together, or it is possible to fold them in Ω^2_k. Either way, they must be identified in Ω^2. It follows from this, and the fact that the only ambiguity in how e_1 and e_2 are folded arises when one of them is an edge loop, that v_1 and v_2 must either have already been identified, or it is possible to identify them in Ω^2_k by a single fold; and so they must be identified in Ω^2.

We have shown that any pair of edges or vertices which are identified by f^1 must also be identified by f^2. Repeating the same argument with the roles of Ω^1 and Ω^2 reversed we see that a pair of edges or vertices in Ω are identified in Ω^1 if and only if they are identified in Ω^2, completing the proof of the claim.

To see that g is well-defined, fix an orientation on the edges of Ω^1, which we can then pull back to an orientation on the edges of Ω by f^1. Let $\bar{e}$ be an edge of Ω^1 and let $\bar{v} = \iota\bar{e}$ be its initial vertex. Then the the initial vertex of e, v', is mapped to $\bar{v}$ by f^1 (since folds are simplicial maps), ie $v' \sim v$. Then

$$g(\iota\bar{e}) = g(\bar{v}) := f^2(v) = f^2(v') = \iota f^2(e) =: \iota g(\bar{e}),$$

where the central equality follows from the proof above. Reversing the orientation on all of the edges of Ω^1 and applying the same argument shows that g is a well-defined simplicial map. Swapping Ω^1 and Ω^2 in the definition of g yields a simplicial map $\Omega^2 \to \Omega^1$ which is the inverse of g, since the equivalence classes of edges and vertices in Ω coming from f^2 are the same as those coming from f^1. Therefore g is a graph isomorphism. That the two diagrams in the statement of the Lemma commute follows directly from the definition of g.

In general it is not true that the map g itself commutes with f^1 and f^2 because these maps may fold an edge along an edge loop at some stage in different orientations.

Proof of Proposition 5.18. We show that if $\Omega \to \widetilde{\Omega}$ is not a free cube attachment, then this is witnessed by any choice of $\widetilde{\Omega}^1$. Let $\widetilde{\Omega} \to \widetilde{\Omega}_1 \to \cdots \to \widetilde{\Omega}_k$ be a sequence of folds such that the cube c attached in $\Omega \to \widetilde{\Omega}$ shares a vertex with another cube c' of the same type. Continue this sequence of folds $\widetilde{\Omega}_k \to \cdots \to \widetilde{\Omega}_l =: \widetilde{\Omega}^2$ until no more folds are possible.

In general $\widetilde{\Omega}^1$ and $\widetilde{\Omega}^2$ are different complexes, however by Lemma 5.19 there is an isomorphism g between their 1-skeleta which commutes with the maps induced on their sets of vertices and edges by $f^1 \colon \; \colon \widetilde{\Omega} \to \widetilde{\Omega}^1$ and $f^2 \colon \widetilde{\Omega} \to \widetilde{\Omega}^2$. Let v be a vertex of c and v' a vertex of c' in $\widetilde{\Omega}$ which get identified in $\widetilde{\Omega}_k$, then these vertices get identified by f^2, and so by f^1 as well. In other words, c and c' share a vertex in $\widetilde{\Omega}^1$ showing that $\Omega \to \widetilde{\Omega}$ is not a free cube attachment. $\qquad\square$

5.3.3   Free cube attachments in freely folded complexes

Lemma 5.20: *Let Ω be a freely folded cube complex, and let $\overline{\Omega}$ be a folded cube complex obtained by performing all possible folds and cube identifications. Then the map $\Omega \to \overline{\Omega}$ is the identity map when restricted to the vertex set of Ω.*

Proof. Cube identifications do not change the vertex set so the only way this Lemma could fail is if some fold identifies two vertices. There are essentially four types of fold shown in Figure 5.3. Types (a) and (b) change the vertex set, but are also homotopy equivalences, and so are always free folds when performed on free complexes. On the other hand (c) and (d) leave the vertex set unchanged, but may fail to be free folds.

We claim that no folds of type (a) or (b) are applied in $\Omega \to \overline{\Omega}$. Indeed, since Ω is freely folded, no (a) or (b) folds are possible to begin with. Writing out the sequence of folds and identifications explicitly

$$\Omega = \Omega_0 \to \Omega_1 \to \cdots \to \overline{\Omega},$$

if such a fold becomes possible subsequently, let Ω_i be the first complex where this happens. Then the map $\Omega_{i-1} \to \Omega_i$ must change the 1-skeleton of Ω_{i-1} which rules out it being a cube identification. In fact the two edges involved in the type (a) or (b) fold must have a common endpoint in Ω_i, but since the fold is not possible in Ω_{i-1} they cannot have a common endpoint before this point. In other words $\Omega_{i-1} \to \Omega_i$ changes the vertex set of Ω_{i-1}, and hence must be a fold of type (a) or (b) itself. But this contradicts the minimality of i. $\hfill\square$

It follows immediately from this Lemma that in the case of freely folded complex, detecting whether a cube attachment is free is even simpler than Proposition 5.18.

Proposition 5.21: *Let Ω be a freely folded cube complex, and $\Omega \to \widetilde{\Omega}$ a cube attachment at some vertex v. This cube attachment is free if and only if there is not another cube in Ω with the same type as the attached cube which has v as a vertex.*

5.4 Standard free completion sequences

As with normal completion sequences, we can define a *standard* free completion sequence to be a free completion sequence where, roughly speaking, we alternately perform all possible cube identifications and free edge folds, followed by all possible free cube attachments. In this Section we rigorously define these sequences and characterise when they are finite.

Standard free completion sequence: Suppose that some finite portion of a standard free completion sequence of Ω has already been constructed

$$\Omega = \Omega_0 \to \Omega_1 \to \cdots \to \Omega_i,$$

it follows by induction that Ω_i is a finite complex. Suppose Ω_i is not freely folded. Because it is finite, it contains finitely many edges and cubes. Let j the the finite number of free folds and cube identifications possible. Performing each of these in turn, the standard free completion sequence extends

$$\Omega_i \to \Omega_{i+1} \to \cdots \to \Omega_{i+j},$$

and Ω_{i+j} is still a finite cube complex.

On the other hand, if Ω_i is a freely folded complex, then for each vertex v in Ω_i perform all possible free cube attachments at v. Since Ω_i is finite this involves adding only a finite number of cubes j, so the standard free completion sequence extends

$$\Omega_i \to \Omega_{i+1} \to \cdots \to \Omega_{i+j}.$$

Remark 5.22 (Uniqueness of standard free completion sequences). As with standard completion sequences, a standard free completion sequence is not unique. Here we are only concerned with *finite* free completion sequences, so there is no need to work with direct limits of complexes.

5.4.1 Finite standard free completion sequences

The somewhat convoluted definition of *freely cube full* is required because otherwise standard free completion sequences are always infinite except in a few trivial cases. With this definition we want to prove the following.

THEOREM 5.23: *Let Ω be a finite connected free Γ-complex. Then any standard free completion sequence of Ω is finite if and only if Ω has a finite completion sequence.*

Recall that by Theorem 5.12, Ω has a finite completion sequence if and only its associated subgroup is quasiconvex in W_Γ.

We prove the two directions of this Theorem separately, but the method of proof follows the same idea in each case. For the only if direction we prove the contrapositive, so assume that every completion sequence of Ω is infinite. Taking an infinite standard completion sequence we construct a standard free completion sequence in parallel by performing the same sequence of folds, identifications, and attachments, but skipping those folds and attachments which are not free. Since the standard completion sequence is infinite, the standard free completion sequence we build in this way is also infinite.

To prove the if direction, we start with the assumption that Ω has a finite completion, which implies that every standard completion is finite. Then we take a standard free completion sequence and construct a standard completion sequence

in parallel by performing the same sequence of free folds, identifications, and free cube attachments, and periodically adding in any extra non-free folds which can be performed. Lemma 5.20 guarantees that these extra folds do not make the two parallel completion sequences differ too much from each other, and its consequent, Proposition 5.21, ensures that all possible cube attachments in the completion sequence actually arise as a result of free cube attachments in the free completion sequence. Since the standard completion sequence we construct in this way must by finite by assumption, the standard free completion sequence we started with (which was arbitrary) must also have been finite.

PROPOSITION 5.24: *Let Ω be a finite connected free Γ-complex. If every completion sequence of Ω is infinite then Ω has an infinite standard free completion sequence.*

Proof. Let

$$\Omega = \Omega_0 \xrightarrow{f_0} \Omega_1 \xrightarrow{f_1} \Omega_2 \xrightarrow{f_2} \cdots \tag{5.1}$$

be an infinite standard completion sequence. We construct a new sequence based on this

$$\Omega = \Omega'_0 \xrightarrow{f'_0} \Omega'_1 \xrightarrow{f'_1} \Omega'_2 \xrightarrow{f'_2} \cdots \tag{5.2}$$

where each map is a (composition of) free fold(s) and/or identifications; a free cube attachment; or the identity. Roughly speaking, $\Omega'_i \to \Omega'_{i+1}$ agrees with $\Omega_i \to \Omega_{i+1}$ in the sense that it folds the 'same' pair of edges, identifies the 'same' pair of cubes, or attaches the same type of cube at the 'same' vertex so long as $\Omega_i \to \Omega_{i+1}$ is a free fold, a free cube attachment, or a cube identification; and otherwise $\Omega'_i \to \Omega'_{i+1}$ is the identity.

It is not clear that this sequence is well-defined, and in fact it is not well-defined in general using the naïve construction outlined above (in particular, cube identifications in (5.1) are not necessarily be possible in (5.2)). Nevertheless, a rigorous construction based on this idea yields a sequence (5.2) from with we can straightforwardly derive an infinite standard free completion of Ω, as required.

Suppose that we have already constructed

$$\Omega = \Omega'_0 \xrightarrow{f'_0} \Omega'_1 \xrightarrow{f'_1} \cdots \xrightarrow{f'_{k-1}} \Omega'_k$$

for some $k \geqslant 0$, together with maps $g_i : \Omega'_i \to \Omega_i$ for each $0 \leqslant i \leqslant k$ such that g_0 is the identity, and the following diagram commutes:

$$
\begin{array}{ccccccccccc}
\Omega'_0 & \xrightarrow{f'_0} & \Omega'_1 & \xrightarrow{f'_1} & \cdots & \xrightarrow{f'_{k-1}} & \Omega'_k & & & & \\
\downarrow{\scriptstyle g_0} & & \downarrow{\scriptstyle g_1} & & & & \downarrow{\scriptstyle g_k} & & & & \\
\Omega_0 & \xrightarrow{f_0} & \Omega_1 & \xrightarrow{f_1} & \cdots & \xrightarrow{f_{k-1}} & \Omega_k & \xrightarrow{f_k} & \Omega_{k+1} & \xrightarrow{f_{k+1}} & \cdots
\end{array}
$$

We see by induction that g_i is surjective, and g_{iV} is a bijection for each $0 \leqslant i \leqslant k$, where (borrowing notation from Lemma 5.19) g_{iV} is the restriction of g_i to the vertex set of Ω'_i. We now define $\Omega'_k \xrightarrow{f'_k} \Omega'_{k+1} \xrightarrow{g_{k+1}} \Omega_{k+1}$ depending on the type of operation performed by f_k.

f_k **folds edges** e_1 **and** e_2 Let v be a vertex common to e_1 and e_2 in Ω_k, and let $v' = g_{kV}^{-1}(v)$. Since g_k is surjective, there must be non-empty sets edges $g_{kE}^{-1}(e_1) = \{e'_{11}, \ldots, e'_{1m_1}\}$ and $g_{kE}^{-1}(e_2) = \{e'_{21}, \ldots, e'_{2m_2}\}$ in Ω'_k, all of which have the same label as e_1 and e_2, and which are incident to v'. Hence all the e'_{ij}'s may be folded in Ω'_k.

If it is not possible to fold any pair of edges in $g_{kE}^{-1}(e_1) \cup g_{kE}^{-1}(e_2)$ freely then none of these folds change the vertex set of Ω'_k. Let $\Omega'_{k+1} = \Omega'_k$ and f'_k be the identity map, then we can take $g_{k+1} = f_k \circ g_k$. It follows that g_{k+1} is a surjection which leaves the vertex set unchanged, so g_{k+1} inherits these properties.

Otherwise, fix an orientation on the edge $e := f_k(e_1) = f_k(e_2)$ and pull this back under $f_k \circ g_k$ to an orientation on each edge in $g_{kE}^{-1}(e_1) \cup g_{kE}^{-1}(e_2)$. Let $\Omega'_k \xrightarrow{f'_k} \Omega'_{k+1}$ be a maximal composition of free folds of pairs of edges in $g_{kE}^{-1}(e_1) \cup g_{kE}^{-1}(e_2)$, (if there is ambiguity about how to perform a given fold because the fold is of type (b) or (d) in Figure 5.3, then the fold should be performed to match the orientations of the edges involved).

Let g_{k+1} be $f_k \circ g_k$ away from the images of the edges in $g_{kE}^{-1}(e_1) \cup g_{kE}^{-1}(e_2)$, since f'_{k+1} does not change $\Omega'_k - (g_{kE}^{-1}(e_1) \cup g_{kE}^{-1}(e_2))$. On the edges in $g_{kE}^{-1}(e_1) \cup g_{kE}^{-1}(e_2)$, let g_{k+1} map them all onto the edge $e \in \Omega_{k+1}$ in such a way that the orientations are preserved. This is surjective, and leaves the vertex set unchanged.

f_k **identifies cubes** c_1 **and** c_2 Let d be the dimension of the cubes c_1 and c_2, which have the same attaching maps in Ω_k, and so, in particular, share the same set of vertices. Consider the set C of all d-cubes in Ω'_k which map onto c_1 or c_2 (as above for folds, there is at least one cube mapping onto each). Exhaustively identify all pairs of facets of cubes in C which can be identified (in particular these facets should have dimension $\geqslant 2$), and call f'_k the composition of all of these identifications. If no identifications are possible then f'_k is the identity.

Define g_{k+1} to agree with g_k on all cubes which do not lie in the image of C in Ω'_{k+1}. All the cubes in C get mapped to a single cube c in Ω_{k+1}, so define g_{k+1} to map all cubes in the image of C onto c. Then g_{k+1} is surjective and $g_{k+1} \circ f'_k = f_k \circ g_k$. Cube identifications do not change the vertex set of the complexes, and g_{k+1} agrees with g_k on the vertex set.

Before considering the final case, we define a subsequence of (5.1) which captures the coarse, alternating structure of the standard completion sequence. Define the subsequence

$$\Omega_0 \to \Omega_{p_1} \to \Omega_{q_1} \to \Omega_{p_2} \to \cdots \to \Omega_{p_j} \to \Omega_{q_j} \to \cdots \qquad (5.3)$$

where Ω_{p_j} is the first term in $\Omega_{q_{j-1}} \to \Omega_{q_{j-1}+1} \to \cdots$ (or in (5.1) if $j = 1$) which is folded; while Ω_{q_j} is the first term, after Ω_{p_j} such that every cube attachment which is possible in Ω_{p_j} has been performed. It is possible that $p_j = q_{j-1}$, which happens if after adding cubes no new folds are possible; in this case $\Omega_{q_{j-1}} \to \Omega_{p_j}$ is the identity.

f_k **attaches a cube** c **at** v Note that by the definition of a standard completion, which cubes are attached during a particular cycle of cube attachments is determined by which attachments are possible in a *folded* complex, so a cube attachment is completely determined by the vertex of attachment, and the type (ie label set) of cube which is being attached.

If $k = p_j$ for some j then we first modify $\Omega'_{k-1} \xrightarrow{f'_{k-1}} \Omega'_k \xrightarrow{g_k} \Omega_k$, which has already been defined, before we define $\Omega'_k \xrightarrow{f'_k} \Omega'_{k+1} \xrightarrow{g_{k+1}} \Omega_{k+1}$. If Ω'_k is freely

folded then we do not need to do anything. Otherwise there are some free folds and/or cube identifications possible in Ω'_k. For possible free folds, the edges must be folded by g_k since Ω_k is folded, and moreover they must be folds of type (c) or (d) in Figure 5.3, because $g_k V$ is a bijection. Update f_{k-1} by composing it with these free folds and cube identifications. Since g_k factors through this sequence it can also be updated appropriately. It is still surjective and induces a bijection on the set of vertices.

Now moving on to f_k, it is also possible to attach a cube c' of the same type to Ω'_k at the vertex $v' = g_{kV}^{-1}(v)$. This attachment must also have been possible in Ω'_{p_j}, but as this complex is not necessarily folded, there may be several edges which meet v' which have the same label. Let f'_{k+1} be the composition of all possible attachments of c' at v'. Because c may be attached to Ω_{p_j} which is folded, there is no other cube at v in Ω_k of the same type. As g_k does not identify any other vertices with v', it follows that there is no cube with the same labels as c' which meets v', so these cube attachments are each free.

The cube attachment does not change the rest of Ω'_k, so we can define g_{k+1} to equal $f_k \circ g_k$ on the image of Ω'_k in Ω'_{k+1}, and to send c' to c. This is surjective, and $g_{k+1 V}$ is a bijection.

By induction we can produce the infinite sequence (5.2), we want to extract an infinite standard free completion sequence from this. Define

$$\Omega'_0 \to \Omega'_{p_1} \to \Omega'_{q_1} \to \Omega'_{p_2} \to \cdots \to \Omega'_{p_j} \to \Omega'_{q_j} \to \cdots \tag{5.4}$$

to be the subsequence parallel to (5.3) in (5.2). We claim that this sequence has the following properties:

1. Each $\Omega'_{p_j} \to \Omega'_{q_j}$ is a composition of a non-empty finite sequence of free cube attachments,

2. Each $\Omega'_{q_j} \to \Omega'_{p_{j+1}}$ (and $\Omega'_0 \to \Omega'_{p_1}$) is a composition of a (possibly empty) finite sequence of free folds and cube identifications,

3. Every free cube attachment which can be performed on Ω'_{p_j} has been applied

to produce, and Ω'_{q_j}

4. Each Ω'_{p_j} is freely folded.

By the definition of a standard completion sequence, properties (1) and (2) hold for (5.3). When (5.2) was defined, each map in (5.3) was replaced with either a map of the same type, or the identity. Since each Ω_{p_j} is folded, any cube attachment which is possible is in fact a free cube attachment. Because g_{p_j} is a Γ-labelling preserving map, every cube attachment possible in Ω_{p_j} has a corresponding possible cube attachment in Ω'_{p_j} which is free. Thus every cube attachment in $\Omega_{p_j} \to \Omega_{q_j}$ yields a corresponding cube attachment in $\Omega'_{p_j} \to \Omega'_{q_j}$. Since $\Omega_{p_j} \to \Omega_{q_j}$ must involve at least one cube attachment (otherwise Ω_{p_j} is folded and cube full, contradicting the assumption that (5.1) is an infinite completion sequence), $\Omega'_{p_j} \to \Omega'_{q_j}$ is a non-empty composition of cube attachments. Thus properties (1) and (2) hold.

Every cube attachment which is possible in Ω'_{p_j} projects to a possible cube attachment in Ω_{p_j} since $g_{p_j V}$ is a bijection. When we defined (5.2) we performed all possible free cube attachments which project to a cube attachment in Ω_{p_j}, so property (3) holds.

Finally, property (4) is guaranteed by the modification we sometimes make at the start of a round of free cube attachments to the free folding sequence in the definition of (5.2).

The final step in proving the Proposition is to expand (5.4) out into a standard free completion of $\Omega = \Omega'_0$. To do this, replace each map $\Omega'_{p_j} \to \Omega'_{q_j}$ with a sequence of single free cube attachments, and either delete $\Omega'_{q_j} \to \Omega'_{p_{j+1}}$ if it is the identity, or else replace it with a sequence of single free folds and identifications. By properties (3) and (4) this is a standard free completion sequence since alternately all possible free folds and identifications, and then all possible free cube attachments which are possible in a freely folded complex are made. Since each of the infinitely many maps $\Omega'_{p_j} \to \Omega'_{q_j}$ are replaced with at least one cube attachment, this standard free completion sequence is infinite in length.   ⬛

PROPOSITION 5.25: *Let Ω be a finite connected free Γ-complex. If Ω has a finite completion sequence then every standard free completion sequence of Ω is finite.*

Proof. Assume that Ω has a finite completion sequence. Then by Theorem 5.12 every standard completion sequence of Ω is finite. We start with a standard free completion sequence of Ω

$$\Omega = \Omega'_0 \xrightarrow{f'_0} \Omega'_1 \xrightarrow{f'_1} \Omega'_2 \xrightarrow{f'_2} \cdots \qquad (5.5)$$

and we construct a new sequence based on this

$$\Omega = \Omega_0 \xrightarrow{f_0} \Omega_1 \xrightarrow{f_1} \Omega_2 \xrightarrow{f_2} \cdots \qquad (5.6)$$

where each map is a (composition of) fold(s) and/or identifications, a cube attachment, or the identity. This follows a complementary process to the proof of Proposition 5.24, and as such, we also produce a sequence of functions $g_k : \Omega'_k \to \Omega_k$ which commute with the f_k's and f''_k's, are surjective, and such that the induced map g_{kV} is a bijection. Define g_0 to be the identity, and then by induction assume that we have constructed

$$
\begin{array}{ccccccccc}
\Omega'_0 & \xrightarrow{f'_0} & \Omega'_1 & \xrightarrow{f'_1} & \cdots & \xrightarrow{f'_{k-1}} & \Omega'_k & \xrightarrow{f'_k} & \Omega'_{k+1} & \xrightarrow{f'_{k+1}} & \cdots \\
{\scriptstyle g_0}\downarrow & & {\scriptstyle g_1}\downarrow & & & & {\scriptstyle g_k}\downarrow & & & & \\
\Omega_0 & \xrightarrow{f_0} & \Omega_1 & \xrightarrow{f_1} & \cdots & \xrightarrow{f_{k-1}} & \Omega_k & & & &
\end{array}
$$

and we define $\Omega_k \xrightarrow{f_k} \Omega_{k+1} \xleftarrow{g_{k+1}} \Omega'_{k+1}$ depending on the map f'_k.

f'_k is a freely folds edges e'_1 and e'_2 If g_k folds these edges then let f_k be the identity, and g_k must factor through f'_k so we can find a map $g_{k+1} : \Omega'_{k+1} \to \Omega_{k+1} = \Omega_k$ making the diagram commute. Since g_{kV} is a bijection, the fold does not identify any distinct vertices and so g_{k+1} is a surjection which induces a bijection on the vertex set.

Otherwise $e_1 = g_k(e'_1)$ and $e_2 = g_k(e'_2)$ are distinct edges in Ω_k, but they have the same label, and share an endpoint, and so can be folded. Choose orientations on e'_1 and e'_2 which match up with how they are folded, then these orientations induce orientations on e_1 and e_2 via g_k. Let f_k fold these edges so that the orientations

are matched up. Let g_{k+1} agree with g_k away from $e' = f'_k(e'_1) = f'_k(e'_2)$, and map $e' \mapsto f_k(e_1)$. This map commutes and has the required properties.

f'_k identifies cubes c'_1 and c'_2 If g_k identifies these cube then let f_k be the identity, and g_k must factor through f'_k so we can find a map $g_{k+1} : \Omega'_{k+1} \to \Omega_{k+1} = \Omega_k$ making the diagram commute. Since cube identifications do not affect the vertex set, g_{k+1} is a surjection which induces a bijection on the vertex set.

Otherwise $g_k(c'_1)$ and $g_k(c'_2)$ are distinct cubes in Ω_k, but they have the same type and attaching maps up to an isometry; hence, the can be identified. Let f_k identify these cubes. Let g_{k+1} agree with g_k away from $c' = f'_k(c'_1) = f'_k(c'_2)$, and map $c' \mapsto f_k(g_k(c'_1))$. This map commutes and has the required properties.

Before considering the final case, we define a subsequence of (5.5) in the same spirit as (5.3). Define the subsequence

$$\Omega'_0 \to \Omega'_{p_1} \to \Omega'_{q_1} \to \Omega'_{p_2} \to \cdots \to \Omega'_{p_j} \to \Omega'_{q_j} \to \cdots \tag{5.7}$$

where Ω'_{p_j} is the first term in $\Omega'_{q_{j-1}} \to \Omega'_{q_{j-1}+1} \to \cdots$ (or in (5.5) if $j = 1$) which is freely folded; while Ω'_{q_j} is the first term after Ω'_{p_j} such that every free cube attachment which is possible in Ω'_{p_j} has been performed. It may that $p_j = q_{j-1}$, which happens if after freely adding cubes no new free folds are possible; in this case $\Omega'_{q_{j-1}} \to \Omega'_{p_j}$ is the identity.

f'_k freely attaches a cube c' at v' to edges set $e'_1, \ldots, e'_d$ If $k = p_j$ for some j, then our first job is to possibly modify $\Omega_{k-1} \xrightarrow{f_{k-1}} \Omega_k \xleftarrow{g_k} \Omega'_k$. If Ω_k is folded then nothing needs to be done. Otherwise Ω_k is freely folded, but not folded. Replace f_{k-1} with $f \circ f_{k-1}$ where f performs all possible folds and cube identifications, and replace g_k with $f \circ g_k$. Now Ω_k is folded, g_k is still surjective, and by Lemma 5.20 g_{kV} is a bijection.

Now we define $\Omega_k \xrightarrow{f_k} \Omega_{k+1} \xleftarrow{g_{k+1}} \Omega'_{k+1}$. Since Ω_k is folded and g_k is surjective, there exists a unique edge e_i incident to $v = g_k(v')$ which carries the same label as e'_i for $1 \leqslant i \leqslant d$. Since it is possible to freely attach c' at v', it follows from

Proposition 5.21 that Ω'_k contains no cubes with the same label as c' which have v' as a vertex. Because g_k is surjective, Ω_k also contains no cube of the same type as c' with v as a vertex. Thus, it is possible to attach a cube c of the same type as c' to Ω'_k at v.

Consider the orientations on the edges $\{e'_i\}$ used to attach c', these induce orientations on the edges $\{e_i\}$ via g_k. Let f_k be the cube attachment of c to v according to these orientations on $\{e_i\}$. Let g_{k+1} agree with g_k away from c', and map c' to c; this gives g_{k+1} the required properties.

Now that we have constructed the sequence (5.6), define the subsequence

$$\Omega_0 \to \Omega_{p_1} \to \Omega_{q_1} \to \Omega_{p_2} \to \cdots \to \Omega_{p_j} \to \Omega_{q_j} \to \cdots \tag{5.8}$$

which runs parallel to (5.7). We show that this sequence enjoys the following properties:

1. Each $\Omega_{p_j} \to \Omega_{q_j}$ is a composition of a non-empty finite sequence of cube attachments,

2. Each $\Omega_{q_j} \to \Omega_{p_{j+1}}$ (and $\Omega_0 \to \Omega_{p_1}$) is a composition of a (possibly empty) finite sequence of folds and cube identifications,

3. Each Ω_{p_j} is folded, and

4. Every cube attachment which can be performed on Ω_{p_j} has been applied to produce Ω_{q_j}.

Properties (1) and (2) are inherited directly from (5.7). The reason $\Omega_{p_j} \to \Omega_{q_j}$ is not the identity is that if $\Omega'_{p_j} \to \Omega'_{q_j}$ contained no cube attachments, Ω'_{p_j} is freely folded and freely cube full, and so (5.5) terminates at Ω_{p_j}. Property (3) is guaranteed by the modification made to f_{p_j-1} before performing any cube attachments to Ω_{p_j}.

Finally, suppose it is possible to attach a cube c at some vertex v in Ω_{p_j}, then since Ω_{p_j} is folded there is a unique set of edges at which c map be attached. Choose a lift of this set of edges to a set of edges in Ω'_{p_j} which meet $v' = g_{p_j V}^{-1}(v)$. We claim it is possible to freely attach a cube c' of the same type as c to v' to this

set of edges. Indeed, Ω'_{p_j} is freely folded, so by Proposition 5.21 it suffices to show that there is not already a cube of the same type in $\Omega'_{p)j}$ which meets v'. Recalling that the g_k's commute with the f''_k's and f_k's, there are two possibilities. Either such a cube originated in Ω, in which case its image in Ω_{p_j} would have to meet v. Alternatively the map f'_k which attaches this cube is paralleled by a map f_k which attaches a cube of the same type, and whose image in Ω_{p_j} again meets v. In either case, we obtain a contradiction to the hypothesis.

Thus it is possible to freely attach c' to Ω'_{p_j} at v', so by the definition of a standard free completion sequence, there is some $p_j \leqslant m < q_j$ such that f'_m attaches this cube. It follows that f_m attached c at v, which established property (4).

Finally, we want to expand out (5.8) into a completion sequence for Ω. Factorise each $\Omega_{p_j} \to \Omega_{q_j}$ into a finite sequence of cube attachments, and for each $\Omega_{q_j} \to \Omega_{p_{j+1}}$ (and $\Omega_0 \to \Omega_{p_1}$), if it is the identity, delete it, and otherwise factorise it into a finite sequence of folds and identifications. Properties (3) and (4) guarantee that this completion sequence is in fact a standard completion sequence. But recall that the hypothesis that Ω has a finite completion implies that every standard completion sequence is itself finite. Thus (5.8) must terminate, and so the standard free completion sequence (5.5) is finite.   ⌷

Taken together Proposition 5.24 and Proposition 5.25 prove Theorem 5.23. As a Corollary of the proof of Proposition 5.25 we can conclude more about the structure of finite free completions.

COROLLARY 5.26: *Let (Ω, v_0) be a based connected free Γ-complex with finite standard free completion $(\widehat{\Omega}^{free}, v_0)$ (where we denote the image of the basepoint v_0 in $\widehat{\Omega}^{free}$ by v_0 as well). Let $(\widehat{\Omega}, v_0)$ be a completion of Ω, then there is a based embedding of the core graph $C(\widehat{\Omega}, v_0)$ into $(\widehat{\Omega}^{free}, v_0)$.*

Proof. We carry over the notation from the proof of Proposition 5.25. Denote the image of $v_0 \in \Omega$ in any complex obtained as part of a (free) completion sequence by v_0 as well. Let (5.5) be a standard free completion sequence which terminates at $\Omega'_N = \widehat{\Omega}^{free}$, then (5.6) terminates at Ω_N, which is a completion. By Theorem 5.10, there is a based isomorphism between the core graphs $C(\widehat{\Omega}, v_0)$ and $C(\Omega_N, v_0)$, so

it suffices to find a based embedding $C(\Omega_N, v_0) \hookrightarrow (\widehat{\Omega}^{\text{free}}, v_0) = (\Omega'_N, v_0)$.

We showed that g_N is a surjection which induces a bijection on the set of vertices. For each edge in $C(\Omega_N, v_0)$, pick a lift to an edge in Ω'_N. The union of these edges form an embedded copy of $C(\Omega_N, v_0)$ in Ω'_N. ◻

We are principally interested in the case that $\Omega = \Omega_X$ is the wedge of circles associated to some finite generating tuple X of W_Γ. Since this space certainly has a finite completion we can apply Theorem 5.23 and this Corollary to this case.

COROLLARY 5.27: *Let X be a finite generating tuple for W_Γ, then there is a finite free completion $\widehat{\Omega}^{free}_X$ whose 1-skeleton contains a single vertex, and at least one edge labelled by each $s \in V\Gamma$.*

Proof. We can build a finite completion for W_Γ along the lines of the Salvetti complex for the Artin group A_Γ. Explicitly, start with a wedge of circles labelled by the vertices $V\Gamma$ with basepoint v_0. For each edge in Γ, take a 2-torus T^2 cellulated using a single square and attach it to the wedge of circles by identifying the two edges on T^2 with the edges labelled by the endpoints of the edge. Working inductively on k, for each k-complete graph in Γ, take a k-torus T^k cellulated using a single (k)-cube. Attach it by sending the k edges of T^k to the k edges labelled by the vertices of the complete graph, and then identifying higher dimensional faces with the tori already attached.

The result is a folded and cube full Γ-complex, ie a completion of W_Γ. Its core graph is its 1-skeleton, which contains a single vertex, and one edge for each vertex in Γ. Let $(\widehat{\Omega}^{\text{free}}_X, v_0)$ be a free completion, then by Corollary 5.26 there is a based embedding of the core graph so the completion contains at least one edge loop at v_0 labelled by each $s \in V\Gamma$. Suppose the 1-skeleton of $\widehat{\Omega}^{\text{free}}_X$ contains more than one vertex. Since the free completion is connected, there is a spanning tree which contains at least one edge which is not an edge loop. Therefore, we can apply a free fold of type (b) in Figure 5.3, contradicting the assumption that $\widehat{\Omega}^{\text{free}}_X$ is freely folded. ◻

5.5   Nielsen equivalence in RACGs

We want to use free completions to study Nielsen equivalence in of RACGs. We mention the more general case of quasiconvex subgroups below, see Example 5.32. Let $(\widehat{\Omega}^{\text{free}}, v_0)$ be a based free completion of W_Γ—we want to find the generating tuple of W_Γ which it represents. Recall from Section 5.1.1 that a based Γ-labelled cube complex represents a map $g \colon \widehat{\Omega}^{\text{free}} \to \mathcal{O}_\Gamma$ from the underlying cube complex to the complex of groups which represents W_Γ. The map g_* induced on (complex of groups) fundamental groups is a marking $\mathbb{F}_n \to W_\Gamma$. In order to find a generating tuple corresponding to this marking, we first perform a sequence of *combinatorial retractions* to $\widehat{\Omega}^{\text{free}}$.

5.5.1   Combinatorial retractions

Definition 5.28. Let Ω be a cube complex. A **free face** in Ω is a cube c such that:

- there is some maximal (with respect to inclusion) cube c' which contains c as a codimension one face;

- the only cube which contains c as a face is c'; and

- the attaching map of c' to Ω is injective on the interior of c.

Given a free face c in Ω contained in the maximal cube c', the **combinatorial retraction** of Ω at c is the complex obtained by deleting the interior of c (or the whole of c if it is a vertex) and the interior of c' from Ω. See Figure 5.6 for examples.

Lemma 5.29: *Let (Ω, v_0) be a based connected finite free Γ-complex, and let (Ω', v_0) be obtained from (Ω, v_0) by a combinatorial retraction which does not delete the base point v_0. Then (Ω, v_0) and (Ω', v_0) represent the same marking.*

Proof. Thinking of (Ω, v_0) as a topological space, combinatorial retraction is a retraction and so, in particular, it does not change the fundamental group. If the retraction involves deleting the interiors of a $(k+1)$-cube c' and its k-face c for $k \geqslant 2$, then the 1-skeleton of (Ω, v_0) is left unchanged and the claim is trivial.
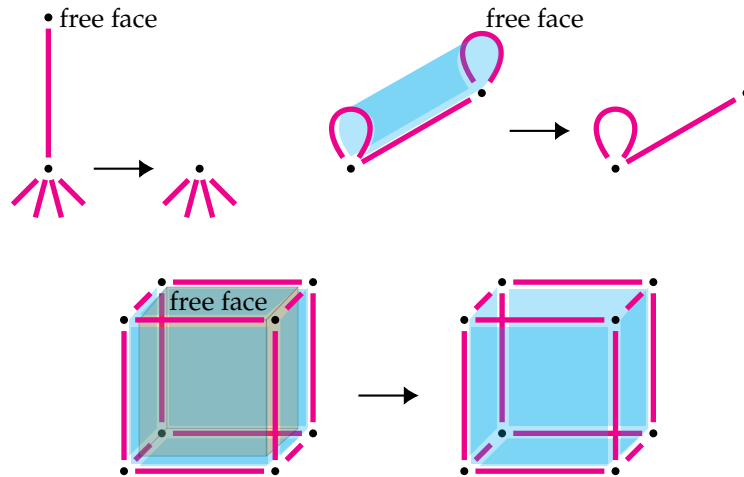
Figure 5.6: Some examples of combinatorial retractions of edges, squares, and cubes.

Suppose $k = 1$, and let ℓ be a based loop in (Ω, v_0) which traverses the edge c from v to v'. Let e_1, e_2, and e_3 be the other three edges of the square c' listed cyclically from v to v'. Then pushing ℓ across c' gives a new loop ℓ' which is homotopic to ℓ, and which traverses $e_1 e_2 e_3$ instead of c. The word labelling ℓ' represents the same element of W_Γ as ℓ, since we have replaced the letter s (which labels c) with the subword tst where t and s commute in W_Γ.

If $k = 0$ then c is a vertex with valence 1 and by assumption $c \neq v_0$. Therefore, any based loop which goes through c must backtrack at that point, and so is homotopic, relative to its endpoint, to a loop which skips c. This combinatorial retraction corresponds to deleting an occurrence of ss in the word labelling the loop, and so does not change the element of W_Γ to which the label corresponds. ⬚

Performing combinatorial retractions which *do* delete v_0 and the edge e containing it, we can update the basepoint to be the other endpoint of e. This has the effect of changing the subgroup of W_Γ associated to Ω by conjugating by the label of e.

Free completions give us a method to test whether a given marking of W_Γ is Nielsen equivalent to (a stabilisation of) a standard marking, ie $\mathbb{F}_{\#V\Gamma} \to W_\Gamma \colon x_i \mapsto v_i$, where $(v_1, \ldots, v_{\#V\Gamma})$ is some ordering of the vertices of Γ. All standard markings of W_Γ are Nielsen equivalent.

THEOREM 5.30: *Let $\phi \colon \mathbb{F}_n \to W_\Gamma$ be a marking of W_Γ such that the generators of $\mathbb{F}_n$ are*

mapped to a tuple X. Let $\widehat{\Omega}_X^{free}$ be a standard free completion of the rose graph Ω_X associated to X. If $\widehat{\Omega}_X^{free}$ retracts onto a graph then ϕ is Nielsen equivalent to (a stabilisation of) a standard marking.

Proof. Notice that if we perform a combinatorial retraction from a free face c which has dimension at least 1, then we do not change the number of hyperplanes in $\widehat{\Omega}_X^{\text{free}}$ (recall Definition 5.1). By Corollary 5.27, $\widehat{\Omega}_X^{\text{free}}$ contains no free faces of dimension 1, nor can any appear as a result of a sequence of retraction. Furthermore, $\widehat{\Omega}_X^{\text{free}}$ contains at least one hyperplane labelled by each $v \in V\Gamma$. Let Ω be the graph onto which $\widehat{\Omega}_X^{\text{free}}$ retracts, then this is a rose graph with at least one edge labelled by each $v \in V\Gamma$. If there are two edges with the same label, these can be freely folded together. As a result, ϕ is Nielsen equivalent to (a stabilisation of) the marking represented by the rose graph with exactly one edge labelled by each $v \in V\Gamma$, which is the standard marking. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ ▱

Since only finitely many cubes in $\widehat{\Omega}_X^{\text{free}}$, only finitely many combinatorial retractions can be performed, and so this test can be made algorithmic.

If $\widehat{\Omega}_X^{\text{free}}$ does not retract onto a graph, we can still read off a generating tuple which it represents. First apply all possible retractions, then read off a presentation for the fundamental group by taking the free fundamental group of its 1-skeleton and adding relations coming from the squares. After removing some redundant generators and relations, the result is a non-free basis for a free group and the relations are primitive elements in the free group generated by this basis. By applying Whiteheads algorithm, we can replace this presentation with a free presentation and then read off from this new presentation, a generating tuple. See the first example below for an illustration of this.

5.5.2 Some non-examples

Theorem 5.30 is not sufficient to prove that all finite generating tuples of a RACG are Nielsen equivalent to some stabilisation of a standard generating tuple since in principle, $\widehat{\Omega}_X^{\text{free}}$ may not retract onto a graph. We give two examples of this phenomenon outside of the class of RACGs.

Example 5.31. We know that in the dihedral group Dih_5, the pair $X = (s_1, s_2s_1s_2)$ is not Nielsen equivalent to the standard one, see Theorem 2.1.

It is possible to generalise the construction of (free) completion sequences to arbitrary Coxeter systems (W, S), where Γ-complexes are replaced by CW complexes whose cells are Coxeter polytopes (see Theorem 4.14) which are edge-labelled by S. We will sketch this in the case of Dih_5. Folds and cell identifications remain unchanged, and cube attachments are replaced by cell attachments which encode Tits moves of type (M1). It is not immediately clear how to generalise all of the theory presented in [31] to this setting, but a straightforward argument allows us to prove a version of Corollary 5.26 in this generality.

Given a finite generating tuple X of W, we can build the rose graph (Ω_X, v_0) and find a 'free completion' of Ω_X, ie a free CW complex (as described above) in which no more free folds, cell attachments, or identifications can be performed—call this $\widehat{\Omega}_X^{\mathrm{free}}$. The idea is that, for each $s \in S$, there is some based loop in (Ω_X, v_0) whose label is a word w representing s. Take a disk diagram whose boundary is labelled by ws and which is tiled by Coxeter polygons, and attach this to (Ω_X, v_0) along the loop labelled w by a sequence of cell attachments. Doing this for each s, the result is a complex Ω' which has a based edge loop labelled by each $s \in S$. Now take a maximal tree in the 1-skeleton of Ω', and we can freely fold this onto the set of based edge loops. Doing any remaining free folds, attachments, and identifications we get $\widehat{\Omega}_X^{\mathrm{free}}$ which has the rose graph (Ω_S, v_0) embedded in it (this takes the place of $C(\widehat{\Omega}, v_0)$) in Corollary 5.26.

Doing this for the pair X given above, Ω_X folds to a 'dumbbell' graph which contains a based edge loop labelled s_1. The relation $(s_1s_2)^5$ allows us to take the word $w = s_1s_2s_1s_2s_1s_2s_1s_2s_1$ representing s_2. Attaching a decagon as shown in Figure 5.7, we can perform a single final free fold to obtain $\widehat{\Omega}_X^{\mathrm{free}}$.

By construction, $\widehat{\Omega}_X^{\mathrm{free}}$ has a free fundamental group, however the attaching map for the decagon traverses every edge at least twice, so there are no free faces from which we can retract onto a graph. This is the topological manifestation of the fact that X is not Nielsen equivalent to S. This complex can be viewed as the presentation complex for the free group $\mathbb{F}_2$ corresponding to the one-relator
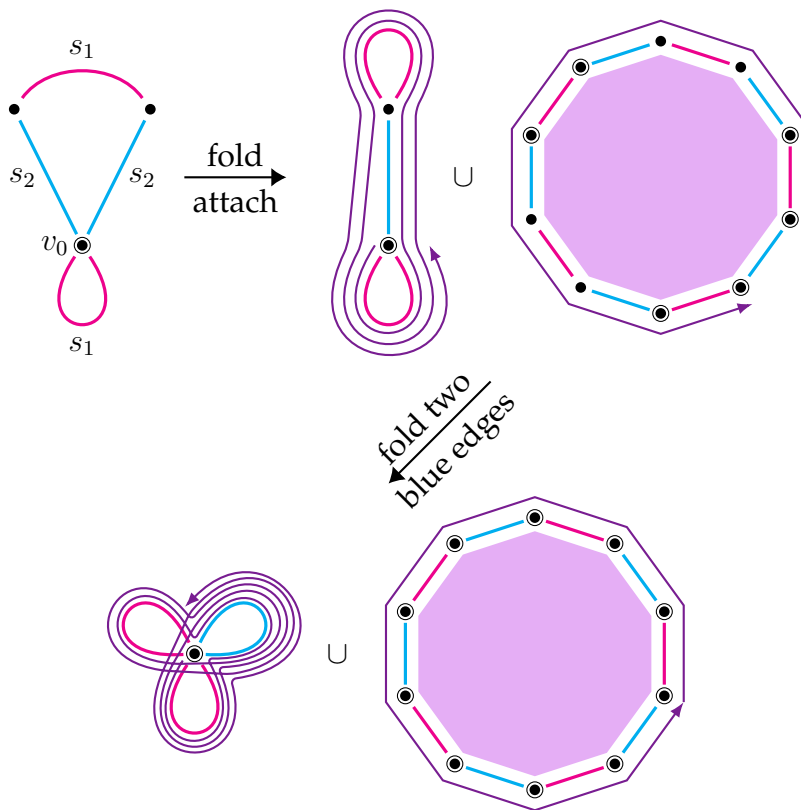
Figure 5.7: The 'free completion' of a generating pair for Dih_5 which is inequivalent to the standard one.

presentation

$$\langle x, y, z \mid xyzyxyzyxy \rangle.$$

The corresponding marking $\phi \colon \mathbb{F}_2 \to \text{Dih}_5$ maps $x, z \mapsto s_1$ and $y \mapsto s_2$. Applying Whitehead's algorithm to this presentation to find a free basis for $\mathbb{F}_2$ yields $(y, xyzyxy)$, so ϕ corresponds to the generating pair $(s_2, s_1 s_2 s_1 s_2 s_1 s_2) = (s_2, s_2 s_1 s_2 s_1)$ which, as is necessarily the case, is Nielsen equivalent to X.

Our second example is more directly related to RACGs as it concerns a quasiconvex subgroup of a RACG. Whereas previously the free completion failed to retract onto a graph because the generating tuple it came from was not Nielsen equivalent to a standard one, in this example the generating tuple *is* Nielsen equivalent to the a stabilisation of a standard generating tuple.

Example 5.32. This example is based on Louder's pathological example of a generating tuple of the closed surface group of genus 2 given in Example 4.1 of [76].

Let $G = \pi_1(S_2)$ by the fundamental group of the closed orientable surface of genus 2, with *standard* presentation

$$\langle a_1, b_1, a_2, b_2 \mid [a_1, b_1][a_2, b_2] \rangle.$$

This group can be realised as a quasiconvex subgroup of some RACG in many ways, for concreteness consider the following. Let Γ be the pentagonal graph:
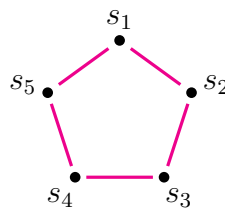
Then W_Γ acts on $\mathbb{H}^2$, generated by reflections in the sides of a right-angled pentagon. Now G can be identified with a subgroup of W_Γ by mapping

$$a_1 \mapsto s_1 s_3,$$
$$b_1 \mapsto s_2 s_4,$$
$$a_2 \mapsto s_5 s_1 s_3 s_5,$$
$$b_2 \mapsto s_5 s_2 s_4 s_5.$$

The G acts with a fundamental domain tiled by eight pentagons, so G is finite index, and hence quasiconvex. Taking the quotient $\mathbb{H}^2/G$ gives the closed orientable genus 2 surface. The tiling of $\mathbb{H}^2$ dual to the pentagonal tiling is a G-invariant square tiling, which descends to a cube complex structure on $\mathbb{H}^2/G$. Each hyperplane in $\mathbb{H}^2/G$ (thought of as a cube complex) lifts to a hyperplane in $\mathbb{H}^2$ (though of as a W_Γ space on which W acts by reflections). The stabiliser of each hyperplane in $\mathbb{H}^2$ is generated by some reflection which is conjugate to exactly one element of $V\Gamma$. Label the hyperplanes of $\mathbb{H}^2/G$ by the element of $V\Gamma$ which is conjugate into the stabiliser of the lift of the hyperplane to $\mathbb{H}^2$. This gives $\mathbb{H}^2/G$ the structure of a Γ-labelled square complex, see Figure 5.8. This is folded and cube full, and hence a completion of G. Its core graph is its entire 1-skeleton.

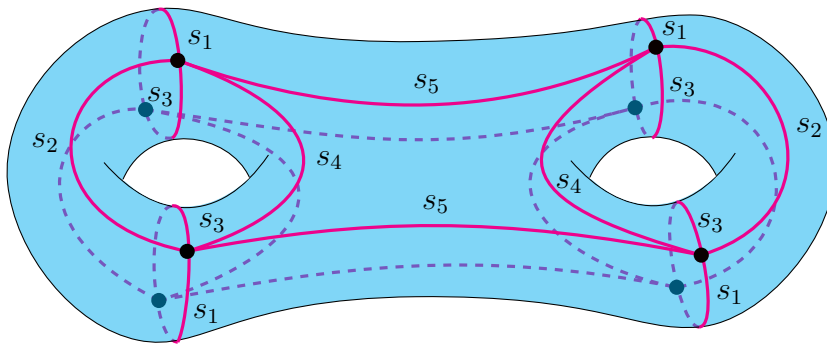We could represent the standard marking of G by a map from the wedge prod-

Figure 5.8: The genus 2 surface as a Γ-complex.

uct of two once-punctured tori, one mapping onto the left-half of the surface, the other mapping onto the right-half of the surface. To get a more interesting marking, take a 3-fold cover of the left-hand torus to get a once-punctured genus 2 surface, and a 5-fold cover of the right-hand torus to get a once-punctured genus 3 surface. Taking again the wedge of these two surfaces (calling the result Ω) and giving them a suitable Γ-complex structure, we can represent this marking by the map of spaces shown in Figure 5.9.
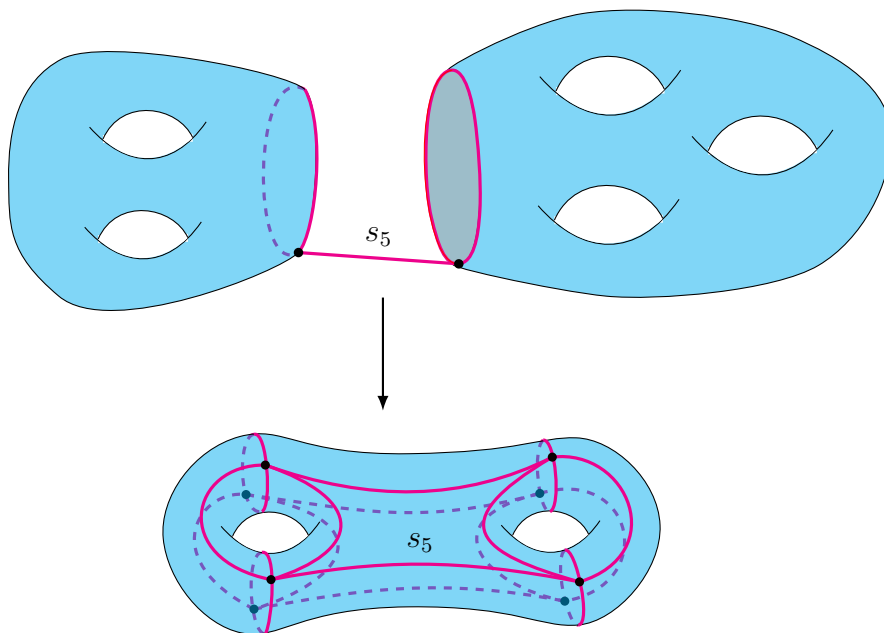
Figure 5.9: A pathological marking of G.

Algebraically, this corresponds to the generating tuple of size ten:

$$(a_1 b_1 a_1^{-1},\ a_1^2 b_1^{-1} a_1^{-1} b_1^{-1},\ b_1^2,\ b_1 a_1 b_1,$$

$$b_2^{-1} a_2 b_2 a_2 b_2 a_2^{-1} b_2^{-1},\ b_2^3, a_2 b_2^2,\ a_2^{-1} b_2^{-1},$$

$$b_2 a_2 b_2^{-1},\ b_2^{-1} a_2 b_2^{-1} a_2^{-2} b_2,\ b_2^{-1} a_2^2 b_2^{-1} a_2^{-1} b_2),$$

which expressed in terms of the generators of W_Γ becomes[]

$$\big(s_1 s_2 s_4 s_1,\ s_1 s_3 s_1 s_4 s_2 s_1 s_4 s_2,\ s_2 s_4 s_2 s_4,\ s_2 s_4 s_1 s_3 s_2 s_4,$$

$$s_5 s_4 s_1 s_3 s_4 s_1 s_2 s_4 s_1 s_4 s_2 s_5,\ s_5 s_1 s_3 s_1 s_3 s_1 s_3 s_5,\ s_5 s_1 s_2 s_4 s_2 s_4 s_1 s_4 s_2 s_5,$$

$$s_5 s_2 s_4 s_1 s_3 s_4 s_2 s_5,\ s_5 s_4 s_2 s_1 s_4 s_1 s_3 s_1 s_4 s_5,\ s_5 s_4 s_2 s_1 s_3 s_1 s_4 s_1 s_4 s_5\big)$$

The two boundary circles of Ω have length 12 and 20 respectively, and writing $w = s_1 s_4 s_1 s_4$, their labels are w^3 and w^5. By attaching a strip of 20 squares labelled alternately by $\{s_1, s_5\}$ and $\{s_4, s_5\}$, we can 'pull' the longer boundary circle across the edge labelled s_5, to get a new complex Ω'.

The boundary of Ω' is now a wedge of two circles which can be thought of as divided up into a triangle and a pentagon, with each edge labelled by w and oriented (since w, unlike the generators of W_Γ, is not an involution). Since 3 and 5 are co-prime, we can perform a sequence of Stallings folds (respecting the orientation) which map the boundary onto the 2-rose. In Figure 5.10, we have illustrated the image of the circle of length five. Notice it traverses each edge of the 2-rose at least twice, so if we call the resulting complex Ω'', it contains no free faces.
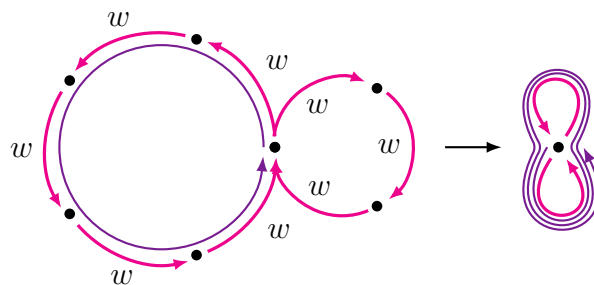

Figure 5.10: A wedge of two circles of length 3 and 5 freely fold onto the 2-rose.

Notice that, at every vertex in Ω'', there is an edge labelled by each element

of $V\Gamma$ and a square labelled by $\{s_i, s_{i+1}\}$ for each $1 \leqslant i \leqslant 5$ (where indices are read modulo 5). It follows that in any (free) completion sequence starting with Ω'', no square attachments are possible which might create a free face. One can also check that no square identifications will create a free face either. As a result, the free completion of Ω'' does not retract onto a graph.

Unlike in the previous example, this is not because the marking represented by Ω'' is not Nielsen equivalent to a stabilisation of a standard marking. This is because Louder proves that all finite generating tuples of closed surface groups are reducible or Nielsen equivalent to a standard marking [76]. Instead, it results from the fact that to see this equivalence, we must at some stage make the marking more complicated before we can fully simplify it.

5.5.3   Questions, conjectures, and code

We have implemented the Tits representation (Theorem 1.8), standard completion sequences (Section 5.1.4), and standard free completion sequences (Section 5.4) in Mathematica [102] which allows us to apply Theorems 5.14 and 5.30 to explore Nielsen equivalence in RACGs.

More specifically, the Tits representation gives an efficient solution to the word problem with which we can enumerate all elements of a given RACG W_Γ up to a certain length N. We can then randomly choose tuples of these elements of size greater than or equal to the rank of W_Γ by selecting elements uniformly without replacement from $\{w \in W \mid \ell(w) \leqslant N\}$. Next check whether they generate using Theorem 5.14, since the procedure will fail to halt if the random tuple generates a nn-quasi-convex subgroup, be put a cap on the number of steps the procedure runs for before it gives up and assumes the tuple does not generate. Finally, we can apply Theorem 5.30 to each random generating tuple we find to see whether it is 'obviously' Nielsen equivalent to a standard generating tuple (obvious in the sense that the standard free completion retracts onto a graph), or, if this test fails, it needs to be studied more closely.

In particular, we have done this for several choices of Γ and most of all for Γ the pentagonal graph of Example 5.32. Every generating tuple we have found has

been 'obviously' Nielsen equivalent to some stabilisation of a standard generating tuple, and on the strength of this we make the following Conjecture.

CONJECTURE 5.33: *Let (W, S) be a Coxeter system for a RACG, then every generating tuple for W is Nielsen equivalent to (some stabilisation of) S.*

In light of Corollary 4.45, it suffices to show that every generating tuple of a RACG is Nielsen equivalent to a tuple of reflections. Alternatively, one could ask the following question, a positive answer to which would prove the Conjecture.

Question 5.34. For every RACGs W_Γ, and any generating tuple X, does $\widehat{\Omega}_X^{\text{free}}$ always retract onto a graph?

More generally, everything in this Chapter up until Section 5.5 works in the context of quasiconvex subgroups of RACGs. Generalising the definition of a *standard generating tuple* for such a group (with a fixed quasiconvex embedding into W_Γ), and proving that all such standard generating tuples are equivalent is not completely straightforward. Nevertheless, it is reasonable to imagine that some version of Theorem 5.30 holds in this case. Of course, Example 5.32 demonstrates that Question 5.34 does not have a positive answer for arbitrary quasiconvex subgroups, however we can ask the following question.

Question 5.35. Can free completion sequences be used to study Nielsen equivalence in quasiconvex subgroups of RACGs?

Leaving the world of Coxeter groups, Michael Ben-Zvi, Robert Kropholler, and Rylee Alanza Lyman generalise the completion sequences of Dani and Levcovitz to study subgroups of fundamental groups of non-positively curved cube complexes. This includes the class of right-angles Artin groups we mentioned at the start of Chapter 2. Therefore we can ask to what extent the work presented here generalises to that setting.

Question 5.36. Can free completion sequences be generalised to the setting of non-positively curved cube complexes in order to study Nielsen equivalence in the associated groups?

Part II:

Equivariant machine learning

> We can only see a short distance ahead, but we can see plenty there that needs to be done.
>
> ALAN TURING (1912–1954)
>
> *Mathematician and pioneer of computer science*

Chapter 6

Background on machine learning

PART II OF THIS THESIS DISCUSSES joint work with Benjamin Aslan and Daniel Platt, much of which is featured in [3]. We approach the well-studied problem of group invariant and equivariant *supervised machine learning* from the point of view of geometric topology. In this Chapter we give the relevant background on supervised machine learning in general, as well as introducing group equivariant and invariant versions of machine learning. We summarise the literature on this topic, and expand on a couple of specific approaches which have been tried in the past.

In Chapter 7, first we provide an overview of the mathematical underpinnings of our approach to invariant and equivariant machine learning via projection maps. Then we set out a unified model of so-called *intrinsic* approaches and apply our own to several example machine learning problems.

The details of the projections we use are explored in Chapters 8 and 9. In the first case we consider the special, but widely applicable, case of a group acting on $\mathbb{R}^{n_0}$ by permuting coordinates. We define several versions of a projection onto a fundamental domain using combinatorial tools from the study of permutation groups. We give an explicit algorithm to compute such projections in general and analyse the efficiency of this algorithm. We also explicitly compute the projections for several group actions, showing that even more efficient algorithms to compute the projections using sorting can be applied in special cases. The second half of the Chapter is devoted to proving that the combinatorial maps that we define really do project onto a fundamental domain for the action.

In the final Chapter we discuss another, more generally applicable, method of finding projections onto a fundamental domain based on the idea of a Dirichlet fundamental domain and gradient descent. We also give an example computing an isometric embedding for the quotient space of one of the group actions featured in our example machine learning problems, namely $\mathbb{Z}_4$ acting on $\mathbb{R}^4 \otimes \mathbb{R}^n$ by cyclically permuting the coordinates in the first factor. Finally we quantitatively compare our projections with a related approach which uses non-isometric embeddings of quotient spaces.

6.1   Supervised machine learning

Supervised machine learning refers to a large class of algorithms and techniques used to make computers solve complex problems by 'learning' a relationship between a large set of data and labels for that data. A typical example is image recognition, where the data set consists of images of, say, animals, and the labels are the English names for these animals.

As an example to have in mind of how a machine learning algorithm might work, we briefly sketch a simple *neural network* which could be applied to such an image recognition task. For the purpose of illustration, assume that each image consists of an $(n \times n)$-array of pixels, each of which can take a brightness value between 0 and 255 (ie the pictures are monochrome). Thus, any image can be represented by a point in $\mathbb{R}^{n^2}$. Further, assume that the images are to be classified into k categories, then we can represent this classification by a map $\alpha \colon \mathbb{R}^{n^2} \to \mathbb{R}^k$, where an image which falls into the ith category is mapped to the ith standard unit vector in $\mathbb{R}^k$.

A neural network is then a model which can approximate α. Mathematically, a neural network is a variable which takes values in the class of functions

$$\mathfrak{M}_{n_0, n_1, \ldots, n_{\ell+1}}(\sigma)$$

of the form

$$\beta = \mathbb{R}^{n_0} \xrightarrow{\lambda_0 + \tau} \mathbb{R}^{n_1} \xrightarrow{\sigma} \mathbb{R}^{n_1} \xrightarrow{\lambda_1} \cdots \xrightarrow{\lambda_{\ell-1}} \mathbb{R}^{n_\ell} \xrightarrow{\sigma} \mathbb{R}^{n_\ell} \xrightarrow{\lambda_\ell} \mathbb{R}^{n_{\ell+1}}$$

where $n_0 = n^2$, $n_{\ell+1} = k$, each λ_i is linear, τ is an affine translation of $\mathbb{R}^{n_1}$, and σ is a non-linear self map called an *activation function*. Each space $\mathbb{R}^{n_i}$ is called a *layer*— the first and last are the *input* and *output* layers respectively, and the interstitial spaces are *hidden* layers. A commonly used activation function is the *rectified linear unit* (ReLU) which maps component-wise $x \mapsto \max\{0, x\}$.

There are several ways to think about the role of the activation function σ. From a mathematical point of view, without it β could only be an affine map which greatly limits its expressiveness; σ is what allows β to be a more compli- cated function. From a more applied point of view, one can think that each of the coordinates in some intermediate space $\mathbb{R}^{n_i}$ is variables which we only want to turn on above a certain threshold value, say, and the activation function performs this task if it is ReLU. Overall, β is a piece-wise affine map (or for other choices of σ, it can be thought of as a smoothed out piece-wise affine map).

A neural network 'learns' the map α by defining some form of *cost function*, which measures to what extent β approximates α, and then modifying β to de- crease this cost. More precisely, we start with a set of training data $D_{\text{train}} = \{(x, \alpha(x)) \mid x \in X_{\text{train}} \subset \mathbb{R}^{n_0}\} \subset \mathbb{R}^{n_0} \times \mathbb{R}^{n_{\ell+1}}$, which samples the function α. Now we can choose a norm on a suitable class of functions which contains α and $\mathfrak{M}_{n_0, n_1, \ldots, n_{\ell+1}}(\sigma)$ and approximate the norm of $(\alpha - \beta)$ by restricting these functions to the set X_{train}.

Once each of the linear maps λ_i has been expressed in terms of bases, they are given by matrices whose entries (called *weights*) are variables. Starting with some randomly chosen $\beta \in \mathfrak{M}_{n_0, n_1, \ldots, n_{\ell+1}}(\sigma)$, we perform gradient descent with respect to the norm of $(\alpha - \beta)$, at each step updating the weights to approach some local minimum. In this way, we compute algorithmically a piece-wise linear approximation for α.

It is natural to wonder how expressive neural networks can be. In other words, can they be used to approximate any 'reasonable' function $\alpha \colon \mathbb{R}^{n_0} \to \mathbb{R}^{n_{\ell+1}}$? A positive answer is given by the *Universal Approximation Theorems*, a collection of fundamental results in the theory of machine learning which state that any 'suf- ficiently nice' function $\alpha \colon \mathbb{R}^n \to \mathbb{R}^m$ can be arbitrarily well approximated by a

neural network with one hidden layer.

To state these Theorems precisely, we need to specify the norm with respect to which we are measuring how well functions approximate each other. Versions of the Universal Approximation Theorem hold weighted Sobolev norms (Theorems 3 and 4 in [65]). Here we will state the versions for L^p norms and the supremum norm.

Definition 6.1. Given a function $f\colon \mathbb{R}^n \to \mathbb{R}^m$ and a measure μ on X, its L^p **norm** (for $1 \leqslant p < \infty$) is

$$\|f\|_p^\mu := \left(\int_{\mathbb{R}^n} \sum_{i=1}^{m} |f_i(x)|^p \mathrm{d}\mu(x) \right)^{\frac{1}{p}},$$

where $f_i(x)$ is the projection of $f(x)$ to the ith coordinate of $\mathbb{R}^m$. We write $L^p(\mu, \mathbb{R}^m)$ for the set of functions $f\colon \mathbb{R}^n \to \mathbb{R}^m$ such that $\|f\|_p^\mu < \infty$.

Then the distance between α and β is $\|\alpha - \beta\|_p^\mu$. Denote by $\mathfrak{M}_{n,*,m}(\sigma)$ the set of functions $\mathbb{R}^n \to \mathbb{R}^m$ implemented by a neural network with activation function $\sigma\colon \mathbb{R} \to \mathbb{R}$ (which we extend to ant $\mathbb{R}^k$ by applying it component-wise) and one hidden layer with arbitrarily many neurons. In particular, a function $\beta \in \mathfrak{M}_{n,*,m}(\sigma)$ has the form

$$\beta\colon \mathbb{R}^n \to \mathbb{R}^m \colon x \to \lambda_1 \sigma(\lambda_0 x + \tau),$$

where $\lambda_0 \in \mathbb{R}^{k \times n}$, $\tau \in \mathbb{R}^k$, and $\lambda_1 \in \mathbb{R}^{m \times k}$ for some $k \in \mathbb{N}$. The Universal Approximation Theorem, as stated in Theorem 1 of [65] is as follows.

THEOREM 6.2 (L^p Universal Approximation Theorem): *For a bounded and non-constant activation function σ, $\mathfrak{M}_{n,*,m}(\sigma)$ is dense in $L^p(\mu, \mathbb{R}^m)$ for all finite measures μ on $\mathbb{R}^n$.*

A measure is *finite* if $\mu(\mathbb{R}^n) < \infty$. A prototypical situation to bear in mind is as follows. Let $X \subset \mathbb{R}^n$ be compact, and suppose we want to approximate a function $\alpha\colon X \to \mathbb{R}^m$ with a neural network. Extend the definition of α to $\mathbb{R}^n$ by setting it equal to 0 on $\mathbb{R}^n \backslash X$ and consider the measure space $(\mathbb{R}^n, \mathcal{L}, \mu)$ where $\mathcal{L}$ is the σ-algebra of Lebesgue measurable sets. Let μ be the restriction of the Lebesgue measure μ_{L} on $\mathbb{R}^n$ to X, ie for $A \in \mathcal{L}$, $\mu(A) = \mu_{\mathrm{L}}(A \cap X)$. Theorem 6.2 implies that α can be arbitrarily well approximated by a neural network in $L^p(\mu, \mathbb{R}^m)$.

The example of σ being ReLU does not satisfy the hypothesis of this version of the Universal Approximation Theorem, but another example of a function often used which does is the sigmoid function which is applied component-wise by $x \mapsto (1 + e^{-x})^{-1}$.

For the second version of the Universal Approximation Theorem we state, we need to define a second norm.

Definition 6.3. Given a function $f \colon X \to \mathbb{R}^m$ on a compact set $X \subset \mathbb{R}^n$, its **supremum norm** is $\|f\|_\infty^X := \sup_{x \in X} |f(x)|$.

Let us denote by $\mathfrak{C}(X, \mathbb{R}^m)$ the set of continuous functions $X \to \mathbb{R}^m$ from a compact subset $X \subset \mathbb{R}^n$. The Universal Approximation Theorem, as stated in Theorem 2 of [65] is as follows.

THEOREM 6.4 (Supremum Universal Approximation Theorem): *For a continuous, bounded, and non-constant activation function σ, $\mathfrak{M}_{n,*,m}(\sigma)$ is dense in $\mathfrak{C}(X, \mathbb{R}^m)$ for all compact subsets $X \subset \mathbb{R}^n$.*

From now on we work in a somewhat more general setting than that described for neural networks. We wish to approximate a function $\alpha \colon X \to Y$ between an input (or *feature*) space and an output space. Typically, these may be subsets of $\mathbb{R}^n$, but could be more complicated like Riemannian manifolds. We consider the problem in the case that α is invariant under some left action of a group G on X (or, more generally, equivariant with respect to actions of G on X and Y).

Machine learning models such as neural networks or random forests can approximate α, but the resulting function β is generally not invariant under the action G. The key task is to define machine learning algorithms producing functions $\beta \colon X \to Y$ which are necessarily invariant.

6.2   Literature on invariant and equivariant machine learning

Machine learning models which are invariant (or equivariant) under the action of a group G have been extensively studied in the machine learning literature.

In [115], Dmitri Yarotsky distinguishes two different approaches to the problem: *symmetrisation based* and *intrinsic* approaches. The first involves averaging some non G-invariant model over the action of G to produce an (approximately) G-invariant model; whereas intrinsic approaches involve designing the model to be G-invariant *a priori* by imposing conditions coming from the group action.

A standard symmetrisation based approach is *data augmentation*, which was used in early works such as [72], and is surveyed in [22]. Given training data $D_{\text{train}} = \{(x,y) \mid x \in X_{\text{train}} \subset X, y = \alpha(x) \in Y\}$, it involves adding more training data by applying sample elements $G_0 \subset G$ to the inputs. The new training data is then $D_{\text{train}}^{\text{aug}} := \{(g \cdot x, y) \mid (x,y) \in D_{\text{train}} \text{ and } g \in G_0\}$ in the invariant case, or $\{(g \cdot x, g \cdot y)\}$ in the equivariant case. A similar approach is to take a machine learning architecture β and apply it to several G-translates of an input, before applying a pooling map to these different outputs. This yields a G-invariant map and was studied in [5].

We now turn to examples of *intrinsic* approaches. For neural networks, one can impose restrictions on the weights (ie the coefficients of the linear maps) so that the resulting network is invariant under a group action on the input. This was done using group equivariant hidden layers, for example, in [59, 116]. The same idea is also used in [81, 96, 97].

Convolutional layers in neural networks are a standard tool to impose translational symmetry in image classification tasks. This idea has been generalised to group equivariant convolutional neural networks in [25] for actions by arbitrary discrete groups. Another intrinsic approach is proposed in Section 2 of [115] based on the theory of *polynomial invariants* of G. All of these approaches are concerned with discrete symmetries. Below we discuss two of these approaches in slightly more detail. The study of continuous symmetries was initiated in [71] and expanded in [108, 27]; and the case of Euclidean transformations has received additional attention, for example in [52, 112].

6.3   Equivariant layers in neural networks

A G-equivariant neural network (see for example [81]) consists of a series of G-equivariant linear maps λ_i separated by some non-linear activation function σ, yielding $\beta = \lambda_\ell \circ \sigma \circ \cdots \circ \sigma \circ \lambda_1$. Restrictions are placed on the learnable parameters of each λ_i to ensure they are G-equivariant. For example, if $\lambda_i \colon \mathbb{R}^n \to \mathbb{R}^n$ is equivariant with respect to S_n acting on each copy of $\mathbb{R}^n$ by permuting coordinates, then it was shown in Lemma 3 of [116], that it must have the form

$$\lambda_i(x) = (a\mathbb{I} + b\mathbb{1}^T\mathbb{1})x,$$

where $a, b \in \mathbb{R}$ are learnable parameters, $\mathbb{I}$ represents the identity matrix, and $\mathbb{1} = (1, 1, \ldots, 1)$. For arbitrary groups, the main task is to describe the space of all G-equivariant linear maps $\lambda \colon \mathbb{R}^{n_1} \to \mathbb{R}^{n_2}$ which could map between layers in the neural network. One approach to this problem is based on decomposing the representation of G on $\mathbb{R}^{n_1}$ and $\mathbb{R}^{n_2}$ into irreducible components and applying Schur's Lemma. Other methods to determine all equivariant linear layers were proposed in [26, 46].

6.4   Equivariant maps from polynomial invariants

We discuss briefly the approach proposed by Yarotsky in Section 2 of [115] based on the theory of polynomial invariants of a group G acting on $X = \mathbb{R}^n$. A polynomial $p(x) \in \mathbb{R}[x_1, \ldots, x_n]$ is called a *polynomial invariant* if it defines an invariant map $\mathbb{R}^n \to \mathbb{R}$. Similarly, if G also acts on $\mathbb{R}^m$, then $q(x) \colon \mathbb{R}^n \to \mathbb{R}^m$ is a *polynomial equivariant* if it is a polynomial such that $q(g \cdot x) = g \cdot q(x)$ for all $g \in G$ and $x \in \mathbb{R}^n$.

   The following, proved in [92] in the invariant case, was generalised to compact Lie groups in Theorem 8.14.A of [113].

Theorem 6.5 (Section 4 of [114])**:** *If a finite group G acts on $\mathbb{R}^n$ and $\mathbb{R}^m$, then there are finite sets of invariants $\{p_i(x)\}_{i=1}^k$, and equivariants $\{q_j(x)\}_{j=1}^l$ such that any polynomial equivariant $q(x)$ can be written as $q(x) = \sum_{j=1}^l q_j(x) r_j(p_1(x), \ldots, p_k(x))$ for some $r_j(x) \in \mathbb{R}[x_1, \ldots, x_k]$. In the invariant case, when G acts trivially on $\mathbb{R}^m$, we can take $l = 1$, and $q_1(x) = 1$.*

Yarotsky shows in Proposition 2.4 of [115] that any continuous G-equivariant function $\alpha \colon \mathbb{R}^n \to \mathbb{R}^m$ can be approximated on a compact set by a G-equivariant neural network of the form

$$\beta(x) = \sum_{j=1}^{l} q_j(x) \sum_{h=1}^{d} a_{jh}\, \sigma\left(\sum_{i=1}^{k} b_{jhi} p_i(x) + c_{jh}\right) \tag{6.1}$$

for some $a_{jh}, b_{jhi}, c_{jh} \in \mathbb{R}$, where $d \in \mathbb{N}$, and σ is a continuous non-polynomial activation function.

Notice that all the learnable parameters are contained in the inner two sums, which also constitute the neural network in the G-invariant case

$$\sum_{h=1}^{d} a_{h}\, \sigma\left(\sum_{i=1}^{k} b_{hi} p_i(x) + c_{h}\right), \tag{6.2}$$

see Proposition 2.3 in [115].

Chapter 7

Geometric approach to equivariant machine learning

THERE ARE MANY VARIATIONS ON this basic model of a neural network such as convolutional networks; as well as many other models for supervised machine learning, for example random forests or support vector machines (SVM). We propose a novel approach to group invariant and equivariant machine learning using a data pre-processing step. This involves projecting the input data into a geometric space which parametrises the orbits of the group, either a fundamental domain, or the quotient space. A significant advantage of this method over several others, is that this new data can then be the input for an arbitrary machine learning model (neural network, random forest, support-vector machine etc).

We give algorithms to compute the projection onto a fundamental domain, one which works very generally, and another which is specialised to groups acting on $\mathbb{R}^{n_0}$ by permuting coordinates. These algorithms are efficient to implement, and we illustrate our approach on some example machine learning problems (including the well-studied problem of predicting Hodge numbers of CICY matrices), in each case finding an improvement in accuracy versus others in the literature. The geometric topology viewpoint also allows us to give a unified description of so-called *intrinsic* approaches to group equivariant machine learning, which encompasses many other approaches in the literature.

7.1 Overview

We focus on the case of *G-invariant* machine learning—The extension of this to *G-equivariant* machine learning is explained in Remark 7.3 below and is fundamentally no harder than the invariant case. Our approach to the problem is intrinsic, based on the fact that composing a G-invariant map with any other map, results in a G-invariant map. One way of getting a G-invariant self-map of the feature space is to map to a *fundamental domain* $\mathcal{F}$, which preserves the local geometry of the feature space. The set $\mathcal{F} \subset X$ comes with a G-invariant map $\pi \colon X \to \overline{\mathcal{F}}$ onto its closure. Let $\overline{\alpha}$ be the restriction of α to $\overline{\mathcal{F}}$, then by G-invariance $\alpha = \overline{\alpha} \circ \pi$. Instead of fitting a machine learning model $\beta \colon X \to Y$ to the training data D_{train}, we train the model $\overline{\beta} \colon \overline{\mathcal{F}} \to Y$ with $D_{\text{train}}^{\pi} := \{(\pi(x), y) \mid (x, y) \in D_{\text{train}}\} \subset \overline{\mathcal{F}} \times Y$ which approximates $\overline{\alpha}$. The resulting map $\beta = \overline{\beta} \circ \pi \colon X \to Y$ is G-invariant. Figure 7.1 shows the difference between the pre-processing approaches of augmentation and our method.
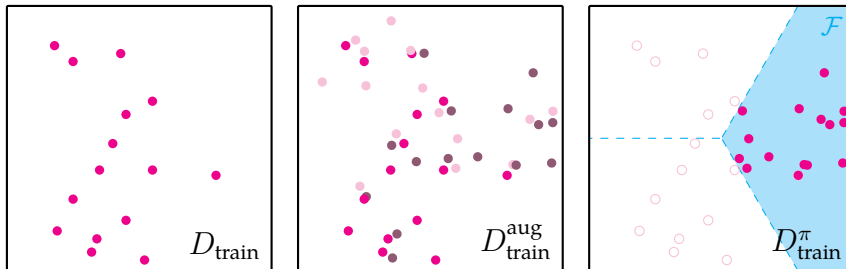
Figure 7.1: Example training data D_{train} for a problem invariant under rotations of $2\pi/3$. Also the processed training data after augmentation $D_{\text{train}}^{\text{aug}}$, and our approach D_{train}^{π} (mapping all the data to the blue fundamental domain for the action).

One advantage our approach has is that it can be applied directly to any supervised machine learning model. In contrast, many existing methods, such as [59, 116, 81, 96, 97, 46, 26] only work for neural networks. The computational cost of data augmentation and many equivariant machine learning approaches scales with the size of the symmetry group. This is not the case for our approach, which we implement for groups of size $6 \cdot 10^{20}$ in Section 7.4.2.

7.2   Mathematical approach

In principle, our approach works in a very general setting, however we restrict ourselves to the following setting.

Assumption 7.1. For the remainder of this thesis, unless stated otherwise, the feature space X is a Riemannian manifold on which the G acts discretely by isometries.

We want to approximate a G-invariant function α, ie a function which takes the same value on every element of a G-orbit. A set $R \subset X$ is a *set of orbit representatives* if for all $x \in X$, $R \cap (G \cdot x) \neq \emptyset$. If we approximate α on a set of orbit representatives then we have essentially approximated it everywhere. A nice choice of orbit representatives which takes into account the geometry of the group action is given by a fundamental domain.

Definition 7.2. Let a group G act on X discretely by isometries. A subset $\mathcal{F} \subset X$ is called a **fundamental domain** for G if

- It is open and connected;

- Every G-orbit intersects $\overline{\mathcal{F}}$, the closure of $\mathcal{F}$, in at least one point; and

- Whenever a G-orbit intersects $\overline{\mathcal{F}}$ at a point in $\mathcal{F}$, then this is the unique point of intersection with $\overline{\mathcal{F}}$.

Given G acting on X, we find a G-invariant map $\pi \colon X \to \overline{\mathcal{F}}$, defined as $\pi(x) = \phi(x) \cdot x$, where $\phi \colon X \to G$ is some suitable function. We call such a map a *projection onto the fundamental domain $\mathcal{F}$*. We can now apply a machine leaning architecture to approximate the function $\alpha|_{\overline{\mathcal{F}}} \colon \overline{\mathcal{F}} \to Y$ trained on the data D_{train}^{π} yielding a function $\overline{\beta}$. This can then be used to compute the G-invariant approximation for α defined on the whole of X by defining $\beta = \overline{\beta} \circ \pi$.

$$
\begin{array}{ccc}
X & \xrightarrow{\ \beta\ } & Y \\
{\scriptstyle \pi}\Big\downarrow & \nearrow_{\overline{\beta}} & \\
\overline{\mathcal{F}} & &
\end{array}
$$

If $\overline{\beta}$ is defined using a neural network, then the natural universal approxima-
tion property is satisfied, namely β can approximate any continuous, G-invariant
map α arbitrarily closely. This follows from the standard Universal Approxima-
tion Theorem and is proved in Section 7.2.1.

Remark 7.3 (Equivariant machine learning). Our method of producing a G-invariant
architecture can easily be modified to the G-equivariant setting. Let G act on X
and Y, then a G-equivariant map $\alpha \colon X \to Y$ satisfies $\alpha(g \cdot x) = g \cdot \alpha(x)$ for all
$x \in X$ and $g \in G$. Let $\pi \colon X \to \overline{\mathcal{F}}$ be the fundamental domain projection as above,
and define $\phi \colon X \to G$ be a function such that $\pi(x) = \phi(x) \cdot x$. Then we define the
β model via
$$\beta(x) := \phi(x)^{-1} \cdot \overline{\beta}(\pi(x)) = \phi(x)^{-1} \cdot \overline{\beta}(\phi(x) \cdot x),$$
which is indeed G-equivariant.

Remark 7.4 (Invariance on the boundary). The projection π is continuous on the
preimage of $\mathcal{F}$ in X, but may fail to be continuous on the preimage of the bound-
ary $\partial \mathcal{F}$ of $\mathcal{F}$. Here ϕ is not necessarily G-invariant, and so the function β may
not be strictly G-invariant/equivariant on $\partial \mathcal{F}$. This only presents a problem if a
significant portion of D_{train}^{π} lies in $\partial \mathcal{F}$. We give an example of this in Section 7.4.

The way to redress this issue with the boundary is to identify points in $\partial \mathcal{F}$
which lie in the same orbit. The resulting space is the *quotient space* for the action
of G on X. We can define this quotient intrinsically as follows.

Definition 7.5. Let G be a group acting on X, then the **quotient space** X/G is the
set of all G-orbits of points in X, $\{G \cdot x \mid x \in X\}$. A quotient space is automatically
equipped with a G-invariant map $\pi_X \colon X \to X/G \colon x \mapsto G \cdot x$. If X is a subset of
$\mathbb{R}^n$ and the action is by isometries, X/G inherits a metric from the Riemannian
metric on X, and π_X is a local isometry.

The quotient space is defined abstractly, but to be useful as input for a neural
network, say, we must realise it as a subset of a vector space. In order to pre-
serve the geometry of the quotient space, this embedding should be isometric,
and so we suppose we have such an isometric embedding $X/G \hookrightarrow \mathbb{R}^{n_0'}$, and view

π as a projection onto this image. Note that in general finding an explicit embedding can be very difficult, and the dimension n_0' may be significantly greater than n_0. In this case we train the machine learning model $\overline{\beta}\colon \mathbb{R}^{n_0'} \to \mathbb{R}^{n_{\ell+1}}$ on $D_{\text{train}}^{\pi} = \{(\pi(x), \alpha(x)) \mid (x, \alpha(x)) \in D_{\text{train}}\}$, and define $\beta = \overline{\beta} \circ \pi$.

7.2.1 Universal Approximation Theorem

To obtain a G-invariant version of Theorem 6.2, let $L_G^p(\mu, \mathbb{R}^m)$ be the set of G-invariant functions $f\colon \mathbb{R}^n \to \mathbb{R}^m$ such that $\|f\|_p^{\mu} < \infty$, where μ is a G-invariant finite measure. Note that, for a group acting by isometries on $\mathbb{R}^n$ with fundamental domain $\mathcal{F}$, the existence of a non-zero G-invariant finite measure implies that $\#G$ is finite. Indeed, since G must me countable,

$$
\begin{aligned}
\infty > \mu(\mathbb{R}^n) &= \mu\left(\bigsqcup_{g \in G}(g \cdot \mathcal{F}) \sqcup \left(\bigcup_{g \in G} g \cdot \partial \mathcal{F}\right)\right) = \sum_{g \in G} \mu(g \cdot \mathcal{F}) + \mu\left(\bigcup_{g \in G} g \cdot \partial \mathcal{F}\right) \\
&\geqslant \sum_{g \in G} \mu(\mathcal{F}) = \#G\mu(\mathcal{F}).
\end{aligned}
$$

Let $\mathfrak{M}_{n,*,m}^G(\sigma)$ be the set of functions of the form $\overline{\beta} \circ \pi$ where $\pi\colon \mathbb{R}^n \to \mathbb{R}^n$ is a projection onto a fundamental domain $\mathcal{F}$, and $\overline{\beta} \in \mathfrak{M}_{n,*,m}(\sigma)$.

THEOREM 7.6 (*G-invariant L^p Universal Approximation Theorem*)**:** *If σ is bounded and non-constant, then $\mathfrak{M}_{n,*,m}^G(\sigma)$ is dense in $L_G^p(\mu, \mathbb{R}^m)$ for all finite G-invariant measures μ on $\mathbb{R}^n$.*

Proof. The claim is trivial for the zero measure, so assume that μ is not the zero measure, and hence that $\#G < \infty$. We need to show that, given $\varepsilon > 0$, a map $\alpha\colon \mathbb{R}^n \to \mathbb{R}^m$ in $L_G^p(\mu, \mathbb{R}^m)$, and a projection $\pi\colon \mathbb{R}^n \to \mathbb{R}^n$ onto a fundamental domain $\mathcal{F}$, there is a neural network $\overline{\beta} \in \mathfrak{M}_{n,*,m}(\sigma)$ such that $\|\alpha - \overline{\beta} \circ \pi\|_p^{\mu} < \varepsilon$. Let

$$
\overline{\alpha}\colon \mathbb{R}^n \to \mathbb{R}^m : x \mapsto \alpha(x)\mathbb{1}_{\overline{\mathcal{F}}}(x),
$$

where $\mathbb{1}_{\overline{\mathcal{F}}}$ is the characteristic function of $\overline{\mathcal{F}} \subset \mathbb{R}^n$. Since $\alpha \in L_G^p(\mu, \mathbb{R}^m)$, it follows that $\overline{\alpha} \in L^p(\mu, \mathbb{R}^m)$, so by Theorem 6.2, there is a network $\overline{\beta}\colon \mathbb{R}^n \to \mathbb{R}^m$ such that

$\|\overline{\alpha} - \overline{\beta}\|_p^\mu < \varepsilon/(\#G)^{\frac{1}{p}}$. Furthermore, by construction $\alpha = \overline{\alpha} \circ \pi$, which implies that

$$\alpha - \overline{\beta} \circ \pi = \overline{\alpha} \circ \pi - \overline{\beta} \circ \pi = (\overline{\alpha} - \overline{\beta}) \circ \pi.$$

Now, using the definition of $\| \cdot \|_p^\mu$ and the G-invariance of μ we can compute

$$
\begin{aligned}
(\|\alpha - \overline{\beta} \circ \pi\|_p^\mu)^p &= (\|(\overline{\alpha} - \overline{\beta}) \circ \pi\|_p^\mu)^p \\
&= \int_{\mathbb{R}^n} \sum_{i=1}^{m} |((\overline{\alpha} - \overline{\beta}) \circ \pi)_i|^p \mathrm{d}\mu(x) \\
&\leqslant \sum_{g \in G} \int_{g \cdot \mathcal{F}} \sum_{i=1}^{m} |((\overline{\alpha} - \overline{\beta}) \circ \pi)_i|^p \mathrm{d}\mu(x) \\
&= \#G \int_{\mathcal{F}} \sum_{i=1}^{m} |((\overline{\alpha} - \overline{\beta}) \circ \pi)_i|^p \mathrm{d}\mu(x) \\
&= \#G \int_{\mathcal{F}} \sum_{i=1}^{m} |(\overline{\alpha} - \overline{\beta})_i|^p \mathrm{d}\mu(x) \\
&\leqslant \#G \int_{\mathbb{R}^n} \sum_{i=1}^{m} |(\overline{\alpha} - \overline{\beta})_i|^p \mathrm{d}\mu(x) \\
&= \#G(\|\overline{\alpha} - \overline{\beta}\|_p^\mu)^p \leqslant \#G \varepsilon^p / \#G = \varepsilon^p,
\end{aligned}
$$

where the subscript i indicates projection to the ith coordinate of $\mathbb{R}^m$. Hence, $\|\alpha - \overline{\beta} \circ \pi\|_p^\mu \leqslant \varepsilon$, as required. $\hfill\square$

We can also prove a G-invariant version of Theorem 6.4. Let $\mathfrak{C}^G(X, \mathbb{R}^m)$ be the set of G-invariant continuous functions $X \to \mathbb{R}^m$. In order to apply Theorem 6.4, we need to slightly restrict the domain of the functions with which we work.

Definition 7.7. Let π be a projection onto a fundamental domain $\mathcal{F}$ for G acting by isometries on $\mathbb{R}^n$. A subset $X \subset \mathbb{R}^n$ is π-**admissible** if π is continuous when restricted to X.

By Remark 7.4, we know that if $\pi(X) \subset \mathcal{F}$ (ie the image of X does not meet the boundary of $\mathcal{F}$), then X is π-admissible. Since the preimage of $\partial \mathcal{F}$ has measure zero in $\mathbb{R}^n$ (with respect to the standard Lebesgue measure), restricting to π-admissible sets does not limit the usefulness of what follows except in the case that the training data is projected onto $\partial \mathcal{F}$ by π.

Theorem 7.8 (*G*-invariant Supremum Universal Approximation Theorem): *If σ is continuous, bounded, and non-constant, then $\mathfrak{M}_{n,*,m}^{G}(\sigma)$ is dense in $\mathfrak{C}^{G}(X, \mathbb{R}^m)$ for all compact, π-admissible, and G-invariant subsets X of $\mathbb{R}^n$.*

Proof. We need to show that, given $\varepsilon > 0$, a map $\alpha \colon X \to \mathbb{R}^m$ in $\mathfrak{C}^{G}(X, \mathbb{R}^m)$, and a projection $\pi \colon \mathbb{R}^n \to \mathbb{R}^n$ onto a fundamental domain $\mathcal{F}$, there is a neural network $\overline{\beta} \in \mathfrak{M}_{n,*,m}(\sigma)$ such that $\|\alpha - \overline{\beta} \circ \pi\|_{\infty}^{X} < \varepsilon$. Let $\overline{\alpha}$ be the restriction of α to $\pi(X)$.

Since X is π-admissible, $\pi(X)$ is compact. Then, from the fact that $\alpha \in \mathfrak{C}^{G}(X, \mathbb{R}^m)$, it follows that $\overline{\alpha} \in \mathfrak{C}^{G}(\pi(X), \mathbb{R}^m)$. By Theorem 6.4, there is a network $\overline{\beta} \colon \mathbb{R}^n \to \mathbb{R}^m$ such that $\|\overline{\alpha} - \overline{\beta}\|_{\infty}^{\pi(X)} < \varepsilon$. Furthermore, by construction $\alpha = \overline{\alpha} \circ \pi$. Then

$$\|\alpha - \overline{\beta} \circ \pi\|_{\infty}^{X} = \|\overline{\alpha} \circ \pi - \overline{\beta} \circ \pi\|_{\infty}^{X} = \|(\overline{\alpha} - \overline{\beta}) \circ \pi\|_{\infty}^{X}$$
$$= \|(\overline{\alpha} - \overline{\beta}) \circ \pi\|_{\infty}^{\pi(X)}$$
$$= \|\overline{\alpha} - \overline{\beta}\|_{\infty}^{\pi(X)} < \varepsilon.$$

Here, the equality in the second line follows because π is G-invariant, and the equality in the third line follows from the fact that π is the identity on $\pi(X)$. □

There is something of a trade-off between Theorem 7.6 and Theorem 7.8. In the first we put no restriction on the domain the of the function α in order to be able to approximate it with a neural network, but the group acting must be finite. Conversely, in the second, the group acting can be infinite, however we have to restrict to approximating α on π-admissible compact sets.

7.2.2 Comparing approaches to equivariant machine learning

We can compare the various approaches to invariant machine learning discussed in Section 6.2 on a theoretical level; below we compare them experimentally. Augmentation is a data pre-processing step and can be applied to any model. However, the resulting model need not be G-invariant, and for large groups it is computationally impractical to augment by a representative subset of the group.

As for intrinsic approaches, group equivariant neural networks like [25, 46, 81, 116] are model-specific, and there are unavoidable limits on their universality

while using low-order tensors [82]. Additionally, [25] requires the elements of G to be stored in memory, making it impractical for large groups.

The approach in [115] using polynomial invariants is not practical outside of small group actions owing, in part, to the need to compute a basis of polynomial invariants. This basis is, in general, large, increasing the dimension of the feature space dramatically. Another drawback is that the polynomials significantly distort the data geometrically, which we found leads to significant losses in accuracy for the machine learning model when running informal experiments.

Projecting onto a fundamental domain combines the benefits of being computationally easy to use, maintaining the original dimension and geometry of the data, and being compatible with any machine learning model. The resulting model is G-invariant almost everywhere (see Remark 7.4).

In principle, projecting to an isometric embedding of the quotient space resolves the issue raised in Remark 7.4 and it does not distort the data. However, it suffers from the same problems as the polynomial invariants method otherwise. Finding isometric embeddings is very difficult and significantly increases the ambient dimension. Therefore, this approach is not practical except for very special cases, such as when the action is particularly simple, or when the quotient space is equal to a fundamental domain.

7.3 Unifying intrinsic approaches to equivariant machine learning

We have focussed on approximating G-invariant functions by considering them as functions on a fundamental domain. In this Section we show that approximating a G-invariant function is equivalent to approximating it on the *quotient space*. We then pick up two approaches to G-invariant machine learning from the literature alongside our approach, and explain in which sense they can be viewed as machine learning on quotient spaces. It is convenient to treat the G-invariant problem as a special case of the G-equivariant problem. We return to the setting that G acts by isometries on X and Y which are Riemannian manifolds. To simplify the proof

of Theorem 7.11 we make the following stronger assumption.

Assumption 7.9. In this Section, assume that G acts properly discontinuously on X and Y.

The result of this is that the quotient space has the structure of a Riemannian orbifold.

7.3.1 Universality of quotient spaces

For us the key is that quotient spaces are *universal* with respect to G-equivariant maps.

Universal property of quotient spaces   Given two spaces X and Y and a group G acting on both, let $\pi_X \colon X \to X/G$ and $\pi_Y \colon Y \to Y/G$ be the canonical projection maps. Then for any G-equivariant map $\alpha \colon X \to Y$ there is a unique map $\overline{\alpha} \colon X/G \to Y/G$ such that $\pi_Y \circ \alpha = \overline{\alpha} \circ \pi_X$.

$$
\begin{array}{ccc}
X & \xrightarrow{\ \ \alpha\ \ } & Y \\
\downarrow{\scriptstyle \pi_X} & & \downarrow{\scriptstyle \pi_Y} \\
X/G & \dashrightarrow{\overline{\alpha}} & Y/G
\end{array}
$$

If the action of G on X and Y is sufficiently 'nice', then the converse holds: equivariant maps $X \to Y$ are parametrised by certain maps $X/G \to Y/G$. Notice that for any G-equivariant function α, point stabilisers in X and Y have the property that $\mathrm{Stab}_G(x) \subset \mathrm{Stab}_G(\alpha(x))$.

Definition 7.10. A continuous map $\overline{\alpha} : X/G \to Y/G$ is **compatible** with the G actions if

1. We have $(\overline{\alpha} \circ \pi_X)_*(\pi_1(X)) \leqslant (\pi_Y)_*(\pi_1(Y))$, and

2. For any $\overline{x} \in X/G$ let $x \in X$ be a lift of $\overline{x}$, and let $y \in Y$ be a lift of $\overline{\alpha}(\overline{x})$. Then the stabiliser $\mathrm{Stab}_G(x)$ is conjugate in G to a subgroup of $\mathrm{Stab}_G(y)$.

Here $(\pi_Y)_* : \pi_1(Y) \to \pi_1(Y/G)$ denotes the map on fundamental groups induced by the map π_Y, and similarly for $(\overline{\alpha} \circ \pi_X)_*$.

It follows from the observation above that, if $\overline{\alpha}$ is the map coming from the universal property, then $\overline{\alpha}$ is automatically compatible with the G-action. In the special case that G acts trivially on Y (ie the G-invariant case) then every continuous map $\overline{\alpha}\colon X/G \to Y/G = Y$ is compatible.

Theorem 7.11: *Any compatible function $\overline{\alpha}\colon X/G \to Y/G$ lifts to a G-equivariant function $\alpha\colon X \to Y$. Suppose α' is another such lift, and assume there is some $x_0 \in X$ such that $\alpha'(x_0) = \alpha(x_0)$ where $\mathrm{Stab}_G(\alpha(x_0))$ fixes Y point-wise. Then $\alpha'(x) = \alpha(x)$ for all $x \in X$.*

We omit the proof, which can be deduced using Theorem 4.1.6 in [23]. This gives a converse to the universal property, and shows that up to an isometry of Y there is a one-to-one correspondence between G equivariant maps $X \to Y$, and compatible maps of their quotients.

Now suppose $\alpha\colon X \to Y$ is a G-equivariant function we want to approximate using a supervised machine learning algorithm. Using an intrinsic approach, as discussed in Section 6.2, means approximating α by a function β which is *a priori* G-equivariant. By the Theorem above, this is equivalent to approximating $\overline{\alpha}$ by a compatible function $\overline{\beta}$.

7.3.2   Specific intrinsic approaches

We now discuss how different intrinsic approaches to the equivariant machine learning problem fit into this framework.

Equivariant maps from polynomial invariants   The relationship between the method in [115] and quotient spaces is shown by the following result.

Theorem 7.12 ([101]): *The map $p(x) \coloneqq (p_1(x), \ldots, p_k(x))$ factors through $\mathbb{R}^n/G$, and induces a smooth embedding of $\mathbb{R}^n/G$ into $\mathbb{R}^k$.*

We can reinterpret the invariant version of (6.2) as a (fully-connected) neural network $\overline{\beta}$. We then train this network on the data $D_{\text{train}}^p = \{(p(x), y) \mid (x, y) \in D_{\text{train}}\}$ which has been projected to the quotient space by p. Thus $\overline{\beta}$ learns the map $\overline{\alpha}$ directly.

Fundamental domain projections Our approach of projecting onto a fundamental domain fits very naturally in this general framework. We want to try to approximate the function $\overline{\alpha}$ rather than approximating α. Instead of working directly with the quotient spaces, one can think of the map from the fundamental domain to the quotient space $\pi_X|_{\mathcal{F}_X} : \mathcal{F}_X \to X/G$ as a *chart* in the sense of differential geometry, so $\mathcal{F}_X$ locally parametrises X/G, see Figure 7.2.
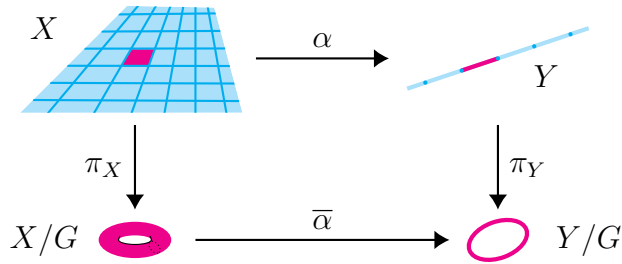
Figure 7.2: A fundamental domain can be thought of as a chart for the quotient space.

In the invariant case, we can approximate $\alpha = \overline{\alpha} \circ \pi_X|_{\mathcal{F}_X}$ by approximating $\overline{\alpha}$. In the equivariant case, we can also view $\pi_Y|_{\mathcal{F}_Y} : \mathcal{F}_Y \to Y/G$ as a chart, and because $\pi_Y|_{\mathcal{F}_Y}$ is a bijection onto its image we can apply its inverse and approximate $\alpha = (\pi_Y|_{\mathcal{F}_Y})^{-1} \circ \overline{\alpha} \circ \pi_X|_{\mathcal{F}_X}$ by approximating $\overline{\alpha}$. Note that $\pi_X|_{\mathcal{F}_X}$ and $\pi_Y|_{\mathcal{F}_Y}$ are not surjective in general, and there is no canonical way to extend their domain to make them so. The fix for this is to perturb points to lie in the preimage of $\mathcal{F}_X$ as discussed in Section 8.1.

Equivariant layers in neural networks On the face of it, the various approaches to equivariant neural networks such as [81, 25, 46] bypass the compatible map $\overline{\alpha}$ by approximation α directly, restricting the space of maps which can be used. However, as stated in Section 6.3, the main task when working with equivariant layers in neural networks is to compute the space of equivariant linear maps. Here we sketch an approach which is combinatorial and involves putting a G-invariant simplicial complex structure on $\mathbb{R}^{n_i}$ and applying the compatibility criterion Theorem 7.11 to the cells in the simplicial structure induced on the quotient space.

To simplify matters slightly for the purpose of exposition, assume that G acts on $\mathbb{R}^{n_1}$ and $\mathbb{R}^{n_2}$ discretely and irreducibly by orthogonal matrices on each space. Because the action is orthogonal (ie it fixes the origin and preserves the Euclidean

metric), it leaves the unit sphere $\mathbb{S}^{n_i-1}$ invariant in $\mathbb{R}^{n_i}$. The sphere is compact and the action is properly discontinuous, so it is possible to find some G-invariant triangulation of $\mathbb{S}^{n_i-1}$. Moreover, possibly after subdividing once, any two points in the interior of some k-simplex have the same stabiliser in G, and if some simplex τ is a face of another simplex τ', then $\mathrm{Stab}_G(\tau') \leqslant \mathrm{Stab}_G(\tau)$.

Now, the projection map $\pi_i \colon \mathbb{S}^{n_i-1} \to \mathbb{S}^{n_i-1}/G =: Q_i$ induces a simplicial structure on the quotient, and we can label each simplex $\overline{\tau}$ in the quotient by the set of stabilisers of all simplices $\tau \in \mathbb{S}^{n_i-1}$ which are mapped to $\overline{\tau}$, $\pi_i(\tau) = \overline{\tau}$. In fact the label of $\overline{\tau}$ defined in this way is exactly the conjugacy class of $\mathrm{Stab}_G(\tau)$ in G for some (equivalently any) τ such that $\pi_i(\tau) = \overline{\tau}$. We can now try to construct a compatible map $\overline{\lambda'} \colon Q_1 \to Q_2$ by mapping simplices $\overline{\tau}_1 \in Q_1$ to simplices $\overline{\tau}_2 \in Q_2$ such that some element of the label $\overline{\tau}_1$ is a subgroup of some element of the label of $\overline{\tau}_2$ (note $\overline{\tau}_1$ and $\overline{\tau}_2$ need not have the same dimension) so that these maps glue together in a continuous way. This reduces the problem of checking the compatibility criterion on every point in Q_1 to only checking it on a finite number of simplices, and checking that the maps glue together.

7.4 Applications to machine learning problems

In this Section we show how G-invariant pre-processing can be applied to examples of classification tasks in group theory, string theory, and image recognition. In each case, the symmetry group acts differently on the input space. We account for this by appropriately choosing different projection maps which are defined in detail in the next two Chapters. Experiments 7.4.1 and 7.4.3 are chosen as proof of concept, not to reach the state of the art, the main application of our approach is experiment 7.4.2. Implementation details may be found in [3].

7.4.1 Cayley tables

The following model problem was introduced in Section 3.2.3 of [62]: up to isomorphism, there are 5 groups with 8 elements. Separate their Cayley tables into two classes and apply random permutations until $20\,000$ tables in each class exist.

The problem is then to assign the correct one of two classes to a given table, and this map is invariant under the action of $S_8 \times S_8$ acting on $\mathbb{R}^8 \otimes \mathbb{R}^8$ by row and column permutations.

Let $\pi_\uparrow \colon \mathbb{R}^8 \otimes \mathbb{R}^8 \to \mathbb{R}^8 \otimes \mathbb{R}^8$ be the ascending projection map from Section 8.1, in particular as defined in Section 8.1.3. This has an explicit description as follows: given a choice of total order on the group elements, permute the columns so that the first row is ordered smallest to biggest and then permute the rows so that the first column is ordered smallest to biggest. Then, $\pi_\uparrow$ is invariant under the action of $S_8 \times S_8$ and can be efficiently computed for Cayley tables. This pre-processing effectively 'undoes' the permutations, which makes the machine learning problem trivial. Consequently, we achieve nearly perfect accuracy using a linear support vector machine (SVM), see Table 7.1.

We compare our approach with the neural network from [62], with the Deep Sets architecture from [116], and with the $S_8 \times S_8$-invariant neural network from [59]. The Deep Sets architecture is invariant under the action of the full $S_{8 \cdot 8} = S_{64}$ on $\mathbb{R}^8 \otimes \mathbb{R}^8$. As all Cayley tables are in the same orbit under this group action, the performance of this architecture can only be as good as random guessing. Note that the general purpose architectures described in [81, 46] in this case are identical to [59]. Other architectures from the literature, such as [25], are difficult to apply to this problem, since they require keeping a non-sparse map $S_8 \times S_8 \to \mathbb{R}$ in memory. This group has size $8! \cdot 8! \approx 1.6 \cdot 10^9$.

Table 7.1: Accuracy of predicting the group isomorphism type of a Cayley table.

	Accuracy
MLP [62]	0.501 ± 0.015
Deep Sets [116]	0.504 ± 0.010
G-inv MLP [59, 81, 46]	0.498 ± 0.012
$\pi_\uparrow$+SVM	$\mathbf{0.994 \pm 0.008}$

7.4.2 CICY

In [54], a dataset of complex three-dimensional *complete intersection Calabi-Yau* manifolds (CICYs) and their basic topological invariants is given. In [61], a neural network was used to predict (among other tasks) the first Hodge number of a given CICY. Here, CICYs are represented by matrices of size up to 12×15, and the first Hodge number is an integer. The same problem was subsequently studied in [18, 17, 43] using more sophisticated machine learning models. The problem is invariant under row and column permutations, ie an action of $S_{12} \times S_{15}$ on $\mathbb{R}^{12} \otimes \mathbb{R}^{15}$, but none of the machine learning models which have been implemented previously for the Hodge number classification satisfy this invariance.

We compare two pre-processing maps: the map $\pi_{\mathrm{Dir}} \colon \mathbb{R}^{12} \otimes \mathbb{R}^{15} \to \mathbb{R}^{12} \otimes \mathbb{R}^{15}$ defined in Section 9.1, which we computed by performing discrete gradient descent; and $\pi_\uparrow \colon \mathbb{R}^{12} \otimes \mathbb{R}^{15} \to \mathbb{R}^{12} \otimes \mathbb{R}^{15}$ defined in the same way as in Section 7.4.1. We found that composing π_{Dir} with existing neural networks slightly improves performance, but not significantly. We also considered an alternative training task in which input matrices first had their rows and columns randomly permuted. In this case, our model outperforms models from the literature by a large margin. We also compare our model with the group invariant model from [59] in both training tasks, see Table 7.2. Again, the approaches of [81] and [46] reduce to [59]. As for Cayley tables, the approach in [25] is impractical due to the large group size $12! \cdot 15! \approx 6 \cdot 10^{20}$.

As our approach is intrinsic, it is well suited for problems with a large symmetry group. For all networks but the G-invariant multi layer perceptron (MLP) the accuracy decreases on the permuted dataset. This suggests that the rows and columns of the CICY matrices are already systematically ordered in the original dataset. The map $\pi_\uparrow$ can be computed efficiently but need not be G-invariant on the boundary of the fundamental domain, see Remark 7.4. This is a potential problem since the input data, which consists of integer-valued matrices, is discrete. Indeed, a substantial proportion of the CICY matrices are very sparse and do lie on the boundary, which could be the reason why $\pi_\uparrow$ performs relatively poorly on

Table 7.2: Accuracies for the task of predicting the second Hodge number of a CICY matrix. Models are compared on the original training task and on randomly permuted input matrices. The last three rows are group invariant models, the first three rows are not group invariant models.

	Original dataset	Randomly permuted
MLP [61]	0.554 ± 0.015	0.395 ± 0.029
MLP+pre-processing [17]	0.858 ± 0.009	0.417 ± 0.086
Inception [43]	0.970 ± 0.009	0.844 ± 0.117
G-inv MLP [59, 81, 46]	0.895 ± 0.029	0.914 ± 0.023
π_{Dir}+Inception	$\mathbf{0.975 \pm 0.007}$	$\mathbf{0.963 \pm 0.016}$
$\pi_{\uparrow}$+Inception	0.969 ± 0.009	0.539 ± 0.020

the permuted data set. The projection map π_{Dir} can only be approximated but is fully G-invariant which is a crucial advantage on the permuted dataset.

7.4.3   Classifying rotated handwritten digits

As an instructional example, we use the MNIST dataset of 28×28 pixel images of handwritten digits from [73], on which $\mathbb{Z}_4$ acts by rotating the images by multiples of $\pi/2$. We use the ascending averaging combinatorial projection defined in Section 8.1.2, $\pi_{\uparrow\mathrm{av}} \colon \mathbb{R}^4 \otimes \mathbb{R}^{196} \to \mathbb{R}^4 \otimes \mathbb{R}^{196}$. The map $\pi_{\uparrow\mathrm{av}}$ rotates each image so that its brightest quadrant is the top-left quadrant. We also use a projection onto the quotient space $\pi_{\mathbb{Z}_4}$ defined in Section 9.3. We then compare performance of a linear classifier, a shallow neural network, and SimpNet (see [60]) which is among the top performers on the original MNIST task; first on their own, then with data augmentation, and finally with the projection map $\pi_{\uparrow\mathrm{av}}$, but without data augmentation, see Table 7.3.

For linear classifiers, data augmentation does not improve accuracy substantially. This is due to their small number of parameters. Unsurprisingly, pre-processing with $\pi_{\uparrow\mathrm{av}}$ improves performance, because it is partially successful at rotating digit pictures into a canonical orientation.

At first look the projection onto the quotient space $\pi_{\mathbb{Z}_4}$ seems to outperform all other methods, however it is difficult to draw conclusions from this. That is

Table 7.3: Accuracy for the task of recognising handwritten digits. We use two different degrees of data augmentation: either add every possible rotation of the input image to the training data (*Augmentation* $\times 4$) or applying data augmentation until the training data reaches 1.5 times its original size (*Augmentation* $\times 1.5$). We also compare two projections, $\pi_{\uparrow\text{av}}$ which maps onto a fundamental domain, and $\pi_{\mathbb{Z}_4}$ which projects onto the quotient space.

	No pre-processing	$\pi_{\uparrow\text{av}}$	$\pi_{\mathbb{Z}_4}{}^{\dagger}$
Linear	0.677 ± 0.001	0.784 ± 0.001	0.905 ± 0.001
MLP	0.939 ± 0.001	0.953 ± 0.003	0.967 ± 0.001
SimpNet [60]	0.979	0.979	
	Augmentation $\times 1.5$	Augmentation $\times 4$	
Linear	0.682 ± 0.001	0.682 ± 0.001	
MLP	0.963 ± 0.002	0.963 ± 0.001	
SimpNet [60]	0.986	0.986	

$\dagger$ In order to make $\pi_{\mathbb{Z}_4}$ practical to compute, we first downsized the images to be 8×8 pixels. Although not stated in the table, performing experiments with a linear classifier and MLP on this downsized data set, both with no pre-processing and with $\pi_{\uparrow\text{av}}$ yielded accuracies comparable to those shown for the full-sized images.

because the projection significantly increases the dimension of the input space, from $8^2 = 64$ (for downsized images) to 128. The effect of this is to greatly increase the number of learnable parameters for the linear map λ_0, which likely contributes to the improved performance.

For neural networks with more than one layer, data augmentation increases accuracy, because the model now has sufficient parameters to include the information from the additional training data. Pre-processing using $\pi_{\uparrow\text{av}}$ yields better accuracy than no pre-processing, but worse accuracy than full data augmentation. If fewer training data points are added during the data augmentation step, the benefit is comparable to applying the map $\pi_{\uparrow\text{av}}$.

This is one example of the fact that data augmentation may be the best pre-processing option if the symmetry group G has few elements and one can augment by the full group. If $\#G$ is very large, this is not possible, and pre-processing using a fundamental domain projection may be better than augmenting with a small, non-representative, subset of G.

Chapter 8

Combinatorial projection maps

IN THIS CHAPTER, WE GIVE an algorithm to compute a projection onto a fundamental domain in a very special setting. While of course this does not cover all machine learning applications by any means, in practice this is sufficient for many use cases.

Assumption 8.1. In this Chapter, assume that $G \leqslant S_n$ is a subgroup of the permutation group, which acts on $X = \mathbb{R}^n$ by permuting coordinates.

More precisely, if $x = (x_i)_i \in \mathbb{R}^n$ and $s \in S_n$ we say s acts on the left by $s \cdot (x_i)_i = \left(x_{s^{-1}(i)}\right)_i$. This induces an action of G on $\mathbb{R}^n$ which we call a *permutation action of G*. This action is discrete (since G is finite) and by isometries since it preserved the Euclidean metric.

In the next Section, we describe a method of finding a similar projection for any G. This method turns out to have a significant degree of flexibility in two senses: first in the initial choice of a *base* for $G \leqslant S_n$, so that projections onto several essentially different fundamental domains is possible for a given subgroup (ie the different domains are not merely translates of one another).

The second is how the projection is applied: each projection is based on permuting the entries of some point $x \in \mathbb{R}^n$ based on their relative size, but one can choose to 'prioritise' small values or large values, and whether the values of individual entries, or of collections of entries are used. This leads to definitions of ascending and descending projections, as well as averaging versions of both of

these. We also explicitly compute projection maps for several examples of groups.

In Section 8.2, we turn the procedure to compute the fundamental domain projection into an implementable algorithm and analyse the time and space complexity of this algorithm. We can compute $\pi(x)$ in $O(n^3(\log \log n)^2)$ time.

Finally, the last and longest Section of the Chapter contains the proof of Theorem 8.5 which states that the procedure we have described for finding π does indeed yield a projection onto a fundamental domain. The proof centres around a method given in [36] for computing a *right transversal* for $G \leqslant S_n$, ie a complete set of unique coset representatives for G in S_n, and the geometry of the action of S_n on $\mathbb{R}^n$. In particular, S_n is a Coxeter group of type A (see Table 1.1), and the action on $\mathbb{R}^n$ essentially coincides with the action on its Davis complex. Thus we can find a fundamental domain for S_n coming from a choice of fundamental chamber (see Definition 4.4), and show that π projects onto a fundamental domain for G which is a union of translates of this fundamental chamber.

8.1   Combinatorial projections

The algorithm to find a fundamental domain projection is based on [36] in which the authors give an efficient algorithm to find a set of unique coset representatives for an arbitrary subgroup $G \leqslant S_n$. A set of coset representatives can be turned into a set of orbit representatives for the permutation action of G on $\mathbb{R}^n$. We modify their algorithm so that this set of orbit representatives is in fact a fundamental domain, and so that it outputs an explicit projection map. This map is easy to implement and efficient to compute.

8.1.1   The ascending projection map

We break the procedure down into several steps.

Finding a base   Let $N = \{1, \ldots, n\}$, which we identify with the set of indices for the standard basis for $\mathbb{R}^n$. S_n acting on $\mathbb{R}^n$ corresponds to the right action of S_n on N by $i \cdot s = s^{-1}(i)$. The first step of the algorithm is to find a *base* for $G \leqslant S_n$.

Definition 8.2. A **base** for $G \leqslant S_n$ is an ordered subset $B = (b_1, \ldots, b_k)$ of N such that $\bigcap_{i=1}^{k} \text{Stab}_G(b_i) = \{1\}$, where $\text{Stab}_G(b_i)$ is the *stabiliser* of b_i in G. Given a base let $G_0 = G$ and for $1 \leqslant i \leqslant k$, define $G_i = \text{Stab}_{G_{i-1}}(b_i) = G_{i-1} \cap \text{Stab}_G(b_i)$.

It follows from this definition that $G_k = \{1\}$. One can always choose $B = (1, \ldots, n-1)$ as the base, although an algorithm to compute a more efficient base is given in Section 8.3.3.

Definition 8.3. Given a base B and the groups G_i, we also define Δ_i to be the orbit of b_i under the action of G_{i-1}, $\Delta_i := b_i \cdot G_{i-1}$.

Example 8.4. Let G be the subgroup in S_4 generated by the elements $(1\ 2)$ and $(3\ 4)$. Then $B = (1, 3)$ is a base and we have stabilisers $G_0 = \{e, (1\ 2), (3\ 4), (1\ 2)(3\ 4)\}$, $G_1 = \{e, (3\ 4)\}$, and $G_2 = \{e\}$; and orbits $\Delta_1 = \{1, 2\}$ and $\Delta_2 = \{3, 4\}$.

Perturbing points in $\mathbb{R}^n$ We now need to define the map $\phi_\uparrow \colon X \to G$ used in the definition of $\pi_\uparrow$. The map $\phi_\uparrow$ is only uniquely defined on points $x = (x_i)_i \in \mathbb{R}^n$ all of whose entries are distinct. We first perturb x slightly to get a point with this property. Choose a *perturbation vector ε* which has all distinct entries, for example $\varepsilon = \frac{1}{2n}(1, 2, \ldots, n)$. Let $d = \min_{x_i \neq x_j}\{|x_i - x_j|\}$ (choose $d = 1$ if all entries of x are the same) and define $x' = x + d\varepsilon$, which is guaranteed to have all entries distinct. The entries of x' have the same relative order, ie if $x_i' \leqslant x_j'$ then $x_i \leqslant x_j$, and $\phi_\uparrow$ depends only on this relative ordering of entries. Then we define $\phi_\uparrow(x) = \phi_\uparrow(x')$ where $\phi_\uparrow(x')$ is defined below.

The ascending projection map   We define a sequence of permutations $g_i \in G$ for $1 \leqslant i \leqslant k$ as follows. Assume $g_1, \ldots, g_{i-1}$ have already been found. G_{i-1} acts transitively on Δ_i, choose $j \in \Delta_i$ such that the jth entry of $(g_{i-1} \cdots g_1) \cdot x'$ is minimal among those entries indexed by Δ_i. Choose $g_i \in G_{i-1}$ such that $j \cdot g_i = g_i^{-1}(j) = b_i$. Now define $\phi_\uparrow(x') := g_k \cdots g_1$, note the choice of the g_i's is not unique, but we show in Section 8.3.6 that $\phi(x')$ *is* uniquely defined.

Section 8.3 is devoted to the proof of the following Theorem which says that the map we have defined is a projection onto a fundamental domain.

Theorem 8.5: *Define $\pi_\uparrow \colon \mathbb{R}^n \to \mathbb{R}^n$ by $\pi_\uparrow(x) = \phi_\uparrow(x) \cdot x$, and let $\mathcal{F}$ be the interior of its image. Then $\mathcal{F}$ is a fundamental domain for G acting on $\mathbb{R}^n$. Given a choice of base B and perturbation vector ε, the projection $\pi_\uparrow$ is uniquely defined.*

Example 8.6. For $G = S_n$, let $B = (1, 2, \ldots, n-1)$, so that $G_i = \mathrm{Sym}(i+1, \ldots, n)$ and $\Delta_i = \{i, \ldots, n\}$. Fixing $x' \in X$ and following the algorithm above: $g_1 \in G_0 = S_n$ is a permutation which moves the smallest entry indexed by $\Delta_1 = \{1, \ldots, n\}$ to the first entry indexed by $b_1 = 1$. Repeating this for each i up to $n-1$, g_i moves the ith smallest entry of x' to the ith position. The result is that $(g_{n-1} \cdots g_1) \cdot x'$ has its entries ordered from smallest to largest.

8.1.2 Other combinatorial projection maps

There are three natural variations of the combinatorial projection map $\pi_\uparrow$ we define above which may be more suited to specific applications. We called that projection an *ascending projection*. The variations are a *descending projection* $\pi_\downarrow$, and ascending and descending *averaging projections* $\pi_{\uparrow\mathrm{av}}$ and $\pi_{\downarrow\mathrm{av}}$. These projections each have their own version of Theorem 8.5 whose proof is essentially identical.

The descending projection is defined via $\phi_\downarrow$, which differs from $\phi_\uparrow$ only when we define g_i. In this case G_{i-1} acts transitively on Δ_i, and we choose $j \in \Delta_i$ such that the jth entry of $(g_{i-1} \cdots g_1) \cdot x'$ is *maximal* among those entries indexed by Δ_i. Choose $g_i \in G_{i-1}$ such that $j \cdot g_i = g_i^{-1}(j) = b_i$. If the input data for the machine learning algorithm consisted of vectors containing non-negative entries including many zeros, the descending projection in some sense *prioritises* the non-zero entries, so may yield different results.

For the averaging projections, assume that $G = H_1 \times H_2$ is a direct product of groups $H_j \leqslant S_{n_j}$ which acts on $\mathbb{R}^{n_1} \otimes \mathbb{R}^{n_2}$ by letting H_1 act on the first factor and H_2 act on the second. In this case, identify N with the set of pairs $\{(l, m) \mid 1 \leqslant l \leqslant n_1, \ 1 \leqslant m \leqslant n_2\}$. Define a transformation $\mu \colon \mathbb{R}^{n_1} \otimes \mathbb{R}^{n_2} \to \mathbb{R}^{n_1} \otimes \mathbb{R}^{n_2}$ by

$$\mu \colon (x_{\ell m})_{\ell m} \mapsto \left(\frac{1}{n_1}(x_{1m} + x_{2m} + \cdots + x_{n_1 m}) + \frac{1}{n_2}(x_{\ell 1} + x_{\ell 2} + \cdots + x_{\ell n_2}) \right)_{\ell m}.$$

Thinking of $(x_{\ell m})_{\ell m}$ as a matrix, notice this is a G-equivariant linear map which replaces each entry of $(x_{\ell m})_{\ell m}$ by the sum of the averages of the entries in its row

and column. Now for any $x \in \mathbb{R}^{n_1} \otimes \mathbb{R}^{n_2}$ we define

$$\phi_{\uparrow \mathrm{av}}(x) = \phi_{\uparrow}\left(\mu(x)'\right) \quad \text{and} \quad \phi_{\downarrow \mathrm{av}}(x) = \phi_{\downarrow}\left(\mu(x)'\right).$$

These definitions generalise to the case $G = \prod_{j=1}^{r} H_j$ acting on $\bigotimes_{j=1}^{r} \mathbb{R}^{n_j}$ component-wise, where $H_j \leqslant S_{n_j}$. One might wish to use an averaging projection if, for example, one of the H_j's is trivial, in which case a non-averaging projection ignores most of the entries, since they are not in any of the orbits Δ_i. This is the case in the application discussed in Section 7.4.3.

Example 8.7. Let $G = \mathbb{Z}_3 \times S_3 \leqslant S_3 \times S_3$ act on $\mathbb{R}^3 \otimes \mathbb{R}^3$, thought of as the set of 3×3 matrices, by cyclically permuting the rows and freely permuting the columns. In this case let $N = \{(\ell, m) \mid 1 \leqslant \ell \leqslant 3, \ 1 \leqslant m \leqslant 3\}$ and construct a base. Let $b_1 = (1, 1)$ whose stabiliser is $G_1 = \{1\} \times \mathrm{Sym}(\{2, 3\})$, and the orbit of b_1 under $G_0 = G$ is $\Delta_1 = N$. Now $(2, 1)$ and $(3, 1)$ are both fixed by G_1 and so should not be the next element of the base. Choose $b_2 = (1, 2)$. Then $G_2 = \{1\} \times \{1\}$ and the orbit of b_2 under G_1 is $\Delta_2 = \{(1, 2), (1, 3)\}$. Since $G_2 \cong \{1\}$ we are done and $B = ((1, 1), (1, 2))$.

Let $x' = (x'_{\ell m})_{\ell m}$ be a 3×3 matrix whose entries are distinct, we want to compute $\phi_{\uparrow}(x')$. Let $(p_1, q_1) \in \Delta_1 = N$ be the pair such that $x'_{p_1 q_1}$ is the minimal entry. Then we can choose $g_1 = (s_1, (1\ q_1)) \in \mathbb{Z}_3 \times S_3$ where

$$s_1 = \begin{cases} (1) & p_1 = 1 \\ (1\ 2\ 3) & p_1 = 2 \\ (1\ 3\ 2) & p_1 = 3 \end{cases} \in \mathbb{Z}_3$$

Now let $g_1 \cdot x' = (x''_{\ell m})_{\ell m}$, and let $(1, q_2) \in \Delta_2 = \{(1, 2), (1, 3)\}$ minimise $x''_{1 q_2}$. Define $g_2 = ((1), (2\ q_2)) \in G_1$ and $\phi_{\uparrow}(x') = g_2 g_1$.

Combinatorially we can describe the projection $\pi_{\uparrow}$ as follows: transport the smallest entry of x' to the top left corner by cyclically permuting rows and freely permuting columns. Then order columns 2 and 3 so that the entries in the first row increase.

As an example, consider the matrix x and perturbation matrix ε

$$x = \begin{pmatrix} 5 & 3 & 3 \\ 4 & 0 & 0 \\ 3 & 5 & 1 \end{pmatrix}, \ \varepsilon = \frac{1}{18} \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}.$$

Then

$$x' = x + \varepsilon = \frac{1}{18} \begin{pmatrix} 91 & 56 & 57 \\ 76 & 5 & 6 \\ 61 & 98 & 27 \end{pmatrix}.$$

We can now apply $\pi_\uparrow(x) = \phi_\uparrow(x') \cdot x$ in the two step process described above:

$$\frac{1}{18} \begin{pmatrix} 91 & 56 & 57 \\ 76 & 5 & 6 \\ 61 & 98 & 27 \end{pmatrix} \overset{g_1}{\longmapsto} \frac{1}{18} \begin{pmatrix} 5 & 76 & 6 \\ 98 & 61 & 27 \\ 56 & 91 & 57 \end{pmatrix} \overset{g_2}{\longmapsto} \frac{1}{18} \begin{pmatrix} 5 & 6 & 76 \\ 98 & 27 & 61 \\ 56 & 57 & 91 \end{pmatrix} = \pi_\uparrow(x')$$

$$x = \begin{pmatrix} 5 & 3 & 3 \\ 4 & 0 & 0 \\ 3 & 5 & 1 \end{pmatrix} \overset{g_1}{\longmapsto} \begin{pmatrix} 0 & 4 & 0 \\ 5 & 3 & 1 \\ 3 & 5 & 3 \end{pmatrix} \overset{g_2}{\longmapsto} \begin{pmatrix} 0 & 0 & 4 \\ 5 & 1 & 3 \\ 3 & 3 & 5 \end{pmatrix} = \pi_\uparrow(x).$$

Similarly

$$\pi_\downarrow(x) = \begin{pmatrix} 5 & 3 & 1 \\ 3 & 5 & 3 \\ 0 & 4 & 0 \end{pmatrix}.$$

We can also compute the averaging versions of these projections. Applying μ we get

$$\mu(x) = \frac{1}{3} \begin{pmatrix} 23 & 19 & 15 \\ 16 & 12 & 8 \\ 21 & 17 & 13 \end{pmatrix}.$$

Hence,

$$\pi_{\uparrow\mathrm{av}}(x) = \begin{pmatrix} 0 & 0 & 4 \\ 1 & 5 & 3 \\ 3 & 3 & 5 \end{pmatrix}, \text{ and } \pi_{\downarrow\mathrm{av}}(x) = \begin{pmatrix} 5 & 3 & 3 \\ 4 & 0 & 0 \\ 3 & 5 & 1 \end{pmatrix}.$$

8.1.3   Examples of combinatorial projection maps

In this Section we list combinatorial projection maps for several common examples of permutation groups $G \leqslant S_n$. Notice that in each of the four examples of concrete groups below, implementation via a suitable sorting function circumvents the need to perturb inputs initially.

The symmetric group   If $G = S_n$, let $N = \{1, \ldots, n\}$ and we can choose the base $B = (1, 2, \ldots, n-1)$. The ascending projection $\pi_\uparrow(x)$ permutes the entries so that they increase from left-to-right, and the descending projection $\pi_\downarrow(x)$ permutes the entries so that they decrease.

The alternating group   If $G = A_n < S_n$ is the group of even permutations, we can choose $B = (1, 2, \ldots, n-2)$ and the ascending (respectively descending) projection permutes the entries of x so that the first $n-2$ entries increase (respectively decrease) from left-to-right, and the last two entries are greater than or equal to all the other entries. If x contains repeated entries then the last to entries can also be ordered to be increasing (respectively decreasing); otherwise their relative order depends on whether the permutation which maps $i \mapsto \#\{1 \leqslant j \leqslant n \mid x_j \leqslant x_i\}$ for $1 \leqslant i \leqslant n$, is an even or odd permutation.

The cyclic group   If $G = \mathbb{Z}_n \leqslant S_n$ is the cyclic group generated by the permutation $(1\ 2\ \cdots\ n)$, we can choose the base $B = (1)$. The ascending (respectively descending) projection cyclically permutes the entries of x so that the first entry is less (respectively greater) than or equal to all other entries of x.

The dihedral group   If $G = \mathrm{Dih}_n \leqslant S_n$ is the dihedral group generated by

$$s_1 = (1\ 2\ \cdots\ n) \quad \text{and} \quad s_2 = (2\ n)(3\ (n-1))(4\ (n-2))\cdots,$$

we can choose base $B = (1, 2)$. The ascending (respectively descending) projection cyclically permutes the entries of x via s_1 so that the first entry is less (respectively greater) than or equal to all other entries of x, and then if the final entry is less (respectively greater) than the second entry, it applies the permutation s_2.

Products of groups acting on products of spaces   Suppose $G = \prod_{j=1}^{r} H_j$ where $H_j \leqslant S_{n_j}$ acts on $\bigoplus_{j=1}^{r} \mathbb{R}^{n_j}$ by each H_j acting by permutations on the corresponding space $\mathbb{R}^{n_j}$ and trivially everywhere else. Let $B_j = \left(b_j^{(1)}, \ldots, b_j^{(k_j)} \right) \subset \{1, \ldots, n_j\} = N_j$ be a base for H_j acting on $\mathbb{R}^{n_j}$, then

$$B = \left(b_1^{(1)}, \ldots, b_1^{(k_1)}, b_2^{(1)}, \ldots, b_2^{(k_2)}, \ldots, b_r^{(1)}, \ldots, b_r^{(k_r)} \right)$$

is a base for G. Let $\pi_{j\uparrow} \colon \mathbb{R}^{n_j} \to \mathbb{R}^{n_j}$ be the ascending projection corresponding to B_j. Then define $\pi_\uparrow = \bigoplus_{j=1}^{r} \pi_{j\uparrow}$, to be the projection which equals $\pi_{j\uparrow}$ when restricted to $\mathbb{R}^{n_j}$. Similarly $\pi_\downarrow = \bigoplus_{j=1}^{r} \pi_{j\downarrow}$.

Products of groups acting on tensors of spaces   Suppose $G = \prod_{j=1}^{r} H_j$ where $H_j \leqslant S_{n_j}$ acts on $\bigotimes_{j=1}^{r} \mathbb{R}^{n_j}$ by each H_j acting by permutations on the jth component of $\bigotimes_{j=1}^{r} \mathbb{R}^{n_j}$, and trivially on the other components. For each $1 \leqslant j \leqslant r$ let $B_j = \left(b_j^{(1)}, \ldots, b_j^{(k_j)} \right) \subset \{1, \ldots, n_j\} = N_j$ be a base for H_j acting on $\mathbb{R}^{n_j}$, and furthermore (for convenience) assume that $b_j^{(1)} = 1$. Then choose $B \subset \prod_{j=1}^{r} N_j =: N$ to be

$$B = \left((1, \ldots, 1), \; \left(b_1^{(2)}, 1, \ldots, 1 \right), \ldots, \left(b_1^{(k_1)}, 1, \ldots, 1 \right), \right.$$

$$\vdots \qquad\qquad\qquad \vdots$$

$$\left. \left(1, \ldots, 1, b_r^{(2)} \right), \ldots, \left(1, \ldots, 1, b_r^{(k_r)} \right) \right),$$

where a 1 in the jth position of an element of B should be thought of as $b_j^{(1)}$. Suppose $x = (x_{\ell_1 \cdots \ell_r})_{\ell_1 \cdots \ell_r} \in \bigotimes_{j=1}^{r} \mathbb{R}^{n_j}$, and let x' be defined as in Section 8.1. Choose $(m_1, \ldots, m_r) \in N$ to be the index in the G-orbit of $\mathbb{1} = (1, \ldots, 1)$ with minimal entry in x'. For $1 \leqslant j \leqslant r$ define $x'_j := (x'_{m_1 \cdots \ell_j \cdots m_r})_{1 \leqslant \ell_j \leqslant n_j} \in \mathbb{R}^{n_j}$, which is the restriction of x' to the $\mathbb{R}^{n_j}$-vector containing the entry $x'_{m_1 \cdots m_j \cdots m_r}$. Then define

$$\phi_\uparrow \colon \bigotimes_{j=1}^{r} \mathbb{R}^{n_j} \to G \colon x \mapsto (\phi_{1\uparrow}(x'_1), \ldots, \phi_{r\uparrow}(x'_r)),$$

where $\phi_{j\uparrow} \colon \mathbb{R}^{n_j} \to H_j$ is the function defined for H_j acting on $\mathbb{R}^{n_j}$, and similarly define $\phi_\downarrow(x)$. Then as before, $\pi_\uparrow(x) := \phi_\uparrow(x) \cdot x$ and $\pi_\downarrow(x) := \phi_\downarrow(x) \cdot x$.

8.2   Algorithm to compute combinatorial projections

In this Section we give algorithms to compute combinatorial projection maps for permutation group actions and analyse the time and space complexity of these algorithms. These work for any permutation group G, although they are not the ones we used in Section 7.4 which employed more efficient *ad hoc* methods described in Section 8.1.3. The general algorithms here fall into two parts: first are the algorithms which are applied as a one-off to compute data like a base and the orbits Δ_i; and which run in $O(k^2 n^3)$ time, and $O(n^2 \log n)$ space, where n is the dimension of the input space and k is the size of the base. Second are the algorithms which actually implement the projection $\pi_\uparrow$ and which must therefore be run for

each input datum. They do this in $O(k^2 n^2)$ time and $O(n^2 \log n)$ space. Since $\pi_\uparrow$ merely permutes the entries of a datum, it does not change the space required to store the input data.

Throughout we maintain the same notation as before, where we are working with a subgroup of S_n which acts by permuting the coordinates of $\mathbb{R}^n$ indexed by $N = \{1, \ldots, n\}$. As initial data we assume we have a subgroup G of S_n given by a generating set of permutations. Moreover, we assume that these permutations are given in *cycle* or *one-line* notation, so that each can be stored in $O(n \log n)$ space, and multiplying two permutations together can be performed in $O(n)$ time. Similarly, given $x \in \mathbb{R}^n$ and a permutation g, the point $g \cdot x$ can be computed in $O(n)$ time.

8.2.1 Computing initial data

We make use of the method of representing permutation groups introduced by Mark Jerrum in [67], which we summarise. First we explain the notation. We work with directed simple graphs which have N as their vertex set. If their directed edge set is E, we write the graph as a pair (N, E). We use $\ell - 1\ell_2$ to denote an edge which starts at ℓ_1 and ends at ℓ_2.

Definition 8.8. A **(directed) path** in (N, E) is a sequence of vertices $\ell_0 \ell_1 \cdots \ell_m$ such that $\ell_j \ell_{j+1} \in E$ for each j. Such a path is said to have **length** $m \geqslant 0$.

A directed graph is called a **branching** if it contains no paths of length $m \geqslant 1$ with the same start and end points, and if each vertex has at most one incoming edge. If (N, E) is a directed graph, an **edge labelling** is a map $\omega \colon E \to S_n \colon b_\ell b_m \mapsto \omega_{\ell m}$ which assigns to each edge a permutation of N. This labelling extends to a labelling of paths by setting $\omega_P = \omega_{b_{\ell_0} b_{\ell_1}} \cdots \omega_{\ell_{m-1} \ell_m} \in S_n$, where $P = \ell_0 \ell_1 \cdots \ell_m$.

Definition 8.9. Let $G \leqslant S_n$ be a permutation group. A **Jerrum representation** of G is an edge labelled directed graph $\Upsilon(G) = (N, E, \omega)$ satisfying the following properties:

1. Υ is a branching
2. For all $b_\ell b_m \in E$

    (a) $\ell < m$ and $\ell \leqslant k$

    (b) $\omega_{b_\ell b_m} \in G_{\ell-1}$

    (c) $b_\ell \cdot \omega_{b_\ell b_m} = b_m$

3. The set $U_i := \{\omega_P \mid P$ is a path in Υ starting at $b_i\}$ is a right transversal for
G_i in G_{i-1} for each 1

   $ellei \leqslant k.$

THEOREM 8.10 (Theorem 3.3 and Section 4 of [67])**:** *Let $G \leqslant S_n$ be a permutation group given by a set of generators, then there is an algorithm which yields a small base $B = (b_1, \ldots, b_k)$ for G (see Section 8.2.3 for a quantitative discussion of what a small base is) together with a Jerrum representation. This algorithm runs in in $O(k^2 n^3)$ time and $O(n^2 \log n)$ space. It also computes the orbits $\Delta_i = b_i \cdot G_{i-1}$ for $1 \leqslant i \leqslant k$.*

The algorithm presented in [67] in fact assumes that $(1, 2, \cdots, n)$ has been chosen *a priori* to be the base, but it is straightforward to amend the algorithm to compute a more efficient base using the greedy algorithm mentioned at the start of Section 8.3.3, compare with [11].

8.2.2 Applying $\pi_\uparrow$ to input data

Fix a permutation group G and let $\Upsilon = \Upsilon(G)$ be a Jerrum representation for G. First we prove a useful characterisation of the orbit Δ_{i+1}.

LEMMA 8.11: *The orbit Δ_i is the set of $b_\ell \in N$ such that there exists a path P in Υ which starts at b_i and ends at b_l.*

Proof. Note that by induction on m, for any path $P = b_{\ell_0} \cdots b_{\ell_m}$, (2) in Theorem 8.10 generalises to say

  (a) $\ell_0 < \ell_m$ and $\ell_{m-1} \leqslant k$

  (b) $\omega_P \in G_{\ell_0-1}$

  (c) $b_{\ell_0} \cdot \omega_P = b_{\ell_m}.$

If P starts at b_i, then $\omega_P \in G_{i-1}$ and $b_i \cdot \omega_P = b_\ell$ so $b_\ell i$. Conversely, if $b_\ell \in \Delta_i$, there is some $g \in G_{i-1}$ such that $b_i \cdot g = b_l$. Note that the cosets of G_i in G_{i-1} are exactly the sets of the form $\{g \in G_{i-1} \mid b_i \cdot g = b\}$ for fixed $b \in \Delta_i$. Indeed $g, g' \in G_{i-1}$ are in the same coset of G_i if and only if $b_i \cdot g' = b_i \cdot g$. Since U_i is a complete set of representatives, it contains an element from every coset, and hence an element which maps b_i to b_ℓ. Call this element u_{b_ℓ}, then by definition there is some path P which starts at b_i such that $u_{b_\ell} = \omega_P$, and by the observation above, the end point of P must be b_ℓ. $\qquad\square$

With this Lemma we can give an algorithm to perform the main task in computing $\pi_\uparrow$, computing $\phi_\uparrow$ as a product of permutations $g_i \in G_{i-1}$.

PROPOSITION 8.12: *Given $x' \in \mathbb{R}^n$, all of whose entries are distinct, and a Jerrum representation $\Upsilon = \Upsilon(G)$ for G, there is an algorithm to compute $\phi_\uparrow(x')$ in $O(k^2 n^2)$ time and $O(n^2 \log n)$ space.*

> **Data:** A point $x' \in \mathbb{R}^n$, a Jerrum representation $\Upsilon = \Upsilon(G)$, and the orbits Δ_i.
> **Result:** $\phi_\uparrow(x')$.
> 1 **for** $1 \leqslant i \leqslant k$ **do**   // Loop runs k times
> 3 $\quad$ Set j to be the index in Δ_i such that $x'_j \leqslant x'_\ell$ for all $\ell \in \Delta_i$; // This is
> $\quad\quad$ the current working vertex in Υ
> 4 $\quad$ Set $g_i = e$;   // This accumulates edge labels from Υ
> 6 $\quad$ **while** $j \neq b_i$ **do**   // Loop runs at most $|\Delta_i|$ times
> 8 $\quad\quad$ Set ℓ to be the unique index in Δ_i such that ℓj is an edge of Υ;
> 10 $\quad\quad$ Set $g_i = \omega_{\ell j} g_i$;
> 11 $\quad\quad$ Set $j = \ell$;
> 12 $\quad$ **end**
> 14 $\quad$ Set $x' = g_i \cdot x'$;
> 15 **end**
> 17 Set $\phi_\uparrow(x') = g_k \cdots g_1$;

Algorithm 1: Computing $\phi_\uparrow(x')$ given a Jerrum representation Υ and the orbits Δ_i.

Proof. We use Algorithm 1. Recall the definition of $\phi_\uparrow(x')$. Assume $g_1, \ldots, g_{i-1}$ have already been found, G_{i-1} acts transitively on Δ_i, choose $j \in \Delta_i$ such that the jth entry of $(g_{i-1} \cdots g_1) \cdot x'$ is minimal among those entries indexed by Δ_i. Choose $g_i \in G_{i-1}$ such that $j \cdot g_i = g_i^{-1}(j) = b_i$. Then define $\phi_\uparrow(x') := g_k \cdots g_1$.

The job of finding each g_i is made easy by Lemma 8.11, since we just need to find a path P in Υ joining b_i to j, which is guaranteed to exist. Since Υ is branching, each vertex has at most one incoming edge, and hence starting from j and working backwards we are guaranteed to reach b_i. Making note of each edge label as we construct P, we can choose $g_i = \omega_P$. This is achieved by the loop starting in Line 6.

We consider the time complexity of this algorithm. Following [67], Υ can be represented by an $n \times n$ array, whose pq-entry is NULL if pq is not an edge of Υ, and ω_{pq} otherwise. Finding ℓ in Line 8 requires searching the jth column of this array for the unique non-NULL entry, whose index row index is ℓ, and so takes $O(n)$ steps. As mentioned above Line 10 also takes $O(n)$ steps, so the while loop at Line 6 takes $O(n|\Delta_i|)$ steps.

Finding j in Line 3 requires $O(|\Delta_i|)$ steps, searching through each entry of x' indexed by Δ_i and comparing it with the current minimal entry found; while Line 14 takes $O(n)$ steps. Thus the dominant step in the main for loop is the while loop. Overall then this for loop takes $O(kn \sum_{i=1}^{k} |\Delta_i|)$ steps, which is greater than the $O(kn)$ steps to compute Line 17. Noting that $|\Delta_i| \leqslant n$, this algorithm runs in $O(k^2 n^2)$ time.

As for space, Υ is an n^2 array, containing at most $n - 1$ non-NULL entries (since the graph is has no cycles, its number of edges is bounded by $n-1$), each of which takes $O(n \log n)$ space to store. An efficient encoding can then use $O(n^2 \log n)$ space. The other significant space cost is storing the g_i's. This takes $O(kn \log n)$ which is at most $O(n^2 \log n)$. □

THEOREM 8.13: *Given $x \in \mathbb{R}^n$, a perturbation vector ε, and a Jerrum representation Υ for G, there is an algorithm to compute $\pi_{\uparrow}(x)$ in $O(k^2 n^2)$ time and $O(n^2 \log n)$ space.*

Proof. Algorithm 2 follows exactly the procedure outlined in Section 8.1. Line 8 dominates in terms of both time and space complexity. □

8.2.3   Final time complexity analysis

Naïvely one may assume that the size k of the base B is $O(n)$, indeed this is the case for $G = S_n$ or A_n for example, but in practice we can do a lot better. Write

Data: A point $x \in \mathbb{R}^n$, a perturbation vector ε, a Jerrum representation
$\quad\quad \Upsilon = \Upsilon(G)$, and the orbits Δ_i.
Result: $\pi_\uparrow(x)$.

1 **if** $x \in \mathbb{R}\mathbb{1}$ **then**
2 $\quad$ Set $d = 1$;
3 **else**
4 $\quad$ Set $d = \min\{|x_i - x_j| \mid x_i \neq x_j\}$;
5 **end**
6 Set $x' = x + d\varepsilon$;
8 Compute $\phi_\uparrow(x')$;   `// Algorithm 1`
9 Set $\pi_\uparrow(x) = \phi_\uparrow(x') \cdot x$;

Algorithm 2: Computing $\pi_\uparrow(x)$ given a perturbation vector and initial data.

$b(G)$ for the size of the smallest base for G, then in [11] Blaha showed that the greedy algorithm used in Theorem 8.10 to find a base yields a base whose size k is $O(b(G) \log \log n)$.

When looking at permutation groups, it is natural to focus on the case of so-called *primitive* permutation groups, ie subgroups $G \leqslant S_n$ which act transitively on $N = \{1, \ldots, n\}$ such that there are no non-trivial G-invariant partitions. This is because arbitrary permutation groups can be built up out of primitive ones. In this setting Liebeck proved the following in [74].

THEOREM 8.14: *Let G be primitive and not S_n or A_n, then there is some absolute constant c such that $b(G) < c\sqrt{n}$.*

It follows that in this case, the base found above has size $O(\sqrt{n} \log \log n)$. Combining this with the observation in Section 8.1.3 that for S_n and A_n, $\pi_\uparrow$ can be computed using a sorting algorithm, we get the following.

THEOREM 8.15: *Let $G \leqslant S_n$ be primitive, then either*

- $G = S_n$ *or* A_n: *no initial data needs to be computed and $\pi_\uparrow$ can be computed in $O(n^2)$ time per datum (using worst case for quicksort); or*

- *Initial data can be computed in $O(n^4 (\log \log n)^2)$ time, and $\pi_\uparrow$ can be computed in $O(n^3 (\log \log n)^2)$ time per datum.*

8.3   Proof of Theorem 8.5

The idea of the proof is as follows. In Section 8.3.1, we outline an equivalence be-
tween subgroups of S_n acting on $\mathbb{R}^n$ by permuting coordinates, and them acting on
S_n by multiplication. This provides a dictionary between certain combinatorially
defined fundamental domains and sets of coset representatives satisfying simple
algebraic properties, see Proposition 8.22. We then outline the work from [36] in
Section 8.3.3 which gives an algorithm to find a set of coset representatives for an
arbitrary subgroup of S_n. The main work is then to show this algorithm, with
modifications, can produce a set of coset representatives with the desired alge-
braic properties so that it corresponds to a fundamental domain. This culminates
in Corollary 8.31. Finally we show in Proposition 8.32 that the algorithm outlined
in Section 8.1 indeed produces a projection onto this fundamental domain.

8.3.1   Actions on $\mathbb{R}^n$ and S_n

Recall we have the group S_n acting on $\mathbb{R}^n$ on the left by $s \cdot (x_i)_i = \left(x_{s^{-1}(i)}\right)_i$. We also
have the normal action of S_n on itself on the left by group multiplication. Here
we show that in some sense these actions are equivalent.

Let $x \in \mathbb{R}^n$ be a point, all of whose entries are distinct, and notice the set of
such points is open and dense in $\mathbb{R}^n$. Define a function which changes the ith
entry x_i of x to the integer $\#\{1 \leqslant j \leqslant n \mid x_j \leqslant x_i\}$. The result is a list of the
integers $1, \ldots, n$ in the same relative order as the entries of x, and we denote the
set of all such points C. We can think of C as a discrete subset of $\mathbb{R}^n$, and the
left action of S_n on $\mathbb{R}^n$ restricts to a left action on C. Notice also that this map
$\mathbb{R}^n_{\mathrm{dist}} := \{x \in \mathbb{R}^n \mid \text{all entries are distinct}\} \to C$ is continuous. In other words the
set of connected components of $\mathbb{R}^n_{\mathrm{dist}}$ are in one-to-one correspondence with C.
Indeed each component contains a point in C, its *representative point*.

Definition 8.16. We call these connected components *chambers*; given $c \in C$ we
write $[c] \subset \mathbb{R}^n$ for the corresponding chamber.

The following is straightforward to check.
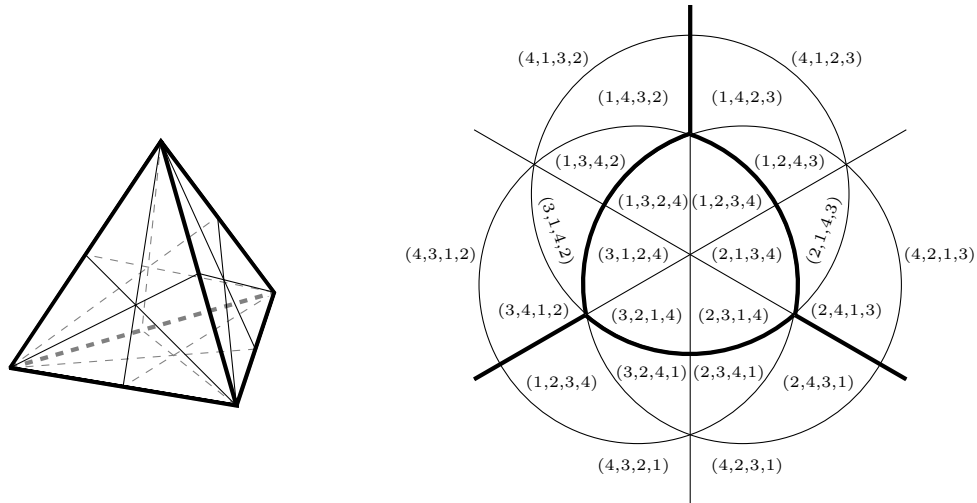
Figure 8.1: On the left is boundary of a 3-simplex, each small triangle corresponds to the intersection of this with a chamber. On the right the picture has been stereographically projected to the plane for the purposes of illustration, and each chamber is labelled by the representative element of C.

LEMMA 8.17: *Each chamber is a fundamental domain for the action of S_n on $\mathbb{R}^n$.*

The action of S_n on $\mathbb{R}^n$ preserves an $(n-1)$-simplex in the orthogonal complement of the vector $(1, \ldots, 1)$. In Figure 8.1 we show the 3-simplex preserved by S_4 and use it to visualise the $24 = \#S_4$ chambers in this case.

On the other hand, we can view each element of C as a permutation in S_n written in *in-line notation*. This means if $c = (c_i)_i$, as a permutation it sends i to c_i for each $i \in \{1, \ldots, n\}$. Thus S_n is in one-to-one correspondence with C. In fact, it is better in our situation to modify this correspondence by inverting elements of S_n via the map $\rho \colon S_n \to C \colon s \mapsto (s^{-1}(i))_i$. The equivalence of the left action of S_n on $\mathbb{R}^n$ and the left action on itself comes in the following form. Let $s, t \in S_n$, and consider the action of s on $\rho(t)$:

$$s \cdot \rho(t) = s \cdot (t^{-1}(i))_i = (t^{-1}(s^{-1}(i)))_i = ((st)^{-1}(i))_i = \rho(st) = \rho(s \cdot t).$$

Given any subgroup $G \leqslant S_n$, the map ρ defines an equivalence between G acting on $\mathbb{R}^n$, which restricts to an action of G on C, and G acting on S_n by left multiplication. We can use this equivalence to convert a set of right coset representatives for G in S_n into a complete set of orbit representatives for G acting on $\mathbb{R}^n$.

PROPOSITION 8.18: *Let R be a set of right coset representatives for $G \leqslant S_n$, then $\overline{\mathcal{F}} = \bigcup_{r \in R} \overline{[\rho(r)]}$ is a complete set of orbit representatives for G acting on $\mathbb{R}^n$, where $\overline{[\rho(r)]}$ is the closure of the chamber containing $\rho(r)$.*

Proof. Since $\mathbb{R}^n_{\mathrm{dist}}$ is dense in $\mathbb{R}^n$ and G acts by continuous maps which leave $\mathbb{R}^n_{\mathrm{dist}}$ invariant as a set, it suffices to show that $\bigcup_{r \in R} [\rho(r)]$ is a complete set of orbit representatives for G acting on $\mathbb{R}^n_{\mathrm{dist}}$. In fact G simply permutes the components of $\mathbb{R}^n_{\mathrm{dist}}$ so it suffices to show that $\bigcup_{r \in R} \rho(r)$ is a complete set of orbit representatives for the induced action of G on C.

But now, ρ is a bijection which exhibits an equivalence between the action of G on C and the action of G on S_n so we just need to show that R is a complete set of orbit representatives for G acting on S_n. The orbits of this action are precisely the right cosets of G, which completes the proof. $\square$

8.3.2   Gallery connectedness and fundamental domains

Given a set of right coset representatives R for $G \leqslant S_n$, the interior of $\overline{\mathcal{F}}$ as defined in the Proposition is not, in general, a fundamental domain because it is not connected. We can reinterpret connectedness in terms of algebraic properties of R. First some geometric definitions.

Definition 8.19. Let $c, c' \in C$ be distinct, we say the chambers $[c]$ and $[c']$ are **adjacent** if $\overline{[c]} \cap \overline{[c']}$ has codimension 1. A **gallery** is a sequence of chambers $[c_1], \ldots, [c_k]$ such that consecutive chambers are adjacent. A set of chambers is called **gallery connected** if any two distinct chambers in the set can be connected by a gallery which is completely contained in the set. As a shorthand, we sometimes call a subset $C' \subset C$ gallery connected if the set $\{[c] \mid c \in C'\}$ is gallery connected.

It turns out that the decomposition of $\mathbb{R}^n_{\mathrm{dist}}$ into chambers corresponds to the chamber system of S_n acting on its Coxeter complex, about which we do not elaborate here, but the interested reader should consult [16]. The upshot of this viewpoint is two characterisations of adjacency of chambers.

LEMMA 8.20: *Let $c \neq c' \in C$ and define $s = \rho^{-1}(c)$, $s' = \rho^{-1}(c')$. Then the following are equivalent:*

1. *The chambers $[c]$ and $[c']$ are adjacent.*

2. *There is $1 \leqslant j \leqslant n - 1$ such that $s' = s(j\ j+1)$, where $(j\ j+1)$ is a transposition in S_n.*

3. *The vectors c and c' differ by swapping exactly two entries which are consecutive integers.*

Proof. The equivalence of (1) and (2) is proved in Theorem I.5A of [16]. To see the equivalence of (2) and (3), notice that $\rho(s(j\ j+1)) = ((s(j\ j+1))^{-1}(i))_i =: (c'_i)_i$. For $i \notin \{j, j+1\}$, $c'_i = s(i) = c_i$ (where $c := (c_i)_i$), whereas $c'_j = c_{j+1}$ and $c'_{j+1} = c_j$. $\square$

The equivalence of (1) and (3) for the example of S_4 can be seen in (Figure 8.1). We use this characterisation to prove Proposition 8.30 which is key to showing that the image of $\pi_\uparrow$ is connected. We are now in a position to upgrade Proposition 8.18 so that it produces a fundamental domain for the action of G.

Definition 8.21. We define a **right transversal** of $G \leqslant S_n$ to be a minimal set of right coset representatives (ie a set containing exactly one element from every right coset).

PROPOSITION 8.22: *Let $R \subset S_n$ be a right transversal for $G \leqslant S_n$ such that $\rho(R)$ is gallery connected. Then $\mathcal{F}$, the interior of $\bigcup_{r \in R} \overline{[\rho(r)]}$, is a fundamental domain for G acting on $\mathbb{R}^n$.*

Proof. By the definition, if $[c]$ and $[c']$ are adjacent, then the interior of $\overline{[c]} \cup \overline{[c']}$ is connected. By induction on the length of galleries in $\{[\rho(r)] \mid r \in R\}$ it follows that $\mathcal{F}$ is connected. It is also open by definition.

By Proposition 8.18 we know that $\overline{\mathcal{F}}$ is a complete set of orbit representatives for G. Finally, suppose that some G-orbit meets $\mathcal{F}$ in at least two points, say x and x', and $g \in G$ is such that $g \cdot x = x'$. Since the G-action permutes the chambers, there are two possibilities:

1. There are coset representatives $r, r' \in R$ such that $x \in [\rho(r)]$ and $x' \in [\rho(r')]$.

2. There are coset representatives $r_1 \neq r_2, r'_1 \neq r'_2 \in R$ such that $x \in \overline{[\rho(r_1)]} \cap \overline{[\rho(r_2)]}$ and $x' \in \overline{[\rho(r'_1)]} \cap \overline{[\rho(r'_2)]}$.

In the first case we must have that $g \cdot [\rho(r)] = [\rho(r')]$, in which case it follows from Lemma 8.17 and the fact that $g \neq 1$, that $r \neq r'$. But then by the equivalence of the action with the action on S_n, we have that $g \cdot r = gr = r'$ and r and r' represent the same right coset of G. This contradicts the assumption that R is minimal. In the second case, we can similarly argue that $\{r_1, r_2\} \neq \{r_1', r_2'\}$ but $g \cdot \{r_1, r_2\} = \{r_1', r_2'\}$, again contradicting the minimality of R. In either case g cannot exist. $\qquad\square$

8.3.3 An algorithm to find coset representatives

In this Section we summarise the main construction of [36] which gives an efficient algorithm to compute a right transversal for an arbitrary subgroup $G \leqslant S_n$. The first step is to find a base $B \subset N$ for $G \leqslant S_n$. Set $B_0 = ()$, the empty tuple. We assume that we have already constructed B_{i-1} and computed G_{i-1}. If $G_{i-1} = \{1\}$, $B = B_{i-1}$ is a base and we are done. Otherwise, pick $b_i \in N$ with the largest orbit under G_{i-1} and let B_i be B_{i-1} with b_i appended.

Let $B = (b_1, \ldots, b_k)$ be a base and recall we define $G_0 = G$ and $G_i = G_{i-1} \cap \mathrm{Stab}_G(b_i)$ for $1 \leqslant i \leqslant k$. We also write $\Delta_i = b_i \cdot G_{i-1}$ for the orbit of b_i under G_{i-1}. Recursively construct a partition Π_i of N, starting with $\Pi_0 = \{N\}$. Denote by Γ_i the element of Π_{i-1} which contains b_i. One can check by induction that Γ_i contains Δ_i as a subset. Define Π_i by replacing Γ_i in Π_{i-1} by the non-empty subsets from the list: $\{b_i\}$, $\Delta_i - \{b_i\}$, and $\Gamma_i - \Delta_i$.

Now let U_i be a right transversal for the group $\mathrm{Sym}(\Delta_i) \times \mathrm{Sym}(\Gamma_i - \Delta_i)$ in $\mathrm{Sym}(\Gamma_i)$, where $\mathrm{Sym}(\Omega)$ is the group of permutations of the set Ω (in the next section we fix a particular choice for U_i), and finally let

$$H_i = \prod_{\Gamma \in \Pi_i} \mathrm{Sym}(\Gamma).$$

Then define $R = H_k U_k U_{k-1} \cdots U_1$, where for subsets $A, B \subset S_n$, $AB := \{ab \mid a \in A, \ b \in B\}$.

Theorem 8.23 ([36] §4): *The set R is a right transversal for $G \leqslant S_n$.*

8.3.4 Gallery connected sets of coset representatives

We now show how the method described above can be used to construct a right transversal R for G such that $\rho(R)$ is gallery connected. This is done by choosing a suitable base B, possibly re-indexing the set N, and choosing appropriate right transversals U_i for $\text{Sym}(\Delta_i) \times \text{Sym}(\Gamma_i - \Delta_i)$ in $\text{Sym}(\Gamma_i)$. We prove Theorem 8.5 assuming B and N have been chosen in this way, and then in Section 8.3.6 show that the assumptions on B and N can be dropped.

We described how to find a base for G by appending more and more elements of N to $B = ()$ until $G_k = \{1\}$ in Section 8.3.3. The first assumption we make is as follows.

Assumption 8.24. Until the final proof of Theorem 8.5 in Section 8.3.6, assume that each new b_i is minimal in the orbit $b_i \cdot G_{i-1}$ with respect to the normal ordering on N. We call such a base *orbit minimal*.

We use the following Lemma to build gallery connected sets out of other gallery connected sets.

LEMMA 8.25: *Let $A_1, \ldots, A_\ell$ be subsets of S_n so that each contains the identity permutation (1), and $\rho(A_i)$ is gallery connected for each i. Then $\rho(A_1 A_2 \cdots A_\ell)$ is gallery connected.*

Proof. Notice that $A_1 A_2$ contains $(1)(1) = (1)$. Let $a \in A_1 A_2$, and choose $a_1 \in A_1$ and $a_2 \in A_2$ such that $a = a_1 a_2$. Since $\rho(A_1)$ and $\rho(A_2)$ are gallery connected, there are galleries $p_{a_1} \subset \rho(A_1)$ and $p_{a_2} \subset \rho(A_2)$ which connect $\rho((1))$ to $\rho(a_1)$ and $\rho((1))$ to $\rho(a_2)$ respectively. Then $a_1 \cdot p_{a_2}$ connects $\rho(a_1)$ to $\rho(a_1 a_2)$ in $\rho(a_1 A_2)$, and the concatenation $p_{a_1} * (a_1 \cdot p_{a_2})$ is a gallery which connects $\rho((1))$ to $\rho(a)$ in $\rho(A_1 A_2)$. Call this gallery $\tilde{p}_a$, its construction is illustrated in Figure 8.2. Now given a, a' in A_1, A_2 the gallery $\tilde{p}_a^{-1} \cup \tilde{p}_{a'}$ (where $\tilde{p}_a^{-1}$ indicates $\tilde{p}_a$ traversed in reverse) connects $\rho(a)$ to $\rho(a')$ in $\rho(A_1 A_2)$, so $\rho(A_1 A_2)$ is gallery connected—the claim follows by induction on ℓ. $\square$

From the definition of R in the previous
section, if we show that H_k and each of the
U_i's satisfy the hypotheses of this Lemma,
then it follows that R is gallery connected.
We first consider H_k. Notice that in fact, if
$\Pi_k = \{N_1, \ldots, N_\ell\}$ then

$$H_k = \prod_{i=1}^{\ell} \mathrm{Sym}(N_i)$$
$$= \mathrm{Sym}(N_1)\mathrm{Sym}(N_2)\cdots\mathrm{Sym}(N_\ell)$$
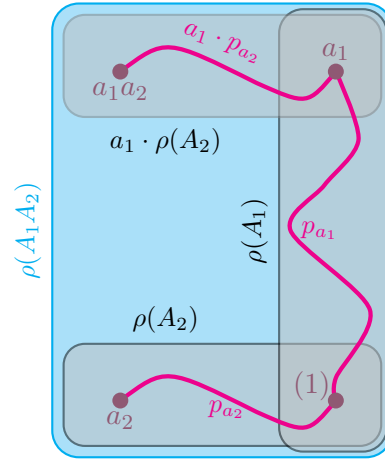
Figure 8.2: Building a gallery in
$\rho(A_1 A_2)$.

can be written as a product of sets, again as in the Lemma. Each $\mathrm{Sym}(N_i)$ con-
tains (1), so we just need $\rho(\mathrm{Sym}(N_i))$ to be gallery connected for each i. It follows
immediately from Lemma 8.20 that this is the case if and only if N_i is a sequence
of consecutive digits from N.

In general this is not the case, however it can be readily achieved by re-indexing
the set N. In fact we can do this so that each part of each partition Π_i is a set of
consecutive digits. This aids in showing that $\rho(U_i)$ is gallery connected.

LEMMA 8.26: *We can re-index N so that b_i remains minimal in Δ_i and each part of Π_i is
a set of consecutive digits for $1 \leqslant i \leqslant k$.*

Proof. We do induction on i: note that in $\Pi_0 = \{N\}$ the only part is a set of consec-
utive numbers. Assume that each element of Π_{i-1} is a set of consecutive digits, in
particular $\Gamma_i \in \Pi_{i-1}$ is a set of consecutive digits. Assume one of the three subsets
$\{b_i\}$, $\Delta_i - \{b_i\}$ or $\Gamma_i - \Delta_i$ is non-empty and does not consist of consecutive digits,
then by the minimality of b_i both $\Delta_i - \{b_i\}$ and $\Gamma_i - \Delta_i$ must be non-empty and
not consist of consecutive digits. Re-index the elements of Γ_i so that overall the
same set of digits is used, but now b_i is the smallest, the next smallest digits are
all in $\Delta_i - \{b_i\}$, and the remaining digits are in $\Gamma_i - \Delta_i$. $\qquad\square$

Assumption 8.27. Until the final proof of Theorem 8.5 in Section 8.3.6, assume
that N is indexed such that each part of the partition Π_i for $1 \leqslant i \leqslant k$ is a set of
consecutive numbers.

8.3.5   Choosing a right transversal U_i

Finally we want to choose the right transversals U_i for $\mathrm{Sym}(\Delta_i) \times \mathrm{Sym}(\Gamma_i - \Delta_i)$ in $\mathrm{Sym}(\Gamma_i)$. Write $\Delta_i = \{d_1, d_2, \ldots, d_m\}$, and $\Gamma_i - \Delta_i = \{d_{m+1}, d_{m+2}, \ldots, d_{m+m'}\}$. For $0 \leqslant \ell \leqslant \min\{m, m'\}$, choose $d_{j_1} < d_{j_2} < \cdots < d_{j_\ell}$ and $d_{m+j'_1} < d_{m+j'_2} < \cdots < d_{m+j'_\ell}$, and consider the product of transpositions

$$(d_{j_1} \ d_{m+j'_1})(d_{j_2} \ d_{m+j'_2}) \cdots (d_{j_\ell} \ d_{m+j'_\ell}) \in \mathrm{Sym}(\Gamma_i). \tag{8.1}$$

Define $\widetilde{U}_i$ to be the set of all such products for any choice of ℓ, and indices j_k and j'_k.

LEMMA 8.28: *([36] Lemma 2) $\widetilde{U}_i$ is a right transversal for $\mathrm{Sym}(\Delta_i) \times \mathrm{Sym}(\Gamma_i - \Delta_i)$ in $\mathrm{Sym}(\Gamma_i)$.*

Let $\tilde{u} \in \widetilde{U}_i$ have the form given in (8.1). If $\ell = 0$, then $\tilde{u}$ is the identity and thinking of it as an element of $S_n \geqslant \mathrm{Sym}(\Gamma_i)$, we get $\rho((1)) = (1, \ldots, n)$. More generally $\rho(\tilde{u})$ is the result of swapping each of the pairs $d_{j_k} \leftrightarrow d_{m+j'_k}$ in this vector, for $1 \leqslant k \leqslant \ell$. Let $g_{\tilde{u}} \in \mathrm{Sym}(\Delta_i) \times \mathrm{Sym}(\Gamma_i - \Delta_i)$ be the permutation such that $\rho(g_{\tilde{u}} \cdot \tilde{u})$ has its first m entries in increasing order, and its last m' entries in increasing order. Define $U_i = \{g_{\tilde{u}} \cdot \tilde{u} \mid \tilde{u} \in \widetilde{U}_i\}$.

LEMMA 8.29: *U_i is a right transversal for $\mathrm{Sym}(\Delta_i) \times \mathrm{Sym}(\Gamma_i - \Delta_i)$ in $\mathrm{Sym}(\Gamma_i)$.*

Proof. $\widetilde{U}_i$ contains exactly one element from each right coset of $\mathrm{Sym}(\Delta_i) \times \mathrm{Sym}(\Gamma_i - \Delta_i)$ in $\mathrm{Sym}(\Gamma_i)$. For $\tilde{u} \in \widetilde{U}_i$, the element $g_{\tilde{u}} \cdot \tilde{u} = g_{\tilde{u}} \tilde{u}$ lies in the same right coset as $\tilde{u}$ since $g_{\tilde{u}} \in \mathrm{Sym}(\Delta_i) \times \mathrm{Sym}(\Gamma_i - \Delta_i)$. Hence U_i contains exactly one element from each right coset of $\mathrm{Sym}(\Delta_i) \times \mathrm{Sym}(\Gamma_i - \Delta_i)$ in $\mathrm{Sym}(\Gamma_i)$. □

We want to show that $\rho(U_i)$ is gallery connected, and for that we use the re-indexing of N provided by Lemma 8.26. Recall that b_i is the ith element of the

base B; it follows from our construction of U_i that

$$
\rho(U_i) = \{(1,\dots,b_i - 1, \overbrace{c_1,\dots,c_m}^{\text{indexed by } \Delta_i}, \overbrace{c_{m+1},\dots,c_{m+m'}}^{\text{indexed by } \Gamma_i - \Delta_i}, b_i + m + m',\dots,n) \mid
$$

$$
c_j \in \Gamma_i = \{b_i, b_i + 1, \dots, b_i + m + m' - 1\} \text{ for all } 1 \leqslant j \leqslant m + m',
$$

$$
c_1 < \cdots < c_m, \text{ and } c_{m+1} < \cdots < c_{m+m'}\} \tag{8.2}
$$

Since this is notationally rather cumbersome, we abbreviate elements of $\rho(U_i)$ by

$$
(c_1, \dots, c_m \mid c_{m+1}, \dots, c_{m+m'}),
$$

where the first 'half' consists of entries indexed by Δ_i, and the second 'half' consists of entries indexed by $\Gamma_i - \Delta_i$.

PROPOSITION 8.30: *Let $G \leqslant S_n$, B be an orbit minimal base, and N indexed so that each part of Π_i is a set of consecutive digits. Then the set $\rho(U_i)$ is gallery connected.*

Proof. To help simplify notation, we do not distinguish between points in C and the chambers they represent. We show that $\rho(U_i)$ is gallery connected by explicitly constructing a gallery which joins an arbitrary chamber

$$
c = (c_1, \dots, c_m \mid c_{m+1}, \dots, c_{m+m'}) \in \rho(U_i)
$$

to the chamber corresponding to the identity in U_i,

$$
\hat{c} = \rho((1)) = (b_i, \dots, b_i + m - 1 \mid b_i + m, \dots, b_i + m + m' - 1).
$$

Lemma 8.20 gives the condition for consecutive chambers in this gallery to be adjacent: they must differ by swapping two entries which are consecutive integers. Furthermore, we ensure this gallery remains in $\rho(U_i)$ throughout. This implies that after swapping the two entries, the two halves of c must remain properly ordered. Taken together, this implies that the only swaps we can perform must switch the position of an entry in the left half with one in the right half, and these entries must be consecutive integers.

Let $c \in \rho(U_i)$ be arbitrary, write $\hat{c}_j = b_i + j - 1$ for the jth entry of $\hat{c}$, and define

$$
\delta(c) = \left(\sum_{j=1}^{m} c_j - \hat{c}_j \right) - \left(\sum_{j=m+1}^{m+m'} c_j - \hat{c}_j \right)
$$

which measures the degree to which c and $\hat{c}$ differ.

Claim 1 $\delta(c) \geqslant 0$.

Let $j \leqslant m$, then since the entries in the left half of c are ordered, distinct integers greater than or equal to b_i, $c_j \geqslant b_i + (j-1) = \hat{c}_j$ so each term in the first sum is non-negative. Similarly, for $j > m$ the entries in the right half of c are ordered, distinct integers less than or equal to $b_i + m + m' - 1$, so $c_j \leqslant b_i + m + m' - 1 - (m + m' - j) = b_i + (j-1) = \hat{c}_j$ so each term in the second sum is non-positive. $\square$

As a remark, it follows from this claim that δ equals the L^1 distance between c and $\hat{c}$. We perform a sequence of swaps as described above which have the effect decreasing the value of $\delta(c)$. Since $\delta(c) = 0$ implies that $c = \hat{c}$, the required gallery can be constructed by induction on $\delta(c)$. Assume $c \neq \hat{c}$, and let j be the minimal index such that $c_j \neq \hat{c}_j$. Since the two halves of c are ordered, c_j is in the left half.

Claim 2 $c_{j'} := c_j - 1$ is in the right half of c.

Indeed suppose it is in the left half, then by the ordering on c, $j' < j$, and by the minimality of j, $c_{j'} = \hat{c}_{j'} = b_i + j' - 1$. But then

$$\hat{c}_j \neq c_j = c_{j'} + 1 = b_i + (j' + 1) - 1 = \hat{c}_{j'+1},$$

so $j \neq j' + 1$ since all entries of $\hat{c}$ are distinct. But now $c_{j'} < c_{j'+1} < c_j$ (by the ordering on c), which is contradiction since these entries are distinct integers, and $c_{j'}$ and c_j differ by 1. $\square$

Thus, c_j and $c_j - 1$ are entries in different halves of c which are consecutive integers. Let c' be the result of swapping these two entries in c, then

$$\delta(c) - \delta(c') = ((c_j - \hat{c}_j) - (c_{j'} - \hat{c}_{j'})) - ((c_{j'} - \hat{c}_j) - (c_j - \hat{c}_{j'}))$$

$$= 2(c_j - c_{j'}) = 2 > 0$$

so performing the swap strictly decreases δ. By induction, there is a gallery in $\rho(U_i)$ joining c and $\hat{c}$, and hence $\rho(U_i)$ is gallery connected. $\square$

It follows directly from this Proposition, Theorem 8.23, and Proposition 8.22 that R as defined in Section 8.3.3 corresponds to a fundamental domain.

Corollary 8.31: *Let $G \leqslant S_n$, B be an orbit minimal base, and N indexed so that each part of Π_i is a set of consecutive digits. Let R be the right transversal for G constructed above. Then $\mathcal{F}$, the interior of $\bigcup_{r \in R} \overline{[\rho(r)]}$, is a fundamental domain for G acting on $\mathbb{R}^n$.*

8.3.6 Completing the proof

We have two things to do to complete the proof of Theorem 8.5: first show that the map $\pi_\uparrow$ as defined in Section 8.1 has image in $\overline{\mathcal{F}} = \bigcup_{r \in R} \overline{[\rho(r)]}$, and therefore indeed projects onto a fundamental domain, and then remove the assumptions of orbit minimality and on how N is indexed.

Proposition 8.32: *Let $G \leqslant S_n$, B be an orbit minimal base, and N indexed so that each part of Π_i is a set of consecutive digits. Then the image of $\pi_\uparrow$ lies in $\bigcup_{r \in R} \overline{[\rho(r)]}$.*

Proof. It suffices to show that the image of $\mathbb{R}^n_{\mathrm{dist}}$ lies in $\overline{\mathcal{F}} = \bigcup_{r \in R} \overline{[\rho(r)]}$. We claim that

$$\rho(R) = \{(c_j)_j \in C \mid \text{for } 1 \leqslant i \leqslant k, \ c_{b_i} \leqslant c_j \text{ for all } j \in \Delta_i\}.$$

The definition of $\pi_\uparrow$ implies that the right hand side of this is the image of $\pi_\uparrow|_C$, so the Proposition follows immediately from this claim.

Call the set on the right hand side C', first we show that $\rho(R) \subseteq C'$. By (8.2) (note that the entries of $(c_j)_j$ there are indexed differently there) we can see

$$\rho(U_i) \subset \{(c_j)_j \in C \mid c_{b_i} \leqslant c_j \text{ for all } j \in \Delta_i\}.$$

Since $U_i \subset \mathrm{Sym}(\Gamma_i)$, which fixes b_{i-1} for $i \geqslant 2$, one can inductively check from the definition of ρ that $\rho(U_k \cdots U_1) \subset C'$. Similarly, in the partition Π_k, each b_i appears as a singleton, so H_k also fixes b_i for $1 \leqslant i \leqslant k$, hence $\rho(R) \subseteq C'$.

To establish the claim we just need to show that $\#C' = \#\rho(R)$, since they are finite sets this implies that they are equal as sets. On the one hand, since ρ is a bijection, and using Lagrange's Theorem

$$\#\rho(R) = \#R = \#\{\text{right cosets of } G \text{ in } S_n\} = \#S_n / \#G.$$

On the other hand, each condition '$c_{b_i} \leqslant c_j$ for all $j \in \Delta_i$' decreases the size of C by a factor of $\#\Delta_i$, so

$$\#C' = \frac{\#C}{\#\Delta_1 \cdots \#\Delta_k}.$$

Since C is the bijective image of S_n under ρ, $\#C = \#S_n$. By the Orbit-Stabiliser Theorem, we also have

$$\#\Delta_i = \#b_i \cdot G_{i-1} = \#G_{i-1}/\#\mathrm{Stab}_{G_{i-1}}(b_i) = \#G_{i-1}/\#G_i,$$

Where the last equality follows from the definition $G_i = \mathrm{Stab}_{G_{i-1}}(b_i)$. Therefore

$$\#\Delta_1 \cdots \#\Delta_k = \frac{\#G_0}{\#G_1}\frac{\#G_1}{\#G_2} \cdots \frac{\#G_{k-1}}{\#G_k} = \frac{\#G_0}{\#G_k} = \frac{\#G}{\#\{1\}} = \#G$$

Hence $\#C' = \#\rho(R)$, which completes the proof. $\square$

Proof of Theorem 8.5. Let $N = \{1, \ldots, n\}$, and choose B a base for $G \leqslant S_n$, and ε satisfying the conditions in Section 8.1. Let $s \in S_n$ be a permutation of N such that $B^s := B \cdot s$ is an orbit minimal base, and each part of each partition $\Pi_i^s := \Pi_i \cdot s$ is a set of consecutive digits. That s exists can be seen by first permuting k times so that B is orbit minimal (note $b_i \notin \Delta_j$ for all $j > i$) and then applying Lemma 8.26. Write $b_i^s = b_i \cdot s$ so that $B^s = (b_1^s, \ldots, b_k^s)$.

Let $G^s = s^{-1}Gs$ be the conjugate of G by s in S_n, then for any $g \in G$ and $m \in N$,

$$(m \cdot s) \cdot g^s = ((g^s)^{-1}s^{-1})(m) = (s^{-1}g^{-1})(m) = (m \cdot g) \cdot s \qquad (8.3)$$

where $g^s = s^{-1}gs$. In other words, permuting by s and then acting by G^s is the same as acting by G and then permuting by s. It follows that $G_i^s := s^{-1}G_i s = G_{i-1}^s \cap \mathrm{Stab}_{G^s}(b_i^s)$, and $\Delta_i^s := \Delta_i \cdot s = b_i^s \cdot G_{i-1}^s$.

Finally define $\phi_\uparrow^s$ and $\pi_\uparrow^s$ as in Section 8.1 with respect to B^s and ε. We claim that for x' as defined in Section 8.1, $\phi_\uparrow^s(x') = (\phi_\uparrow(x'))^s = g_{x'}^s$. Indeed by definition $\phi_\uparrow^s(x') = \tilde{g}_k \cdots \tilde{g}_1$ where $\tilde{g}_i \in G_i^s$ such that $\tilde{\jmath} \cdot \tilde{g}_i = b_i^s$ and $\tilde{\jmath} \in \Delta_i^s$ is chosen such that the $\tilde{\jmath}$th entry of $(\tilde{g}_{i-1} \cdots \tilde{g}_1) \cdot x'$ is minimal among those entries indexed by Δ_i^s. But now $\Delta_i^s = \Delta_i \cdot s$ means $\tilde{\jmath} = j \cdot s$ (where $j \in \Delta_i$ is the index found in the definition of $\phi_\uparrow$). Thus

$$b_i^s = b_i \cdot s = (j \cdot g_i) \cdot s \overset{(8.3)}{=} (j \cdot s) \cdot g_i^s = \tilde{\jmath} \cdot g_i^s,$$

so we can certainly choose $\tilde{g}_i = g_i^s$. Then as claimed

$$\phi_\uparrow^s(x') = \tilde{g}_k \cdots \tilde{g}_1 = g_k^s \cdots g_1^s = (g_k \cdots g_1)^s = g_{x'}^s = (\phi_\uparrow(x'))^s.$$

Expanding out $\phi^s(x') = s^{-1}\phi(x')s$, we can now compute $\pi_\uparrow^s$ in terms of $\pi_\uparrow$ and s:

$$\pi_\uparrow^s(x) = \phi_\uparrow^s(x') \cdot x = s \cdot (\phi_\uparrow(x') \cdot (s^{-1} \cdot x)) = s \cdot \pi_\uparrow(s^{-1} \cdot x).$$

Writing $\mathcal{F}$ for the interior of the image of $\pi_\uparrow$, and $\mathcal{F}^s$ for the interior of the image of $\pi_\uparrow^s$, this implies $\mathcal{F}^s = s \cdot \mathcal{F}$ (because $s^{-1} \cdot \mathbb{R}^n = \mathbb{R}^n$). But Corollary 8.31, together with Proposition 8.32, says that $\mathcal{F}^s$ is a fundamental domain for G^s and $\pi_\uparrow^s$ is a projection onto $\mathcal{F}^s$; so $\mathcal{F} = s^{-1} \cdot \mathcal{F}^s$ is a fundamental domain for G and $\pi_\uparrow$ is a projection onto $\mathcal{F}$.

To prove the final claim of the Theorem, that $\pi_\uparrow\colon \mathbb{R}^n \to \mathbb{R}^n$ is uniquely defined by the choice of B and ε, we just need to show that a different choice of the elements $g_1, \ldots, g_k$ given $x \in \mathbb{R}^n$ does not change $\phi_\uparrow$. In fact $\phi_\uparrow$ is determined completely by what it does to the points $x' \in \mathbb{R}^n_{\text{dist}}$, and $\phi_\uparrow(x')$ lies inside the fundamental domain (not on its boundary). By the definition of a fundamental domain, any different choice $g_1', \ldots, g_k'$ must necessarily combine to give the same element $g_{x'}$ (no non-trivial element of G acts trivially), and hence $\phi_\uparrow$ is uniquely determined. $\qquad\square$

Chapter 9

Other projection maps

THE COMBINATORIALLY DEFINED PROJECTIONS DISCUSSED in the previous Chapter are useful as far as they go, but are only defined for finite permutation group actions. For other group actions, one could apply *ad hoc* methods to define an analogous combinatorial projection. However in the first Section of this Chapter we give a method for finding a projection onto a fundamental domain for any group acting discretely by isometries on a connected Riemannian manifold. This uses the idea of a *Dirichlet* fundamental domain, and can be implemented using a discrete version of gradient descent on the Cayley graph of the group which acts.

We then go on to discuss in more detail an alternative approach to projecting onto a fundamental domain: projecting onto the quotient space. We explicitly compute a (nearly) isometric embedding of the quotient space for $\mathbb{Z}_4$ acting on $\mathbb{R}^4 \otimes \mathbb{R}^n$ by cyclically permuting the coordinates in the first factor.

In Section 9.4, we give a quantitative way to measure the extent to which various pre-processing approaches to group invariant machine learning distort the input data and use this to compare these methods. In the final Section we offer some directions in which our approach can be generalised, say to actions by Lie groups.

9.1   Dirichlet projections

This method of computing a projection map follows from a classical proof of the existence of a fundamental domain for a sufficiently nice action of a group G by isometries on a metric space (X, d) (for example, a discrete action on a Riemannian manifold). The idea is to fix some point x_0 which is only fixed by the kernel of the action of G, and then define as follows

Definition 9.1. Let $\overline{\mathcal{F}}$ be

$$\{x \in X \mid d(x, x_0) \leqslant d(g \cdot x, x_0) \text{ for all } g \in G\}.$$

see for example Section II.1.4 of [110]. Such an $\mathcal{F}$ is called an *Dirichlet fundamental domain*. By its definition, we can rephrase the problem of finding a projection $\pi_{\mathrm{Dir}} \colon X \to \overline{\mathcal{F}}$ as a minimisation problem for the metric on X: given $x \in X$ find $g \in G$ which minimises $d(g \cdot x, x_0)$. In practice, this can be approximated using a discrete gradient descent algorithm.

Let us focus on the special case that G acts on $\mathbb{R}^n$ by orthogonal matrices. Then the inner product $\langle \cdot, \cdot \rangle$ is invariant. It is efficient to compute $\langle \cdot, \cdot \rangle$, which varies *inversely* with the Euclidean distance d between points, so we can perform gradient descent to minimize

$$\langle g \cdot x, -x_0 \rangle = \frac{1}{2} \left(d(g \cdot x, x_0)^2 - |g \cdot x|^2 - |x_0|^2 \right). \tag{9.1}$$

Choose a point $x_0 \in \mathbb{R}^n$ whose stabiliser is in the kernel of the G action. The map ϕ maps $x \in \mathbb{R}^n$ to the element in G which minimises the Euclidean distance $d(g \cdot x, x_0)$. Our main application of the discrete gradient descent algorithm is for CICY matrices when $G = S_{12} \times S_{15}$. Since $|G| \approx 6 \times 10^{20}$ one cannot minimise a function over G by simply evaluating it at all elements of G.

It is natural to compute the minimiser of (9.1) on a group orbit using gradient descent. The steps in the descent are restricted to the discrete G-orbit of the input point x, so we must define what a gradient is in this case. Taking a generating set

T for G, two points $x, x' \in \mathbb{R}^n$ are *adjacent* with respect to T if there is $t \in T$ such that $x' = t \cdot x$, so in particular, adjacent points are in the same G-orbit.

Definition 9.2. Given an action of a finite group G on $\mathbb{R}^n$, a generating set T of G, a function $N \colon \mathbb{R}^n \to \mathbb{R}$ and $x \in \mathbb{R}^n$, the **discrete gradient descent** is an approximation for

$$\min_{g \in G} \langle g \cdot x, -x_0 \rangle$$

and is defined iteratively as follows. Let $x_1 = x$. Given x_i, define

$$x_{i+1} = \min_{t \in T \cup \{e\}} \langle t \cdot x_i, -x_0 \rangle.$$

The output of the algorithm is x_i when $x_{i+1} = x_i$.

Since G acts discretely, this algorithm always terminates. In general, there are many choices for a generating set T resulting in different approximations for ϕ. For $G = S_n$ a natural choice for a generating set is given by $T = \{(1\ 2), (2\ 3), \ldots, (n - 1\ n)\}$. In particular, one has in this case $\#T = n - 1 \ll n! = \#S_n$. By taking the union of these generating sets for S_n and S_m one obtains a generating set of size $n + m - 2$ for $S_n \times S_m$.

Choosing a larger generating set increases the computational cost of the algorithm but potentially also its accuracy. For example, consider the set $T' = \{tt' \mid t, t' \in T \cup \{e\}\}$. This is a generating set for S_n and again yields a generating set for $S_n \times S_m$ in a similar way. When applied to the CICY dataset, we find that choosing T' instead of T leads to a significant increase in computation cost, but not so in accuracy.

Instead, we use discrete gradient descent starting with different seeds. For a 12×15 CICY matrix x, The seeds are $x_{km} := C_{12}{}^k x C_{15}{}^m$, where C_i is a cyclic permutation matrix, $1 \le k \le 12$, and $1 \le m \le 15$. To each x_{km}, apply the discrete gradient descent algorithm above and pick the minimum over all seeds. This increases the computation cost by a constant factor $11 \times 14 + 1 = 155$ but has led to a significant accuracy boost.

We are unable to give a bound for the number of generators applied to an input until a local minimum is reached. Experiments on the CICY dataset show that this number is very low compared to the size of the group. On the original CICY

dataset the average number of iterations is ≈ 17.4, with standard deviation ≈ 15.6 and maximum 163. On an augmented dataset, which contains 10 permutations of each matrix, the average number of iterations is ≈ 22.5, with standard deviation ≈ 16.9 and maximum 198.

9.2   Quotient space projections

One potential problem with projecting onto a fundamental domain $\pi : \mathbb{R}^n \to \overline{\mathcal{F}}$ is that this map is in general not strictly G-invariant, see Remark 7.4. Every point is mapped to a point in its G-orbit, however G-orbits which intersect the boundary $\partial \mathcal{F}$ may do so in several points and π does not necessarily pick out a unique one of these.

For many applications this turns out not to be a problem, because generically the training data lies in the preimage $\pi^{-1}(\mathcal{F})$, on which π *is* G-invariant. There are cases, however, where it might cause a significant issue, if most of the training data lies in the preimage $\pi^{-1}(\partial \mathcal{F})$. This is that case for our CICY example in Section 7.4.2 where the input data is typically a sparse integer matrix.

The way to deal with this mathematically is to project to the quotient space instead. Since G acts by isometries, the quotient space inherits the metric from $\mathbb{R}^n$, although it has the structure of an *orbifold* rather than a manifold. In order to run a machine learning algorithm with inputs in $\mathbb{R}^n/G$, we need to map it to a vector space $\mathbb{R}^n/G \hookrightarrow \mathbb{R}^k$ for some k, so that this map preserves the metric on the quotient. We write $\pi : \mathbb{R}^n \to \mathbb{R}^n/G \to \mathbb{R}^k$ for the composition of these maps, and call the result a *projection onto the quotient space*. The key point is that π is now truly G-invariant. Ideally, π should be an *isometric embedding*, but it is not always practical to find a map which is injective or isometric, so instead we content ourselves with finding a map which locally approximates an isometry and does not identify 'too many' G-orbits. We call such maps *locally near-isometric projections*.

We can now proceed in the same way as we did when π was a projection onto a fundamental domain, and train any machine learning model $\overline{\beta} : \mathbb{R}^k \to \mathbb{R}^m$ on the data D_{train}^{π}.

Finding an explicit projection onto the quotient space, π, for an arbitrary group G is an extremely difficult problem. In very special cases like Coxeter group actions where one has a *strict* fundamental domain, a projection onto a fundamental domain turns out to be a projection onto the quotient space. In the next Section we give an explicit computation in a relatively simple case.

9.3 An embedding of the quotient space for $\mathbb{Z}_4$ acting on $\mathbb{R}^4 \otimes \mathbb{R}^n$

In Section 7.4.3 we can think of $\mathbb{Z}_4$ acting on a square image by rotations as $\mathbb{Z}_4 \times \{(1)\}$ acting on $\mathbb{R}^4 \otimes \mathbb{R}^n$, where $\mathbb{Z}_4$ is generated by the permutation $s = (1\ 2\ 3\ 4)$, and n is the number of pixels in each quadrant of the image. In this Section we construct an almost-isometric embedding $\mathbb{R}^4/\mathbb{Z}_4 \hookrightarrow \mathbb{R}^8$ and then extend this to a locally almost-isometric map $(\mathbb{R}^4 \otimes \mathbb{R}^n)/(\mathbb{Z}_4 \times \{1\}) \to \mathbb{R}^{8n}$. We construct the map in five stages.

Stage 1 Change of basis

The first stage is an orthogonal change of basis of $\mathbb{R}^4$ which reveals the decomposition of $\mathbb{R}^4$ into invariant subspaces. The new basis, written in terms of the standard one, is

$$\left\{ e_1 = \frac{1}{2}(1,1,1,1), e_2 = \frac{1}{2}(1,1,-1,-1), e_3 = \frac{1}{2}(1,-1,-1,1), e_4 = \frac{1}{2}(1,-1,1,-1) \right\}.$$

Call the change of basis matrix P. Notice that $e_1 \cdot s = e_1$; $e_2 \cdot s = -e_3$ and $e_3 \cdot s = e_2$; and $e_4 \cdot s = -e_4$, so s acts by fixing $\mathbb{R}e_1$, rotating the (e_2, e_3)-plane by $\pi/2$, and reflecting in the hyperplane $e_4^\perp$. This shows that the action is trivial on the one dimensional subspace $\mathbb{R}e_1$, and preserves the unit sphere $\mathbb{S}^2$ in the orthogonal complement. We now use this decomposition to focus on finding an embedding of $\mathbb{S}^2/\mathbb{Z}_4$. We can write $\mathbb{Z}_4$ as a group extension $1 \to H \to \mathbb{Z}_4 \to Q \to 1$ where $H, Q \cong \mathbb{Z}_2$. It follows that $\mathbb{S}^2/\mathbb{Z}_4 = (\mathbb{S}^2/H)/Q$ where H acts on $\mathbb{S}^2$ by rotating by π in the (e_2, e_3)-plane, and Q acts on $\mathbb{S}^2/H$ by the antipodal map.

Stage 2 Embedding $\mathbb{S}^2/H \hookrightarrow \mathbb{R}^3$

Let $\mathbb{S}^2$ be the unit sphere in the copy of $\mathbb{R}^3$ spanned by $\{e_2, e_3, e_4\}$. Let (x, y, z) be a point written in these coordinates and, set $\eta_1(0, 0, \pm 1) = (0, 0, \xi(\pm 1))$ and otherwise define the map

$$\eta_1 : \mathbb{S}^2 - \{(0, 0, \pm 1)\} \to \mathbb{R}^3 : (x, y, z) \mapsto \left(\frac{1}{2} \frac{x^2 - y^2}{\sqrt{x^2 + y^2}}, \frac{1}{2} \frac{2xy}{\sqrt{x^2 + y^2}}, \xi(z) \right),$$

where

$$\xi(z) = \int_0^z \sqrt{\frac{1 - \frac{1}{4}t^2}{1 - t^2}} \, \mathrm{d}t$$

is an elliptic integral of the second kind. It is a routine calculation to check that η_1 indeed defines an isometric embedding $\mathbb{S}^2/H \hookrightarrow \mathbb{R}^3$.

Stage 3 Embedding $(\mathbb{S}^2/H)/Q \hookrightarrow \mathbb{R}^3$

Notice that the image of η_1 is indeed invariant under the antipodal map $(x, y, z) \mapsto (-x, -y, -z)$, so the action of Q is well-defined. There is a reasonably well-known isometric embedding of $\mathbb{R}^3/Q \hookrightarrow \mathbb{R}^6$ given by $\tilde{\nu} : \mathbb{R}^3 \to \mathbb{R}^6$ called the *Veronese embedding* (see Example 2.4 in [58], for example). We denote by $\eta_2 = \tilde{\nu}|_{\mathbb{S}^2/H}$ the restriction of this embedding to $\mathbb{S}^2/H$. First we define an isometric embedding $\nu : \mathbb{S}^2 \hookrightarrow \mathbb{R}^6$ of $\mathbb{S}^2/Q$, the unit sphere in $\mathbb{R}^3$ quotiented by Q,

$$\nu : \mathbb{S}^2 \to \mathbb{R}^6 : (x, y, z) \mapsto \left(\frac{1}{\sqrt{2}}x^2, \frac{1}{\sqrt{2}}y^2, \frac{1}{\sqrt{2}}z^2, xy, yz, zx \right).$$

This extends to $\mathbb{R}^3$ by taking the cone of the map as follows: write $r = \sqrt{x^2 + y^2 + z^2}$, then $\tilde{\nu}(\underline{0}) = \underline{0}$ and

$$\tilde{\nu} : \mathbb{R}^3 - \{\underline{0}\} \to \mathbb{R}^6 : (x, y, z) \mapsto r \left(\nu \left(\frac{(x, y, z)}{r} \right) + \frac{1}{3\sqrt{2}}(1, 1, 1, 0, 0, 0) \right).$$

This works because the image of ν lies in a sphere in a codimension 1 linear hyperplane of $\mathbb{R}^6$, so we have one spare dimension in which we can send the cone of the map ν. Stages 2 and 3 are illustrated in Figure 9.1.

Stage 4 Extending the embedding to $\mathbb{R}^4/\mathbb{Z}_4$

We have constructed $\eta_2 \circ \eta_1$ which isometrically embeds $\mathbb{S}^2/\mathbb{Z}_4$ in $\mathbb{R}^6$. Let $(a, b, c, d) \in \mathbb{R}^4$ be expressed in standard coordinates, and let $P(a, b, c, d) = (w, x, y, z)$. As
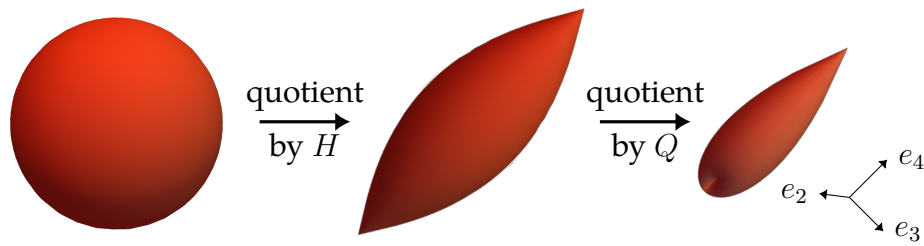
Figure 9.1: Constructing an isometric embedding of $\mathbb{S}^2/\mathbb{Z}^4 \subset \mathbb{R}^6$, which has been projected into three dimensions in the final picture for the purposes of illustration.

above define $r = \sqrt{x^2 + y^2 + z^2}$. Then let $\eta(\underline{0}) = \underline{0}$ and

$$\eta : \mathbb{R}^4 - \{\underline{0}\} \to \mathbb{R}^8 : (a, b, c, d) = \left(w, \frac{\sqrt{3}}{2}r\right) \oplus \eta_2 \left(r\eta_1 \left(\frac{(x, y, z)}{r}\right)\right). \qquad (9.2)$$

This is an embedding of $\mathbb{R}^4/\mathbb{Z}_4$, but it is not isometric, it does however very closely approximate an isometric embedding. We quantify this in Example 9.6.

Stage 5   The general case

We now have an *almost*-isometric embedding $\eta : \mathbb{R}^4 \to \mathbb{R}^8$ of the quotient space $\mathbb{R}^4/\mathbb{Z}_4$. We want to upgrade this to a map $\pi : \mathbb{R}^4 \otimes \mathbb{R}^n \to \mathbb{R}^k$ for some k which can reasonably be called a projection onto the quotient space $(\mathbb{R}^4 \otimes \mathbb{R}^n)/(\mathbb{Z}_4 \times \{(1)\})$. The following Proposition is straightforward to prove.

PROPOSITION 9.3: *Let G be a finite abelian group and X a Riemannian manifold on which G acts by isometries. Let $H = \bigtimes_{i=1}^n G$ act on $Y = \prod_{i=1}^n X$ component-wise and define $G_0 := \{(g, \ldots, g) \mid g \in G\} \leqslant H$ to be the diagonal subgroup which is isomorphic to G. Then the quotient spaces Y/G_0 and Y/H inherit orbifold structures from Y, and there is a canonical orbifold covering map $Y/G_0 \to Y/H$ of degree $\#G^{n-1}$ induced by the action of $Q = H/G_0$ on Y/G_0.*

Proof. The group H is finite and acts by isometries on Y. Point stabilisers are therefore finite groups, hence Y/G_0 and Y/H are orbifolds. The diagonal subgroup of a product is normal if and only if it is abelian, so in our case the quotient Q is well-defined and Q acts by isometries on Y/G_0 with $(Y/G_0)/Q = Y/H$. The quotient map is the required orbifold covering map. $\qquad \square$

Once can also check that $Y/H = \bigtimes_{i=1}^{k} X/G$. In our present case, let $X = \mathbb{R}^4$ and $G = \mathbb{Z}_4$. Then we can identify $Y = \bigoplus_{i=1}^{n} \mathbb{R}^4$ with $\mathbb{R}^4 \otimes \mathbb{R}^n$, and the action of G_0 on Y is the same as the action of $\mathbb{Z}_4 \times \{(1)\}$ on $\mathbb{R}^4 \otimes \mathbb{R}^n$. If we define

$$\pi := \bigoplus_{i=1}^{n} \eta : \bigoplus_{i=1}^{n} \mathbb{R}^4 \to \bigoplus_{i=1}^{n} \mathbb{R}^8$$

then this is an almost-isometric embedding of Y/H, which has the quotient space we are actually interested in, $(\mathbb{R}^4 \otimes \mathbb{R}^n)/(\mathbb{Z}_4 \times \{(1)\})$ as a 4^{n-1}-fold orbifold cover. In other words π is a map of this quotient space into $\mathbb{R}^{8n}$ which is locally almost an isometry, and which identifies orbits in sets of size 4^{n-1}.

9.4   Distortion of metrics

The chief difference between our approaches and that of [115] using polynomial invariants is that we choose our projection onto a fundamental domain or quotient space to be either a local isometry, or where this is impractical for quotient spaces, as close to a local isometry as possible. In this Section we introduce a measure of the extent to which a map fails to be a local isometry called the *distortion* of the map. We compute it for a number of the examples we have been considering.

Definition 9.4. Let a finite group G act by linear isometries on $\mathbb{R}^n$, and let $X \subset \mathbb{R}^n$ be a compact, measurable, G-invariant subset of $\mathbb{R}^n$. Suppose $\pi \colon \mathbb{R}^n \to \mathbb{R}^k$ is a smooth projection onto the quotient space. Let g_0 be the Riemannian metric on the non-singular locus of $\mathbb{R}^n/G$ inherited from the flat metric on $\mathbb{R}^n$, and let g_π be the metric induced by the flat metric on $\mathbb{R}^k$. Then we define the **distortion** of the map π on X to be

$$\mathrm{Dist}_X(\pi) := \int_{(X/G)^0} \|g_\pi - g_0\| \, \mathrm{d}g_0,$$

where $(X/G)^0$ is the non-singular locus of the quotient space, and $\|g_\pi - g_0\|$ is the 2–norm of the difference of the metrics expressed as matrices with respect to the standard coordinates on $\mathbb{R}^n$.

Compare this with the function **Loss** in [86]. There the authors find it convenient to integrate over $\|g_\pi - g_0\|^2$ when performing gradient descent, but as an

absolute measure of the difference between metrics it makes more sense not to square the integrand.

Example 9.5. If G acts on X with fundamental domain $\mathcal{F}$ such that each orbit intersects $\partial\mathcal{F}$ at most once, then we have already remarked that $\pi\colon X \to \mathcal{F}$ is also a projection onto the quotient space. It follows that π is exactly a local isometry, and $g_\pi = g_0$. Hence $(X/G)^0 = \mathcal{F} \cap X$ and

$$\mathrm{Dist}_X(\pi) := \int_{\mathcal{F} \cap X} 0 \; \mathrm{d}x_1 \cdots \mathrm{d}x_n = 0,$$

where the x_i's represent some coordinates on X. So there is no distortion.

Example 9.6. Let $G = \mathbb{Z}_4 \leqslant S_4$ act by cyclically permuting the coordinates of $\mathbb{R}^4$. In the previous Section we constructed an embedding $\eta\colon \mathbb{R}^4 \to \mathbb{R}^8$, (9.2), of the quotient space $\mathbb{R}^4/\mathbb{Z}_4$ which we claimed closely approximated an isometric embedding. On the other hand, following the approach outlined in Section 6.4, we can compute a generating set of polynomial invariants which gives another embedding $p\colon \mathbb{R}^4 \to \mathbb{R}^7$ of the quotient space. Explicitly

$$p(x_1, x_2, x_3, x_4) = \left(\sum x_i, \sum x_i^2, \; \sum x_i x_{i+1}, \sum x_i^3, \; \sum x_i x_{i+1}^2, \; \sum x_i^4, \; \sum x_i x_{i+1}^3 \right),$$

where each sum runs over $i = 1, \ldots, 4$ and indices are read modulo 4. This group action arises in the application of image recognition, where each coordinate represents the brightness of a pixel taking a value in the interval $[0, 1]$, so we choose as our compact $\mathbb{Z}_4$-invariant set, the unit cube $X = [0, 1]^4$. Then we can compute the distortions of each of these embeddings numerically:

$$\mathrm{Dist}_X(\eta) \simeq 0.29, \; \text{ and } \mathrm{Dist}_X(p) \simeq 8.38.$$

So p has almost 30 times more distortion than η.

9.5 Generalisations

Here we focuss mainly on a very special type of group action, namely permutation actions of finite groups on real vector spaces, and how to design machine learning architectures for problems which are invariant or equivariant with respect to

these actions. We also focus, in our example applications, on neural network architectures. Since projections onto a fundamental domain or the quotient space can be viewed as a pre-processing steps applied to the input data before machine learning, they can be applied to any type of supervised machine learning model.

Before discussing more general types of group actions for which our approach works, we want to mention two additions one could make to our approaches even in the case of finite groups acting by permutations. Notice that the definition of the *averaging* combinatorial projections in Section 8.1.2 simply involves pre-composing a normal combinatorial projection map by a G-equivariant linear map μ. Instead of fixing a particular choice of μ, one could replace this with a G-equivariant neural network with no hidden layers, of the type described in [81].

The second addition applies to quotient space projections, which we noted often artificially increase the input dimension space considerably. One way to mitigate this is described in [86], where the authors give an algorithm based on gradient descent with respect to a loss function similar to the *distortion* which we define in Section 9.4, to find near-isometric embeddings of manifolds. Their method could also take as input an embedding coming from a generating set of invariant polynomials and make it a near-isometric embedding.

For actions which are not properly discontinuous, for example if G is a real Lie group, then a fundamental domain does not exist. Nevertheless, the quotient space X/G does always exist, and in general has smaller dimension than X, meaning finding quotient space projections can be easier than in the case of properly discontinuous group actions. Once could also define analogues of a fundamental domain in this setting, and project onto this.

References

[1] Peter Abramenko and Kenneth S Brown. *Buildings. Theory and Applications*. Springer Science Buisness Media, Inc., 2010. ɪsʙɴ: 978–1–4419–2701–9.

[2] Goulnara N Arzhantseva. "A dichotomy for finitely generated subgroups of word hyperbolic groups". In: American Mathematical society, 2006, pp. 1–10.

[3] Benjamin Aslan, Daniel Platt, and David Sheard. *Group invariant machine learning by fundamental domain projections*. 2022. arXiv: 2202.02164 [cs.LG].

[4] Patrick Bahls. *The isomorphism problem in Coxeter groups*. Imperial College Press, 2005. ᴅᴏɪ: 10.1142/p385.

[5] Erkao Bao and Linqi Song. "Equivariant neural networks and equivarification". In: *CoRR* abs/1906.07172 (2019). arXiv: 1906.07172.

[6] Michael Barot and Bethany J Marsh. "Reflection group presentations arising from cluster algebras". In: *Transactions of the American Mathematical Society* 367.3 (2015), pp. 1945–1967. ɪssɴ: 00029947.

[7] Michael Ben-Zvi, Robert Kropholler, and Rylee Alanza Lyman. "Folding-like techniques for CAT(0) cube complexes". In: *Mathematical Proceedings of the Cambridge Philosophical Society* 173.1 (2022), pp. 227–238. ᴅᴏɪ: 10.1017/S0305004121000645.

[8] Mladen Bestvina and Mark Feighn. "Bounding the complexity of simplicial group actions on trees." In: *Inventiones mathematicae* 103.3 (1991), pp. 449–470.

[9] Sara C Billey and Venkatraman Lakshmibai. *Singular Loci of Schubert Varieties*. Birkhäuser Basel, 2000.

[10] Anders Bjorner and Francesco Brenti. *Combinatorics of Coxeter groups*. Springer Science Buisness Media, Inc., 2005. ɪsʙɴ: 3–540–44238–3.

[11] Kenneth D Blaha. "Minimum bases for permutation groups: the greedy approximation". In: *Journal of Algorithms* 13.2 (1992), pp. 297–306. ɪssɴ: 0196–6774.

[12] Nicolas Bourbaki. *Lie groups and Lie algebras*. Trans. French by Andrew Pressley. Elements of Mathematics. Springer-Verlag Berlin Heidelberg New York, 2002. Chap. IV–VI. ɪsʙɴ: 3–540–42650–7.

[13] Noel Brady et al. "Rigidity of Coxeter groups and Artin groups". In: *Geometriae Dedicata* 94 (2002), pp. 91–109. ᴅᴏɪ: 10.1023/A:1020948811381.

[14] Martin R Bridson and André Haefliger. *Metric Spaces of Non-Positive Curvature*. Grundlehren der mathematischen Wissenschaften. Springer Berlin Heidelberg, 2011. ISBN: 9783540643241.

[15] Martin R Bridson and Daniel T Wise. "Malnormality is undecidable in hyperbolic groups". In: *Israel Journal of Mathematics* 124.1 (2001), pp. 313–316. ISSN: 1565–8511.

[16] Kenneth S Brown. *Buildings*. Springer-Verlag New York Inc., 1989. ISBN: 0–387–98624–3.

[17] Kieran Bull et al. "Getting CICY high". In: *Physics Letters B* 795 (2019), pp. 700–706.

[18] Kieran Bull et al. "Machine learning CICY threefolds". In: *Physics Letters B* 785 (2018), pp. 65–72.

[19] Mathieu Carette and Richard Weidmann. *On the rank of Coxeter groups*. 2009. arXiv: 0910.4997 [math.GR].

[20] James B Carrel. "The Bruhat graph of a Coxeter group, a conjecture of Deodhar, and Rational Smoothness of Schubert varieties". In: *Proceedings of Symposia in Pure Math.* 56 (1994), pp. 53–61.

[21] Ruth Charney. "An introduction to right-angled Artin groups". In: *Geometriae Dedicata* 125.1 (2007), pp. 141–158. ISSN: 1572–9168.

[22] Shuxiao Chen, Edgar Dobriban, and Jane H Lee. "A group-theoretic framework for data augmentation". In: *Journal of Machine Learning Research* 21.245 (2020), pp. 1–71.

[23] Weimin Chen. "A homotopy theory of orbispaces". In: *arXiv preprint* (2001). URL: arXiv:math/0102020v3.

[24] William Chen. *Nonabelian level structures, Nielsen equivalence, and Markoff triples*. 2021. arXiv: 2011.12940 [math.NT].

[25] Taco Cohen and Max Welling. "Group equivariant convolutional networks". In: *International conference on machine learning*. PMLR. 2016, pp. 2990–2999.

[26] Taco Cohen and Max Welling. "Steerable CNNs". In: *International conference on learning representations*. 2017.

[27] Taco S Cohen, Mario Geiger, and Maurice Weiler. "A general theory of equivariant cnns on homogeneous spaces". In: *Advances in neural information processing systems* 32 (2019).

[28] Donald J Collins. "Generation and presentation of one-relator groups with centre". In: *Mathematische Zeitschrift* 157.1 (1977), pp. 63–77. ISSN: 1432–1823. DOI: 10.1007/BF01214681.

[29] Marston Conder. "Two element generation of the finite reflection groups". In: *The Quarterly Journal of Mathematics* 46.1 (1995), pp. 95–106. ISSN: 0033–5606. DOI: 10.1093/qmath/46.1.95.

[30] Harold SM Coxeter. "The complete enumeration of finite groups of the form $R_i^2 = (R_iR_j)^{k_{ij}} = 1$". In: *Journal of the London Mathematical Society* s1–10.1 (1935), pp. 21–25.

[31] Pallavi Dani and Ivan Levcovitz. "Subgroups of right-angled Coxeter groups via Stallings-like techniques". In: *J Comb Algebra* (2021). (To appear). DOI: 10.4171/JCA/54.

[32] Mike W Davis. *The geometry and topology of Coxeter groups*. Princeton University Press, 2008. ISBN: 0–691–13138–4.

[33] Max Dehn. "Über unendliche diskontinuierliche Gruppen". In: *Mathematische Annalen* 71.1 (1911), pp. 116–144. ISSN: 1432–1807.

[34] Vinay V Deodhar. "A note on subgroups generated by reflections in Coxeter groups". In: *Archiv der Mathematik* 53 (1989), pp. 543–546.

[35] Persi Diaconis and Ronald Graham. "The graph of generating sets of an abelian group". eng. In: *Colloquium Mathematicae* 80.1 (1999), pp. 31–38.

[36] John D Dixon and Abdul Majeed. "Coset representatives for permutation groups". eng. In: *Portugaliae mathematica* 45.1 (1988), pp. 61–68.

[37] Spencer Dowdall and Samuel J Taylor. "Rank and Nielsen equivalence in hyperbolic extensions". In: *International Journal of Algebra and Computation* 29 (2017). DOI: 10.1142/S0218196719500176.

[38] Martin J Dunwoody. "Nielsen transformations". In: *Computational problems in abstract algebra (Proc. Conf., Oxford, 1967)* (1970), pp. 45–46.

[39] Ederson RF Dutra. *Nielsen equivalence in closed 2-orbifold groups*. 2020. arXiv: 2004.09326 [math.GR].

[40] Ederson RF Dutra. *Nielsen equivalence in triangle groups*. 2021. arXiv: 2101.05344 [math.GR].

[41] Matthew J Dyer. "Hecke algebras and reflections in Coxeter groups". PhD thesis. University of Sydney, 1987.

[42] Matthew J Dyer. "Reflection subgroups of Coxeter systems". In: *Journal of Algebra* 135.1 (1990), pp. 57–73. ISSN: 0021–8693.

[43] Harold Erbin and Riccardo Finotello. "Inception neural network for complete intersection Calabi–Yau 3-folds". In: *Machine Learning: Science and Technology* 2.2 (2021), 02LT03.

[44] Martin J Evans. "Presentations of groups involving more generators than are necessary". In: *Proceedings of the London Mathematical Society* s3–67.1 (1993), pp. 106–126.

[45] Martin J Evans. "Primitive elements in free groups". In: *Proceedings of the American Mathematical Society* 106 (1989), pp. 313–316.

[46] Marc Finzi, Max Welling, and Andrew Gordon Wilson. "A practical method for constructing equivariant multilayer perceptrons for arbitrary matrix groups". In: *International conference on machine learning*. PMLR. 2021, pp. 3318–3328.

[47] Sergey Fomin and Andrei Zelevinsky. "Cluster algebras I: Foundations". In: *Journal of the American Mathematical Society* 15.2 (2002), pp. 497–529.

[48] Sergey Fomin and Andrei Zelevinsky. "Cluster algebras II: Finite type classification". In: *Inventiones mathematicae* 154.1 (2003), pp. 63–121. ISSN: 1432–1297.

[49] Ralph H Fox. "Free differential calculus. I: derivation in the free group ring". In: *Annals of Mathematics* 57.3 (1953), pp. 547–560.

[50] William N Franzsen. "Automorphisms of Coxeter groups". PhD thesis. University of Sydney, 2001.

[51] William N Franzsen and Robert B Howlett. "Automorphisms of nearly finite Coxeter groups." eng. In: *Advances in Geometry* 3.3 (2003), pp. 301–338.

[52] Fabian Fuchs et al. "SE(3)-transformers: 3D roto-translation equivariant attention networks". In: *Advances in neural information processing systems*. Ed. by Hugo Larochelle et al. Vol. 33. Curran Associates, Inc, 2020, pp. 1970–1981.

[53] Joseph Grant and Bethany J Marsh. "Braid groups and quiver mutation". In: *Pacific Journal of Mathematics* 290.1 (2017), pp. 77–116. DOI: 10.2140/pjm.2017.290.77.

[54] Paul Green and Tristan Hübsch. "Calabi-Yau manifolds as complete intersections in products of complex projective spaces". In: *Comm. Math. Phys.* 109.1 (1987), pp. 99–108.

[55] Igor Aleksandrovich Grushko. "On the bases of a free product of groups". In: *Matematicheskii Sbornik* 8.50 (1940), pp. 169–182.

[56] André Haefliger. *Complexes of groups and orbihedra*. World Scientific, 1991. ISBN: 981–02–0442–6.

[57] Frédéric Haglund and Daniel T Wise. "Special cube complexes". In: *Geometric and functional analysis* 17 (2008), pp. 1551–1620.

[58] Joe Harris. *Algebraic geometry: a first course*. Graduate texts in mathematics; 133. Springer-Verlag, 1992.

[59] Jason S Hartford et al. "Deep models of interactions across sets". In: *Proceedings of the 35th International Conference on Machine Learning*. Ed. by Jennifer G Dy and Andreas Krause. Vol. 80. Proceedings of Machine Learning Research. PMLR, 2018, pp. 1914–1923.

[60] Seyyed Hossein Hasanpour et al. "Towards principled design of deep convolutional networks: introducing SimpNet". In: *arXiv preprint* (2018). URL: arXiv:1802.06205.

[61] Yang-Hui He. "Machine-learning the string landscape". In: *Physics Letters B* 774 (2017), pp. 564–568.

[62] Yang-Hui He and Minhyong Kim. "Learning algebraic structures: preliminary investigations". In: *arXiv preprint arXiv:1905.02263* (2019).

[63] Graham Higman. "The units of group-rings". In: *Proceedings of the London Mathematical Society* s2–46.1 (1940), pp. 231–248.

[64] Pat Hill and Stephen J Pride. "Commutators, generators and conjugacy equations in groups". In: *Archiv der Mathematik* 44.1 (1985), pp. 1–14. ISSN: 1420–8938. DOI: 10.1007/BF01193774.

[65] Kurt Hornik. "Approximation capabilities of multilayer feedforward networks". In: *Neural Networks* 4.2 (1991), pp. 251–257.

[66] James E Humphreys. *Reflection groups and Coxeter groups*. Cambridge University Press, 1990. ISBN: 0–521–37510–X.

[67] Mark Jerrum. "A compact representation for permutation groups". In: *Journal of Algorithms* 7.1 (1986), pp. 60–78. ISSN: 0196–6774.

[68] Ilya Kapovich and Richard Weidmann. "Nielsen equivalence in a class of random groups". In: *Journal of Topology* 9.2 (2016), pp. 502–534. ISSN: 1753–8416. DOI: 10.1112/jtopol/jtw001.

[69] Ilya Kapovich and Richard Weidmann. *Nielsen equivalence in small cancellation groups*. 2012. arXiv: 1011.5862 [math.GR].

[70] Ilya Kapovich and Richard Weidmann. "Nielsen methods and groups acting on hyperbolic spaces". In: *Geometriae Dedicata* 98 (2002). DOI: 10.1023/A:1024064029186.

[71] Risi Kondor and Shubhendu Trivedi. "On the generalization of equivariance and convolution in neural networks to the action of compact groups". In: *International Cconference on machine learning*. PMLR. 2018, pp. 2747–2755.

[72] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. "Imagenet classification with deep convolutional neural networks". In: *Advances in neural information processing systems* 25 (2012), pp. 1097–1105.

[73] Yann LeCun et al. "Gradient-based learning applied to document recognition". In: *Proceedings of the IEEE* 86.11 (1998), pp. 2278–2324.

[74] Martin W Liebeck. "On minimal degrees and base sizes of primitive permutation groups". In: *Archiv der Mathematik* 43.1 (1984), pp. 11–15. ISSN: 1420–8938. DOI: 10.1007/BF01193603.

[75] Richard J Lipton and Yechezkel Zalcstein. "Word Problems Solvable in Logspace". In: *J ACM* 24.3 (1977), pp. 522–526. ISSN: 0004–5411.

[76] Larsen Louder. *Nielsen equivalence in closed surface groups*. 2015. arXiv: 1009.0454 [math.GR].

[77] Martin Lustig. "Nielsen equivalence and simple-homotopy type". In: *Proceedings of the London Mathematical Society* s3–62.3 (1991), pp. 537–562. ISSN: 0024–6115. DOI: 10.1112/plms/s3-62.3.537.

[78] Martin Lustig and Yoav Moriah. "Generating systems of groups and Reidemeister-Whitehead torsion". In: *Journal of Algebra* 157 (1993). DOI: 10.1006/jabr.1993.1096.

[79] Martin Lustig and Yoav Moriah. "Nielsen equivalence in Fuchsian groups". In: *Algebraic and Geometric Topology* 22.1 (2022), pp. 189–226. DOI: 10.2140/agt.2022.22.189.

[80] Martin Lustig and Yoav Moriah. "Nielsen equivalence in Fuchsian groups and Seifert fibered spaces". In: *Topology* 30 (1991), pp. 191–204. DOI: 10.1016/0040-9383(91)90005-0.

[81] Haggai Maron et al. "On learning sets of symmetric elements". In: *Proceedings of the 37th international conference on machine learning*. Vol. 119. Proceedings of Machine Learning Research. PMLR, 2020, pp. 6734–6744.

[82] Haggai Maron et al. "On the Universality of Invariant Networks". In: *Proceedings of the 36th International Conference on Machine Learning*. Ed. by Kamalika Chaudhuri and Ruslan Salakhutdinov. Vol. 97. Proceedings of Machine Learning Research. PMLR, 2019, pp. 4363–4371.

[83] Timothée Marquis and Bernhard Mühlherr. "Angle-deformations in Coxeter groups". In: *Algebraic and Geometric Topology* 8.4 (2008), pp. 2175–2208. ISSN: 1472–2747. DOI: 10.2140/agt.2008.8.2175.

[84] George Maxwell. "The Normal Subgroups of Finite and Affine Coxeter Groups". In: *Proceedings of the London Mathematical Society* 76.2 (1998), pp. 359–382.

[85] Darryl McCullough. "Generating pairs and group actions". In: *Journal of Pure and Applied Algebra* 218.5 (2014), pp. 777–783. ISSN: 0022–4049.

[86] James McQueen, Marina Meila, and Dominique Joncas. "Nearly Isometric Embedding by Relaxation". In: *Advances in Neural Information Processing Systems*. Ed. by Daniel Lee et al. Vol. 29. Curran Associates, Inc, 2016.

[87] Michael L Mihalik and John G Ratcliffe. "On the rank of a Coxeter group". In: *Journal of Group Theory* 12.3 (2009), pp. 449–464. DOI: doi:10.1515/JGT.2008.089.

[88] Bernhard Mühlherr. "On Isomorphisms between Coxeter Groups". In: *Designs, Codes and Cryptography* 21.1 (2000), pp. 189–189. ISSN: 1573–7586.

[89] Aglaia Myropolska. "Nielsen equivalence in Gupta-Sidki groups". In: *Proceedings of the American Mathematical Society* 145 (2014). DOI: 10.1090/proc/13612.

[90] Bernhard H Neumann. "On a question of Gaschütz". In: *Archiv der Mathematik* 7 (1956), pp. 87–90.

[91] Jakob Nielsen. "Die isomorphismengruppe der freien gruppen". In: *Mathematische Annalen* 91.3 (1924), pp. 169–209. ISSN: 1432–1807. DOI: 10.1007/BF01556078.

[92] Emmy Noether. "Der endlichkeitssatz der invarianten endlicher gruppen". In: *Mathematische Annalen* 77.1 (1915), pp. 89–92.

[93] Gennady A Noskov. "Primitive elements in a free group". In: *Mat. Zametki* 30 (4 1981), pp. 497–500.

[94] Daniel Oancea. "A note on Nielsen equivalence in finitely generated ableian groups". In: *Bulletin of the Australian Mathematical Society* 84.1 (2011), pp. 127–136. DOI: 10.1017/S0004972711002279.

[95] David Radcliffe. "Rigidity of right-angled Coxeter groups". In: *arXiv: Group Theory* (1999).

[96] Siamak Ravanbakhsh, Jeff G Schneider, and Barnabás Póczos. "Deep learning with sets and point clouds". In: *CoRR* abs/1611.04500 (2016).

[97] Siamak Ravanbakhsh, Jeff G Schneider, and Barnabás Póczos. "Equivariance through parameter-sharing". In: *Proceedings of the 34th international conference on machine learning*. Ed. by Doina Precup and Teh Yee Whye. Vol. 70. Proceedings of Machine Learning Research. PMLR, 2017, pp. 2892–2901.

[98] Melissa Rodd. *Tree-scapes*. 2021. URL: https://art2021lsgnt.weebly.com/transforming-trees.html.

[99] Gerhard Rosenberger. "Minimal generating systems for plane discontinuous groups and an equation in free groups". In: *Groups-Korea 1988*. Ed. by Ann C Kim and Bernhard H Neumann. Lecture notes in mathematics 1398. note. Springer, Berlin, 1988, pp. 170–186.

[100] Gerhard Rosenberger. "Zum rang-und isomorphieproblem für freie produkte mit amalgam". PhD thesis. Univ, 1974.

[101] Gerald W Schwarz. "Smooth functions invariant under the action of a compact lie group". In: *Topology* 14.1 (1975), pp. 63–68.

[102] David Sheard. *Coxeter groups software*. Version 0.1.0. URL: github.com/trakkabase/Coxeter-groups.

[103] Juan Souto. "The rank of the fundamental group of certain hyperbolic 3–manifolds fibering over the circle". In: *The Zieschang Gedenkschrift* (2008). ISSN: 1464–8997. DOI: 10.2140/gtm.2008.14.505.

[104] John R Stallings. "A topological proof of Grushko's theorem on free products". In: *Mathematische Zeitschrift* 90.1 (1965), pp. 1–8. ISSN: 1432–1823. DOI: 10.1007/BF01112046.

[105] John R Stallings. "Topology of finite graphs". In: *Inventiones mathematicae* 71.3 (1983), pp. 551–565. ISSN: 1432–1297. DOI: 10.1007/BF02095993.

[106] Jacques Tits. "Groupes et géométries de Coxeter". French. unpublished. 1961.

[107] Jacques Tits. "Le problème des mots dans les groupes de Coxeter". French. In: *1969 Simposia Mathematica (INDAM, Rome 1967/8)*. Vol. 1. Academic Press, London, 1969, pp. 175–185.

[108] Soledad Villar et al. "Scalars are universal: Equivariant machine learning, structured like classical physics". In: *Advances in Neural Information Processing Systems* 34 (2021), pp. 28848–28863.

[109] Ernest B Vinberg. "Discrete linear groups generated by reflections". In: 5.5 (1971), pp. 1083–1119. DOI: 10.1070/im1971v005n05abeh001203.

[110] Ernest B Vinberg. *Geometry II: spaces of constant curvature*. eng. 1st ed. 1993. Encyclopaedia of mathematical sciences; Volume 29. Berlin, Heidelberg: Springer-Verlag, 1993. ISBN: 3–540–52000–7.

[111] Patrick Wegener and Sophiane Yahiatene. *Reflection factorizations and quasi-Coxeter elements*. 2021. arXiv: `2110.14581` [`math.GR`].

[112] Maurice Weiler and Gabriele Cesa. "General $E(2)$-equivariant steerable CNNs". In: *Advances in neural information processing systems*. Ed. by Hanna Wallach et al. Vol. 32. Curran Associates, Inc, 2019.

[113] Hermann Weyl. *The classical groups: their invariants and representations*. Princeton University Press, 1966.

[114] Patrick A Worfolk. "Zeros of equivariant vector fields: algorithms for an invariant approach". In: *Journal of Symbolic Computation* 17.6 (1994), pp. 487–511.

[115] Dmitri Yarotsky. "Universal approximations of invariant maps by neural networks". In: *Constr Approx* (2021).

[116] Manzil Zaheer et al. "Deep sets". In: *Advances in neural information processing systems 30: annual conference on neural information processing systems 2017*. Ed. by Isabelle Guyon et al. 2017, pp. 3391–3401.

[117] Heiner Zieschang. "Über die Nielsensche kürzungsmethode in freien produkten mit amalgam". In: *Inventiones mathematicae* 10 (1970), pp. 4–37.

Subject index