2014 PhD Summer School in Discrete Mathematics

Mark Ellingham ▪ *Mariusz Meszka* ▪ *Primož Moravec* ▪ *Enes Pasalic*

FAMNIT LECTURES ▪ FAMNITOVA PREDAVANJA ▪ 3

2014 PhD Summer School in Discrete Mathematics

Famnit Lectures | Famnitova predavanja | 3

2014 PhD Summer School in Discrete Mathematics

Mark Ellingham
Mariusz Meszka
Primož Moravec
Enes Pasalic

Famnit Lectures 3 | ISSN 2335-3708

2014 PhD Summer School in Discrete Mathematics

Dr Mark Ellingham, Vanderbilt University, USA
Dr Mariusz Meszka, AGH University of Science and Technology, Poland
Dr Primož Moravec, University of Ljubljana, Slovenia
Dr Enes Pasalic, University of Primorska, Slovenia

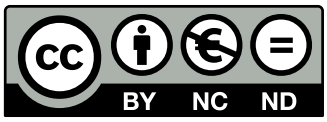
Published by
University of Primorska Press,
Titov trg 4, SI-6000 Koper
Koper · 2014

Editor-in-Chief
Dr Jonatan Vinkler

Managing Editor
Alen Ježovnik

www.hippocampus.si

Print run · 60 · Not for sale



CIP - Kataložni zapis o publikaciji
Narodna in univerzitetna knjižnica, Ljubljana

51(082)(0.034.2)

TWO thousand and fourteen
2014 Phd summer school in discrete mathematics [Elektronski vir] / Mark Ellingham ... [et al.]. - El. knjiga. - Koper : University of Primorska Press, 2014. - (Famnit lectures = Famnitova predavanja, ISSN 2335-3708 ; 3)

ISBN 978-961-6832-92-2 (pdf)
ISBN 978-961-6832-93-9 (html)
1. Ellingham, Mark N.
275179264

REPUBLIKA SLOVENIJA
MINISTRSTVO ZA IZOBRAŽEVANJE, ZNANOST IN ŠPORT

*O*peracijo delno financira Evropska unija, in sicer iz Evropskega socialnega sklada. Projekt se izvaja v okviru Operativnega programa razvoja človeških virov 2007-2013, razvojne prioritete 3: “Razvoj človeških virov in vseživljenjskega učenja”; prednostne usmeritve 3.3 “Kakovost, konkurenčnost in odzivnost visokega šolstva”.

Preface

This is a collection of lecture notes of the PhD Summer School in Discrete Mathematics, held from June 29 to July 5, 2014, by tradition at Rogla, Slovenia. The organization of this summer school came as a combined effort of the Faculty of Mathematics, Natural Sciences and Information Technologies and the Andrej Marušič Institute at the University of Primorska, and the Centre for Discrete Mathematics at the Faculty of Education at the University of Ljubljana.

The Scientific Committee of the meeting consisted of Klavdija Kutnar, Aleksander Malnič, Dragan Marušič, Štefko Miklavič and Primož Šparl. The Organizing Committee of the meeting consisted of Iva Antončič, Ademir Hujdurović, Boštjan Frelih and Boštjan Kuzman.

The aim of this Summer School was to bring together senior researchers, junior researches and PhD students working in Algebraic Graph Theory. The summer school has consisted of lectures given by

- Dave Witte Morris, University of Lethbridge, Canada,
- Joy Morris, University of Lethbridge, Canada,

and four minicourses given by

- Mark Ellingham, Vanderbilt University, USA,
- Mariusz Meszka, AGH University of Science and Technology, Poland,
- Primož Moravec, University of Ljubljana, Slovenia,
- Enes Pasalic, University of Primorska, Slovenia.

Contents

Chapter 1

Construction Techniques for Graph Embeddings

Mark Ellingham
Vanderbilt University, USA

SUMMARY

Mathematicians have been trying to construct embeddings of specific graphs in surfaces since at least the 1890s. However, until the 1960s the construction techniques were usually fairly ad hoc, although some general ideas such as 'schemes of cyclic sequences' had emerged. This changed with the development of current graphs by Gustin and others in the 1960s, which provided a unified framework for many earlier constructions and played an important role in the proof of the Map Colour Theorem. Fifty years later we have a number of useful general tools for constructing embeddings of graphs. These lectures will survey tools of various kinds. We will look at algebraic methods such as current, voltage and transition graphs; surgical tools such as the diamond sum and adding handles or crosscaps around a vertex; lifting constructions due to Bouchet and his collaborators; and techniques that use objects from design theory, such as latin squares, to construct embeddings.

NOTE ON PRESENTATION: These are lecture notes for a course that will survey a lot of material in a short amount of time, so the presentation is often informal and rigorous details are omitted. The figures are taken from a number of different sources. Some are hand-drawn, others are drawn using software packages. The author apologizes for the lack of consistency!

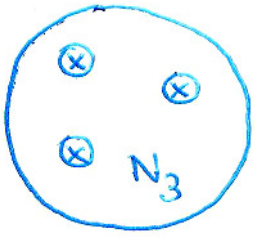
1.1 Embeddings of graphs

Surfaces

Definition: A *surface* is a 2-manifold without boundary. Examples: sphere, torus, projective plane, Klein bottle (all compact); plane, open Möbius strip (not compact).

Theorem, Classification of Surfaces: Every compact surface is homeomorphic to the sphere S_0, a sphere with $h \geq 1$ handles added S_h, or a sphere with $k \geq 1$ crosscaps added N_k.

Definition: *Adding a handle*: delete a disk, glue a punctured torus on to the boundary. *Adding a crosscap*: delete a disk, glue a punctured projective plane (i.e., a Möbius strip) on to the boundary.



Surfaces S_h, $h \geq 0$, are *orientable*: can define consistent clockwise orientation everywhere. Surfaces N_k, $k \geq 1$ are *nonorientable*: can travel in surface, maintaining locally consistent clockwise orientation, in such a way that orientation is reversed when you return to your starting point.

In an orientable surface all closed curves are 2-*sided*; nonorientable surfaces have 1-*sided* closed curves.

Question: What if add mixture of handles and crosscaps? Adding a crosscap and a handle is equivalent to adding three crosscaps. Consequently, if add $h \geq 0$ handles, $k \geq 1$ crosscaps, get N_{2h+k}.

Definition: The *genus* of a surface is the number of added handles or crosscaps: genus of S_h is h, genus of N_k is k.

Convention: From now on 'surface' means 'compact surface' unless otherwise specified.

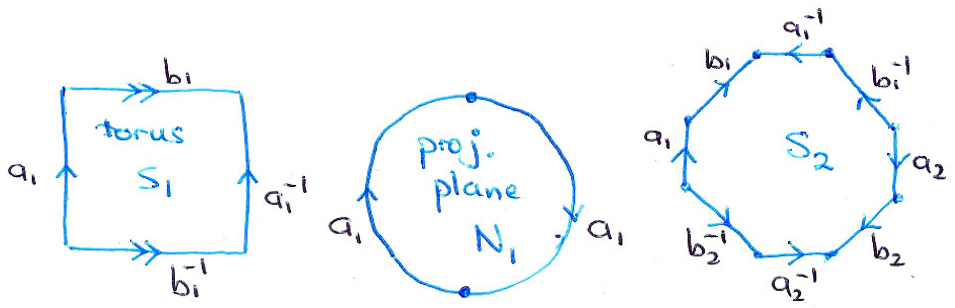
Representing surfaces

Polygon representation: Proof of classification theorem shows that every surface can be represented in a standard way as a polygon (possibly a 2-gon) with sides identified in pairs. Use inverse notation when sides identified in opposite directions.

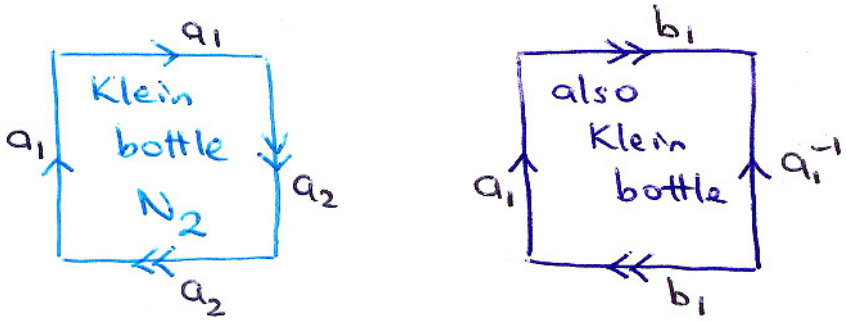
Sphere S_0: (aa^{-1})

S_h, $h \geq 1$: $(a_1 b_1 a_1^{-1} b_1^{-1} \ldots a_h b_h a_h^{-1} b_h^{-1})$

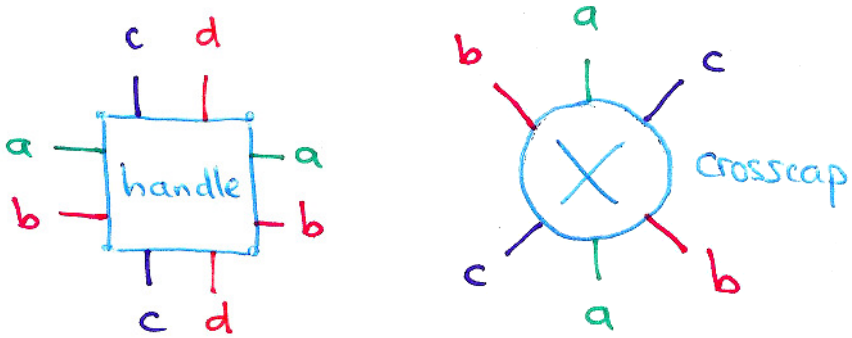
N_k, $k \geq 1$: $(a_1 a_1 a_2 a_2 \ldots a_k a_k)$



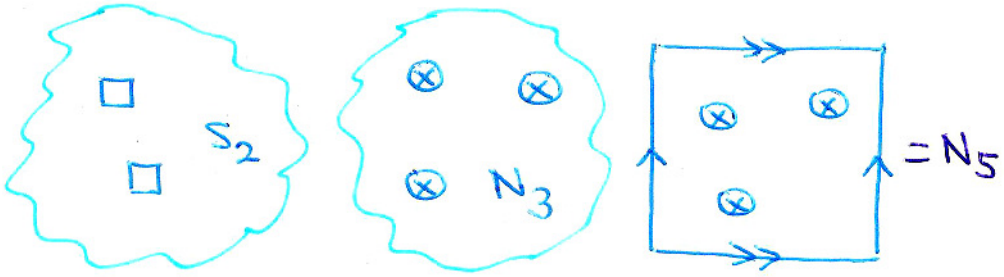
Surfaces can also be represented in other ways as polygons with identified sides, e.g. 'usual' representation of Klein bottle is not standard one.



Planar representation with handle or crosscap gadgets: Can also represent surfaces in plane: think of sphere as plane with implicit point at infinity, then add handles or crosscaps which we treat as 'gadgets' allowing curves to cross in certain ways.

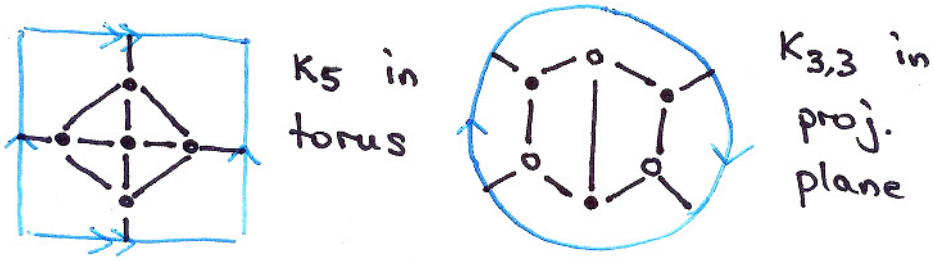


We can mix the above two representations: use polygon representation and then add handles or crosscaps.



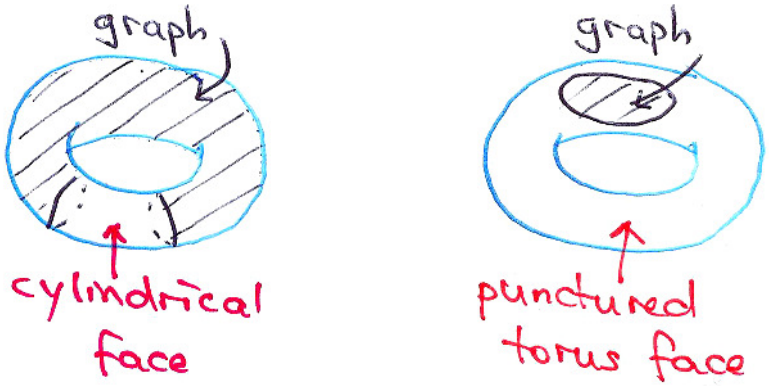
Graph embeddings

Definition: Loosely, an *embedding* Ψ *of graph* G *in surface* Σ, which we denote $\Psi : G \hookrightarrow \Sigma$, is a drawing of G in Σ with no crossing edges. Can make this rigorous, but concept should be clear.



Can represent embedding by drawing on either representation above (polygon with identified sides, or plane plus handle/crosscap gadgets), or on mixed representation. But if embedding nice, can represent in purely combinatorial ways or by simpler drawings.

Definition: Embedding of graph is *cellular* or open 2-cell or just 2-*cell* if every face is homeomorphic to an open disk.



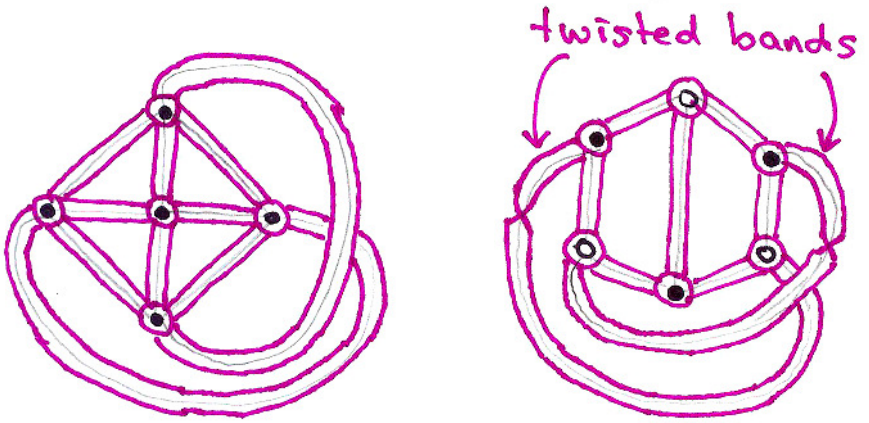
What prevents an embedding being 2-cell? Face has multiple boundary components, or face contains handles or crosscaps.

Even stronger definition: embedding is *closed* 2-*cell* if the closure of every face is homeomorphic to a closed disk. Equivalent to open 2-cell and boundary of every face is a cycle (not just a closed walk) in the graph. Closed 2-cell embeddings give *cycle double covers*. Closed 2-cell is usually a stronger property than we need or want.

Representation of 2-cell embeddings

Since all faces are open disks, just need to know how to glue faces onto graph.

Band decompositions or ribbon graphs: Take small disk around each vertex, small band (or strip) along each edge, throw rest of surface away. Get a 'fattened' version of graph. Can reconstruct entire surface by gluing a disk along each boundary component of resulting complex.



Rotation schemes: If our surface is orientable and we know a consistent global clockwise orientation, we can describe the embedding just by giving the clockwise order (*rotation*) of ends of edges at each vertex. This is a *pure rotation system*. Essentially known by Heffter in 1891, formalized by Edmonds in 1960.

More general definition: If we do not know a consistent global clockwise orientation (always true if our surface is nonorientable, but surface could also be orientable) then we use a *local* clockwise orientation for each vertex to give the order of ends of edges. But then we need to say whether rotations at two ends of an edge match up.

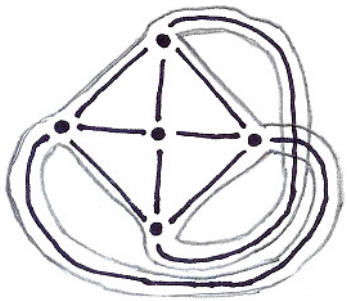
An edge is *type* 0 or *signature* 1 or *untwisted* if the local clockwise rotation of the vertex at one end can be followed along the edge and agrees with the local clockwise rotation of the vertex at the other end. Otherwise the edge is *type* 1 or *signature* -1 or *twisted*.

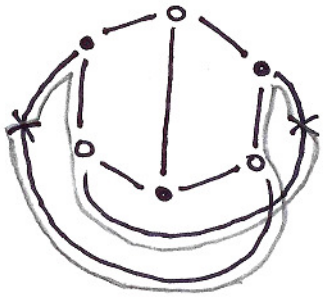
A *rotation scheme* in general consists of the orders of ends of edges around each vertex plus the type of each edge. This is a *purely combinatorial* description.

We can tell if a closed walk in a graph is 1-sided from this. A walk is 1-sided if and only if it contains an odd number of twisted (type 1) edges.

Rotation projections: However, it is convenient to represent a rotation scheme geometrically by a *rotation projection*. We just draw the graph in the plane, with edge crossings allowed, so that the clockwise order of ends of edges around each vertex agrees with the local clockwise orientation of the surface at that vertex. We indicate twisted (type 1) edges by putting an 'X' in the middle of them.

Face tracing for rotation projections: We can determine the face boundaries by following along the sides of edges, taking corners in the natural way, ignoring edge crossings, and switching sides in the middle of a twisted edge (at the 'X').





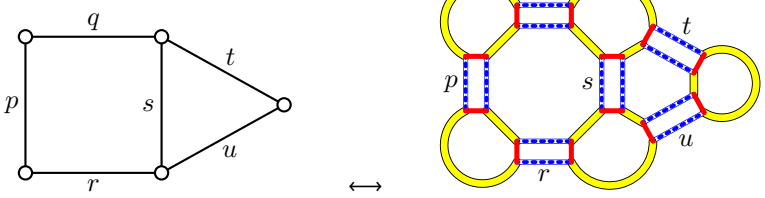
Orientability detection for rotation projections: The presence of twisted edges does not necessarily mean the embedding is nonorientable. Take spanning tree, start at root vertex, flip rotations so that all edges in tree become untwisted. Embedding orientable if and only if *all* edges now untwisted.

Gem representation: Due to Neil Robertson, 1971. Make band decomposition into 3-edge-coloured cubic graph:

Corner → vertex.
Vertex/face boundary → yellow edge.
Vertex/edge boundary → red edge.
Edge/face boundary → blue edge.

Embedded graphs ⟷ 3-edge-coloured graphs in which every red-blue cycle (edge) is a 4-cycle. Red-yellow cycles represent vertices, blue-yellow cycles represent faces.

Theory of gems developed extensively in book by Bonnington and Little [BL].



Facial walk description: Give collection of closed walks that cover every edge exactly twice. Can glue a disk along each such walk to get a surface provided have 'proper rotation' at each vertex, determined using 'rotation graph'.

Definition: The *rotation graph* at v has as vertices the ends of edges incident with v. Join two ends of edges if there is a face that passes through them consecutively. Rotation graph is *proper* if it consists of a single cycle.

Rotation graphs useful for building embeddings, basis of idea of transition graphs later.

Rotation graphs are useful for relative (partial) embeddings. E.g., rephrasing of theorem of Škoviera and Širáň, 1986: Given a graph G, a collection of closed walks using each edge at most twice can be completed to an embedding if and only if each rotation graph is a subgraph of a cycle (so is a spanning cycle, or is a collection of paths possibly including isolated vertices).

Embedding described by collection of facial walks is orientable if and only if can orient each walk so that every edge is used once in each direction.

Euler's formula

We have a fundamental counting relationship for graphs with 2-cell embeddings on surfaces.

Euler's formula: Suppose we have a 2-cell embedding of a connected graph G on a surface Σ, where G has v vertices, e edges, and the embedding has f faces. Then

$$v - e + f = \chi$$

where $\chi = \chi(\Sigma)$ is a constant that depends only on the surface; in particular,
$\chi(S_h) = 2 - 2h$ for $h \geq 0$ and
$\chi(N_k) = 2 - k$ for $k \geq 1$.

Definition: $\chi(\Sigma)$ is the *Euler characteristic* and can often be used to handle both orientable and nonorientable surfaces at the same time. But often convenient and more intuitive to have a nonnegative number with the same property. Define the *Euler genus* $\varepsilon(\Sigma)$ by
$\varepsilon(S_h) = 2h$ for $h \geq 0$ and
$\varepsilon(N_k) = k$ for $k \geq 1$
so that $\chi(\Sigma) = 2 - \varepsilon(\Sigma)$.

Example: K_5 on torus S_1: $v = 5$, $e = 10$, $f = 5$, $v - e + f = 5 - 10 + 5 = 0 = 2 - 2 \times 1$.

Important note: For Euler's formula to work, graph must be connected and embedding must be 2-cell. (There are more general versions that work if we relax these restrictions, but we need them for the basic formula above.)

Euler's formula and face degrees: Euler's formula gives an important implication involving the *degrees* of faces (lengths of facial walks) in an embedding. Since $\varepsilon = 2 - v + e - f$, for a minimum genus embedding of a given graph G (meaning v and e are fixed) we want to maximize f. Since the sum of the face degrees is $2e$, which is fixed, this means we want many faces of small degree. For a simple graph, we want triangular faces.

Based on considerations like this we can often find obvious lower bounds on the genus of embeddings of a given graph G. We then want to show that this lower bound can be achieved by constructing an embedding.

Exercise: Consider the usual drawing of the prism $K_2 \square C_n$, $n \geq 3$, (cartesian product) in the plane.

(a) If we make every K_2 edge twisted, what is the Euler genus of the corresponding embedding (use Euler's formula!), and is it orientable or nonorientable?

(b) Suppose we instead make every edge of one copy of C_n twisted, but leave all other edges untwisted. Answer the same question.

1.2 Voltage graphs

Note: Main reference for this section and next is Gross and Tucker's book [GT]. My notation and setup is similar to [GT], but not exactly the same.

Voltage graphs

Basic construction: Start with *base graph* G, orient each edge e arbitrarily to get directed edge e^+ in oriented graph $\boldsymbol{G}$, reverse of e^+ is e^-. (Oriented graph here refers to putting a direction on each edge, nothing to do with surfaces.)

Have *voltage group* Γ (usually assumed to be finite), every edge assigned a weight or *voltage* $\alpha(e^+)$. Implicitly $\alpha(e^-) = \alpha(e)^{-1}$. Form *derived graph* G^α as follows:

$V(G^\alpha) = V(G) \times \Gamma$.

For each e^+ from u to v in $\boldsymbol{G}$ with $\alpha(e^+) = a$, add an (oriented) edge (e^+, g) in G^α from (u, g) to (v, ga) for every $g \in \Gamma$. (Reverse of (e^+, g) is (e^-, ga).) Edge directions can now be ignored.

Note: We multiply edge weights on right; could equally well define with edge weights multiplying on left.

At this point we have just constructed a *graph*, no embeddings yet.

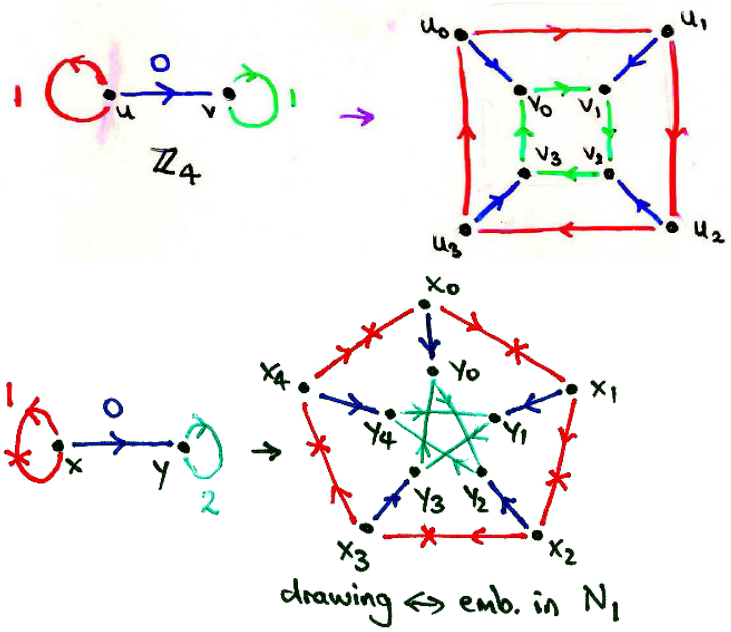
Remark: A *Cayley graph* is just a connected graph derived from a 1-vertex base voltage graph. Since G has only one vertex, vertices of G^α can be identified with elements of Γ.

Embedded voltage graphs

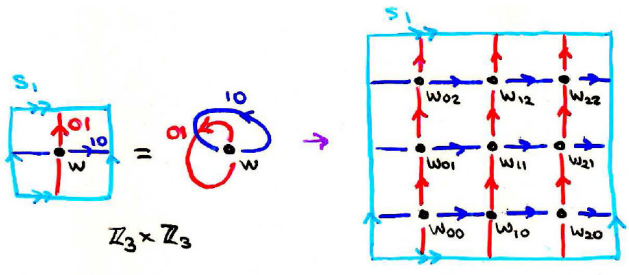
Extension to embedded graphs: Suppose base graph G has 2-cell embedding Ψ in surface. Describe using rotation projection. Construct 2-cell embedding of derived graph with following additional rules:

Around each vertex (v, g) of G^α the edges follow the order of their images in G (rotations are lifted).

Each edge in G^α has the same type (untwisted or twisted) as its image in G.



Can actually describe in more abstract terms without using rotation projection, but equivalent. Resulting *derived embedding* Ψ^α does not depend on specific rotation projection. Objects in Ψ or G are said to *lift* to corresponding objects in Ψ^α or G^α.



Lifting walks: Suppose $u\,v$-walk W in G corresponds to sequence of edges/reverse edges in $\boldsymbol{G}$ that is $f_1 f_2 \ldots f_d$. Say *net voltage* of W is $\alpha(W) = \alpha(f_1)\alpha(f_2)\ldots\alpha(f_d)$. If start at a vertex (u, g) in G^α and follow lifted walk $\tilde{W}$ in G^α, will end at $(v, g\alpha(W))$.

In particular, if W is a facial walk in G starting at u, will end at $(u, g\alpha(W))$. Will come back to original vertex (u, g) if repeat r times where r is the order of $\alpha(W)$ in Γ. Thus, each face of degree d in G becomes a face of degree dr in G^α where r is the order in Γ of the net voltage of the facial walk. Face length does not change exactly when net voltage is the identity (face satisfies *Kirchoff Voltage Law, KVL*).

Orientability: If original embedding of G is orientable, derived embedding will be orientable. If original embedding is nonorientable derived embedding could end up being orientable if all 1-sided walks lift to 2-sided walks.

Gross and Tucker [GT, 4.1.6] have algorithm based on reducing voltages in a spanning tree to the identity; won't discuss details.

But if voltage group has odd order all 1-sided closed walks have net voltage of odd order, must be repeated an odd number of times to close up in derived embedding, stay 1-sided, so embedding stays nonorientable.

More general voltage graphs

Permutation/group action voltage graphs: Gross and Tucker describe 'permutation voltage graphs' using permutation groups. Permutation groups are equivalent to group actions so can also describe that way. Suppose have right action of group Γ on set S: for each $s \in S$, $g \in \Gamma$ can form sg obeying natural rules.

Then given graph G with edges oriented and voltage $\alpha(e^+) \in \Gamma$ for each edge e, can form derived graph with

$V(G^\alpha) = V(G) \times S$.

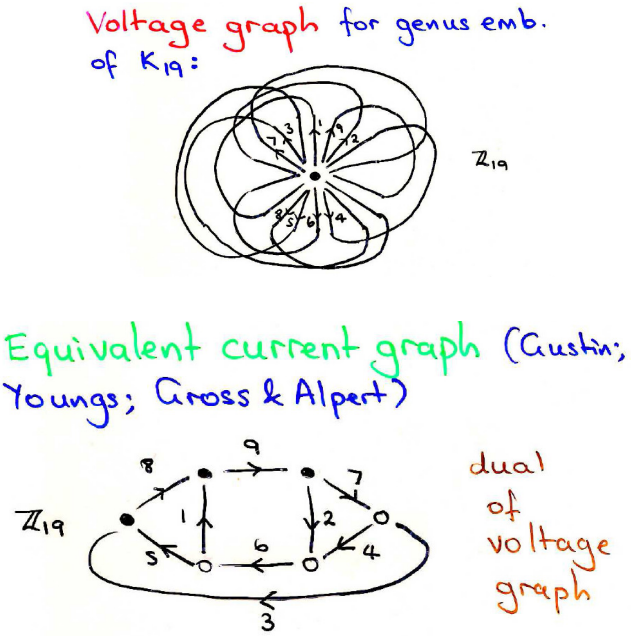
For each e^+ from u to v in G with $\alpha(e^+) = a$, add an edge (e^+, s) in G^α from (u, s) to (v, sa) for every $s \in \Gamma$. (Reverse of (e^+, s) is (e^-, sa).)

Can lift embedding of G to derived embedding of G^α in same way as for ordinary voltage graphs: lift vertex rotations and edge twists.

Final remark: Voltage graphs are straightforward to understand but may not be most convenient representation for particular applications. For very symmetric graphs a voltage graph representation of an embedding may have only one or two vertices and many edges, making it hard to keep track of where the edges go. So will look at alternative, current graphs, and then later another alternative, transition graphs.

1.3 Current graphs

Background: Current graphs were invented before voltage graphs, even though less intuitive. Used in proof of Map Colour Theorem, determination of minimum genus of complete graphs. Equivalent voltage graphs would have very few vertices, so it would be very hard to keep track of where the edges go.



Current graphs are duals of voltage graphs (so apply to *embedded* voltage graphs). Faces of current graph correspond to vertices in a voltage graph, and vice versa. However, tricky to deal with duals when have edge weights: need to turn them 90°, but which way? Hard to decide without globally consistent orientation (which never have in nonorientable case, and may not be given in orientable case).

See Gross and Tucker [GT] for general treatment. For simplicity we will restrict to current graphs given as rotation projections in plane.

Current graphs without twisted edges (hence orientable)

Basic construction: We are given oriented graph with weights or *currents* on edges, from *current group* Γ. For applications convenient to have two sorts of vertices (although only really need one sort):

solid $\bullet$ = clockwise vertices,

open $\circ$ = anticlockwise vertices.

Obtain derived embedding as follows:

Vertices of derived embedding have form (f,t) with f a face of the base graph and $t \in \Gamma$. To get faces of base graph with globally consistent rotations, trace faces in current graph in such a way that every edge is used once in each direction (trace all clockwise, or all anticlockwise). Order of edges along face f specify rotation around each vertex (f,t) in derived embedding.

Edges of derived embedding have form $(f,t)(g,ta)$, where f and g are faces of base graph meeting along an edge of current a; decide which way current applies based on rules below.

Faces of derived embedding come from vertices of base graph. For each vertex multiply currents of incident edges together in direction of vertex to get net current. Order r of net current in current group specifies how many times that sequence of edges is repeated to give a face of the derived graph, so vertex of degree d yields face of degree dr.

If net current of vertex is the identity, say vertex obeys the *Kirchoff Current Law, KCL*. Then vertex of degree d yields face of degree d.

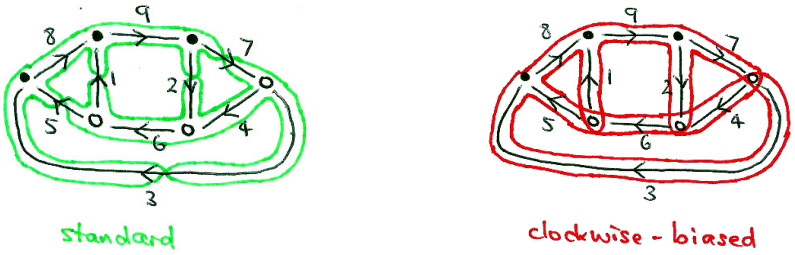
Standard tracing algorithm:

at each vertex follow natural rotation;

if an edge has vertices of different directions at its ends, cross over in the middle;

if we are leaving a clockwise vertex on an edge with face f on left, face g on right, current is a, then in derived graph (f,t) is joined to (g,ta) for each $t \in \Gamma$ (current acts 90° clockwise because vertex is clockwise); for an anticlockwise vertex swap left ⟷ right (current acts 90° anticlockwise).

There are alternative ways to trace faces that are more convenient in some ways.

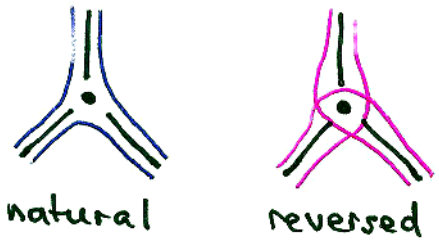


Clockwise-biased tracing algorithm:

at clockwise vertices follow natural rotation;

at anticlockwise vertices follow reversed rotation;

if going along an edge with face f on left, face g on right, current is a, then in derived graph (f,t) is joined to (g,ta) for each $t \in \Gamma$ (currents *always* act 90° clockwise).



Advantage is that we don't have to worry about vertex rotations until we are actually at vertex. Also less complicated when have to deal with twisted edges, later.

Also have *anticlockwise-biased tracing algorithm*: swap clockwise ⟷ anticlockwise, left ⟷ right. Can choose whether to use clockwise-biased or anticlockwise-biased algorithm depending on whether more clockwise or anticlockwise vertices.

All tracing algorithms give same result. In each case, resulting list of edges, destination faces, and (outgoing) currents is called *log* of face.

In above examples, only one face, and edges uniquely identified by current, so can just write log by listing currents:

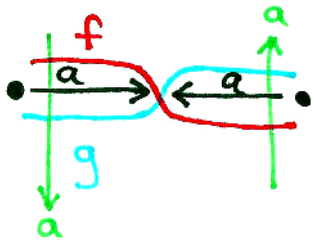
$$8 \quad 9 \quad 7 \quad 4 \quad -2 \quad -9 \quad -1 \quad 5 \quad -3 \quad -7 \quad 2 \quad 6 \quad 1 \quad -8 \quad -5 \quad -6 \quad -4 \quad 3$$

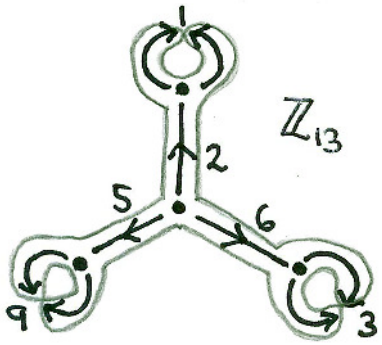
Current graphs with twisted edges

Principle for handling twisted edges: Twisted edges reverse whether we cross over in the middle of the edge or not. To maintain consistency of rules about the way currents act, when we go through a twist on an edge, current must reverse. Current reverses in middle of edge, so twisted edges have same current going in opposite directions on opposite ends.

Modifying standard tracing algorithm: When traverse a twisted edge, cross over in middle if vertices at ends have same direction, do not cross if vertices have opposite directions.

Modifying clockwise- or anticlockwise-biased tracing algorithms: Always cross over in middle of a twisted edge.





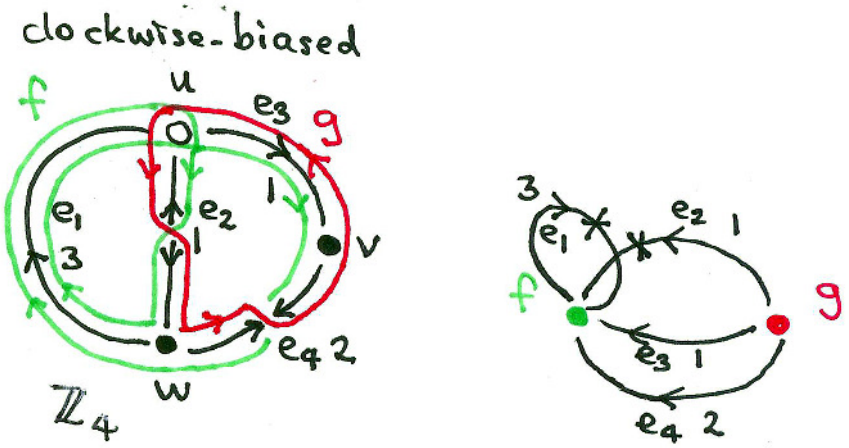
Dealing with lack of global orientation in derived graph: Unless base embedding is actually orientable, cannot trace faces in consistent way, using each edge once in each direction. Give each vertex (f,t) local (clockwise) rotation based on order edges encountered when tracing face f. In derived embedding suppose edge $e' = (f,t)(g,ta)$ is derived from e with f, g on either side. Then e' is twisted if both f and g trace e in the same direction, untwisted if they trace it in opposite directions.

Example: For $\mathbb{Z}_{13}$ current graph above, again just one face, edges uniquely identified by currents. Log of face is

$$2 \quad 1^* \quad -1^* \quad -2 \quad 6 \quad 3^* \quad -3^* \quad -6 \quad 5 \quad 9^* \quad -9^* \quad -5$$

where * denotes twisted edge in derived embedding.

Even with twisted edges derived embedding may be orientable (just as for nonorientable voltage graphs). Will always be nonorientable if base graph is actually nonorientable and current group has odd order.



Example: See figure above. Face logs need to show edge, current applied, face representing other end in derived embedding, and also whether edges are twisted (denoted by *).

f:	e_1^*	e_2^*	e_1^*	e_3	e_4	g:	e_4	e_3	e_2^*
	3	-1	3	-1	-2		2	1	1
	f	g	f	g	g		f	f	f

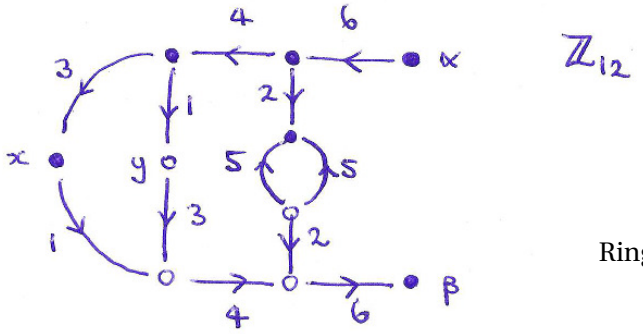
We show the equivalent voltage graph.

Final remarks on current graphs

Map Colour Theorem: Current graphs were used heavily to determine the minimum genus of the complete graph K_n, generally by finding triangular embeddings. This often meant using current graphs with one face ('index one') and with most vertices of degree 3 and net (additive) current 0 (satisfying KCL). Ringel's book on this [Ri] uses current graphs very heavily; presentation sometimes disagrees with modern conventions.

Using both voltages and currents together: Won't go into details, but can use voltages and currents simultaneously on an embedding by applying voltages to edges of gem representation, in such a way that net voltage of each red-blue cycle (corresponding to an edge) is the identity. Equivalent to a construction by Dan Archdeacon [Ar92] that puts voltages on edges of *medial graph*.

Exercise: Consider the current graph from Ringel's Fig. 9.1, shown below. Trace the faces and determine the derived graph. Also determine the distribution of face degrees in the derived embedding.



Ringel's Figure 9.1, slightly modified

In his book Ringel uses this current graph to construct a triangular embedding of $K_{14} - K_2$, the graph obtained by deleting one edge from K_{14}. How does this work? [Hints: keep only one component; add vertices in large faces.]

Exercise: Suppose we have a voltage graph G using group Γ, and we take a vertex v and a constant $g \in \Gamma$. If we right multiply all voltages on edges into v by g, and left multiply all voltages on edges out of v by g^{-1}, then the derived graph stays unchanged, except that vertex (v,h) is now labelled vertex (v,hg) for each $h \in \Gamma$.

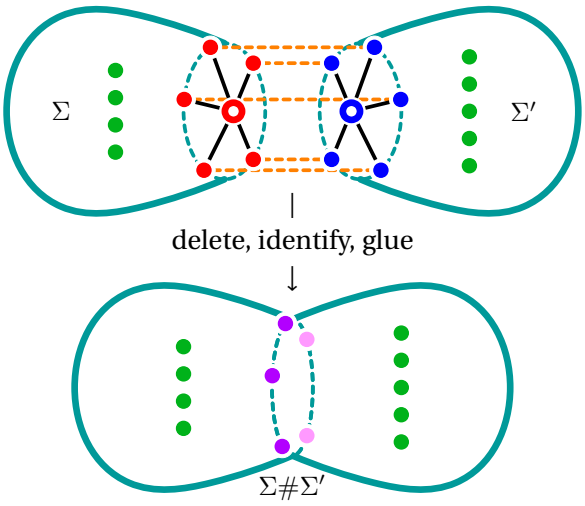
Prove that if we have a gem with assigned voltages, such that the net voltage around every red-blue cycle is the identity, then we can modify the voltages as in the previous paragraph, so that all red and blue edges have identity voltage. [Then all nontrivial voltages are on yellow edges; this is the main step in proving that assigning voltages to gems is effectively the same as Archdeacon's assignment of voltages to the medial graph.]

1.4 Bouchet's diamond sum

Definition: To take the *diamond sum* of two graphs G and G' we take vertices v in G and v' in G', so that $\deg_G(v) = \deg_{G'}(v')$, delete the two vertices, and identify their neighbours together. We denote this as $G \diamond G'$ (where v, v', and the particular identification of their neighbours are understood to be known).

We can extend this to two embeddings Ψ of G and Ψ' of G': when we delete v and v' we cut along a curve through their neighbours, and we glue the surfaces together (to get *connected sum* surface $\Sigma\#\Sigma'$). The neighbours must be identified in rotation order. We denote this as $\Psi \diamond \Psi'$.

Note that if Ψ, Ψ' have Euler genus ε, ε' respectively, then $\Psi \diamond \Psi'$ has Euler genus $\varepsilon + \varepsilon'$. So if both embeddings are orientable, or both are nonorientable, we can just add the genera of the surfaces.



History: Used by Bouchet [Bo78a] in dual form for new proof of minimum genus of $K_{m,n}$, 1978. Primal form used by Mohar, Parsons and Pisanski [MPP85], and Magajna, Mohar and Pisanski [MMP86], mid-1980s. Mohar and Thomassen [MT] give primal version of Bouchet's proof in their book and use diamond notation, hence name 'diamond sum'. General form stated by Kawarabayashi, Stephens and Zha [KSZ04].

Theorem: The minimum genus of an orientable genus embedding of $K_{m,n}$ is $\mathbf{g}(K_{m,n}) = \lceil (m-2)(n-2)/4 \rceil$.

Proof: This was first proved by Ringel, 1965. But we will give a proof based on the diamond sum, following the one in Mohar and Thomassen [MT]. Will use straightforward primal arguments instead of Bouchet's dual arguments.

Euler's formula and the fact that face degrees must be at least 4 (since graph is simple and bipartite) gives a lower bound of $f_0(m,n) = (m-2)(n-2)/4$ on genus, which is achieved if we have a quadrangular embedding (all facial walks are 4-cycles). Since genus is integral, can round up: $\mathbf{g}(K_{m,n}) \geq f(m,n) = \lceil f_0(m,n) \rceil = \lceil (m-2)(n-2)/4 \rceil$.

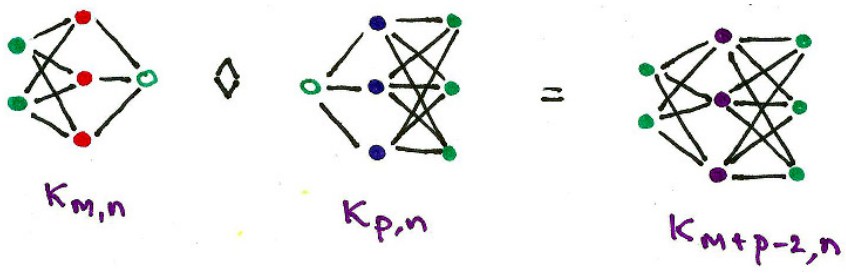
$F(m,n)$ is the statement that $\mathbf{g}(K_{m,n}) = f(m,n)$. We prove it for $m,n \geq 2$ by induction on $m+n$ by constructing an embedding. Note that $F(m,n) \Leftrightarrow F(n,m)$. True if $m=2$ or $n=2$ so suppose $m,n \geq 3$.

Claim D: If $F(m,n)$ and $F(p,n)$ hold and at least one of $f_0(m,n)$ and $f_0(p,n)$ is integral, then $F(m+p-2,n)$ holds.

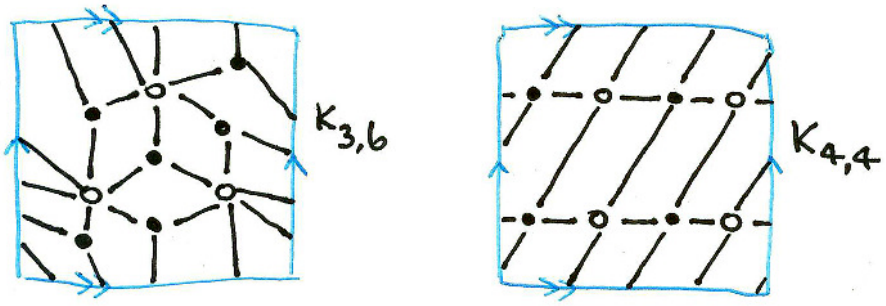
Proof: Take the diamond sum of minimum genus embeddings of $K_{m,n}$ and $K_{p,n}$, deleting a vertex in the first part of each bipartition. The resulting graph is $K_{m+p-2,n}$ with an embedding of genus

$$\begin{aligned} f(m,n)+f(p,n) &= \lceil f_0(m,n) \rceil + \lceil f_0(p,n) \rceil \\ &= \lceil f_0(m,n) + f_0(p,n) \rceil \quad \text{as long as one of } f_0(m,n),\ f_0(p,n) \text{ is integral} \\ &= \lceil \frac{(m-2)(n-2)}{4} + \frac{(p-2)(n-2)}{4} \rceil = \lceil \frac{(m+p-4)(n-2)}{4} \rceil \\ &= \lceil f_0(m+p-2,n) \rceil = f(m+p-2,n) \end{aligned}$$

and so $F(m+p-2,n)$ holds.



Claim B: $F(3,6)$ and $F(4,4)$ hold.
Proof:



Claim S: $F(m,6)$ holds for all m.
Proof: By repeated diamond sums with $K_{3,6}$ we can build up $K_{3,6} \to K4,6 \to K5,6 \to \ldots$, and since $f_0(3,6) = 1$ is integral the result follows from Claim D.

Claim B$^+$: $F(m,n)$ holds if $m,n \leq 6$.
Proof: Use Claim S if $m = 6$ or $n = 6$. Claim B covers $F(4,4)$, and also $F(3,6)$ from which we also get $F(3,3)$, $F(3,4)$ and $F(3,6)$. We get $F(4,5)$ from $F(4,6)$, and $F(5,5)$ from $F(5,6)$.

Now we just use induction. Without loss of generality $m \leq n$ and $n \geq 7$. Now $F(m,n-4)$ and $F(m,6)$ give $F(m,n)$ by Claim D. ■

Nonorientable genus of $K_{m,n}$: In a similar way can prove that $K_{m,n}$ has nonorientable genus $\tilde{\mathbf{g}}(K_{m,n}) = \lceil (m-2)(n-2)/2 \rceil$.

Minimum genus of complete tripartite graphs

Use of diamond sums suggested by Kawarabayashi, Stephens and Zha [KSZ04]. Used by Ellingham, Stephens and Zha [ESZ06] (together with transition graphs and surgical techniques) to find nonorientable genus of all complete tripartite graphs.

Lower bound from Euler's formulas, conjectured to give actual genus: assume $\ell \geq m \geq n$:

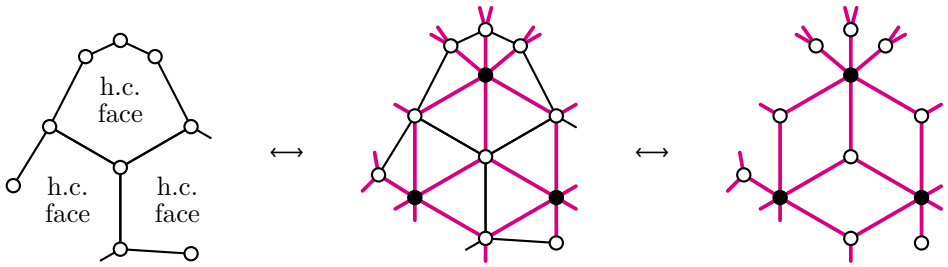
$$\mathbf{g}(K_{\ell,m,n}) \geq \lceil (\ell-2)(m+n-2)/4 \rceil,$$
$$\tilde{\mathbf{g}}(K_{\ell,m,n}) \geq \lceil (\ell-2)(m+n-2)/2 \rceil.$$

Note that lower bound is just genus of $K_{\ell,m+n}$. So if this is really the genus, a minimum genus embedding of $K_{\ell,m,n}$ just consists of a minimum genus embedding of $K_{\ell,m+n}$ with the edges of a $K_{m,n}$ inserted into the faces without changing the surface.

So diamond sum works in a way similar to complete bipartite graphs, but we have extra edges of a $K_{m,n}$ on one side of the diamond sum. Specifically, we can take diamond sum of $K_{\ell,m,n}$ with $K_{p,m+n}$, deleting vertex in first part of partition of each graph. Result is $K_{\ell+p-2,m,n}$. Means that we can start with embeddings of $K_{\ell,m,n}$ for only a small number of values of ℓ close to m (at worst $m, m+1$ in nonorientable case or $m, m+1, m+2, m+3$ in orientable case: stop at first value where no rounding occurs in formula above) and then get all other values of ℓ by diamond sum.

Genus of families of graphs from hamilton cycle embeddings

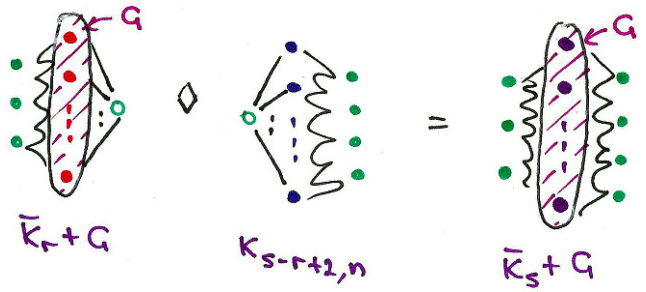
The situation with complete tripartite graphs suggested looking at graphs that look like complete bipartite graphs with some extra edges added on one side of the bipartition. Turns out to be related to embeddings where all facial walks are hamilton cycles.



Hamilton cycle embedding of some *r-regular* n-vertex G
$\longleftrightarrow$ Triangular (hence min. genus) embedding of *join* $\overline{K_r}+G$
$\longleftrightarrow$ Quadrangular (hence min. genus) embedding of $K_{r,n}$.

So the middle step here is a complete bipartite graph with edges added on one side, and the last step tells us that we added edges to a minimum genus embedding of $K_{r,n}$. So we can now proceed as follows:

Hamilton cycle embedding of r-regular n-vertex G
$\updownarrow$
Triangular (hence minimum genus) embedding of join $\overline{K_r}+G$
$|$
because contains min. genus emb. of $K_{r,n}$
$\downarrow$
Minimum genus embedding of join $\overline{K_r}+H$ for any spanning subgraph H of G
$|$
diamond sum with min. genus embedding of $K_{n,s-r+2}$
$\downarrow$
Guaranteed minimum genus embedding of $\overline{K_s}+H$ for all $s \geq r$ and spanning $H \subseteq G$.



Exercise: Prove that the nonorientable genus of $K_{m,n}$ is $\lceil (m-2)(n-2)/2 \rceil$ for $m,n \geq 3$, given that this is known to be a lower bound on the genus. First find an embedding of $K_{3,4}$ in the projective plane N_1. Then use the diamond sum for induction.

Exercise: Suppose we can construct an orientable hamilton cycle embedding of $K_{n,n,n}$ for some particular n. For what family of graphs (as large as possible) can we then use the diamond sum to obtain minimum orientable genus embeddings?

Repeat the question for $K_{n,n,n,n}$.

1.5 Transition graphs

Comment on the name: In retrospect 'transition graph' is not a great name. Should really be called 'global rotation graphs' or something like that: name comes from fact that edges in rotation graph represent 'transitions' between two edges as we pass through a vertex.

General idea: Given an embedded voltage graph, take rotation graph around each vertex R_v. Now for each edge e from u to v identify the vertex of R_u corresponding to an end of e with the vertex of R_v corresponding to the other end of e. Result is actually medial graph of voltage graph. Add some information corresponding to embedding of medial graph, edge twists, voltages.

Will not give formal definition. If desired, see [ESZ06].

Scope and usefulness: This is a general construction, equivalent to embedded voltage graphs (or to current graphs).

We saw that current graphs were more convenient than voltage graphs for finding triangular embeddings of complete graphs. Similarly, transition graphs are more convenient for embeddings of regular complete bipartite graphs $K_{m,m}$ with control over face sizes (usually want faces to be either 4-cycles or bamilton cycles). Play a key role in determining genus of complete tripartite graphs.

Controlled embeddings of $K_{m,m}$

Motivation: For complete tripartite graphs of form $K_{m,m,n}$, may get min. genus embedding from embedding of $K_{m,m}$ with

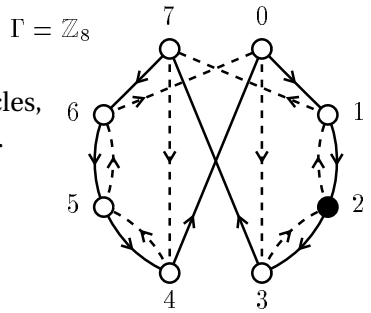
n hamilton cycle faces,
all other faces 4-cycles.

Can then add n new vertices in the hamilton cycle faces.

For joins of edgeless and complete graphs of form $\overline{K_m}+K_m$, may get min. genus embedding from embedding of $K_{m,m}$ with room in faces to add edges of a K_m.

Structure of a transition graph: Construction has

group Γ, directed graph D,
vertices (not edges) labelled by voltages in Γ,
edges partitioned into directed cycles,
each vertex traversed exactly twice by directed cycles,
vertices (not edges) may have twist (solid vertex •).



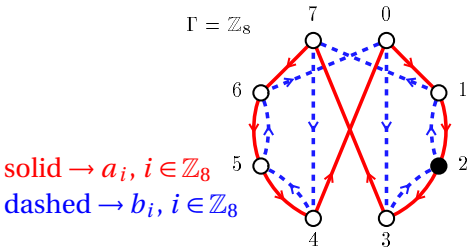
For embeddings of $K_{m,m}$ generally have

group $\Gamma = \mathbb{Z}_m$,
exactly two directed cycles (solid, dashed).

Deriving the embedding:

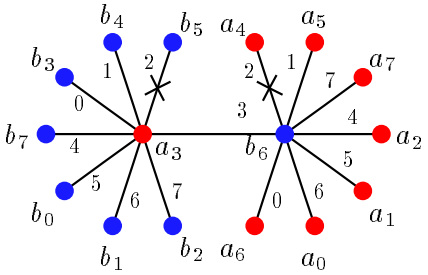
Directed cycle → vertices indexed by Γ,
vertex → class of edges with given "slope",
twisted vertex → twisted edges,
directed cycles show rotations.

Example:



solid → a_i, $i \in \mathbb{Z}_8$
dashed → b_i, $i \in \mathbb{Z}_8$

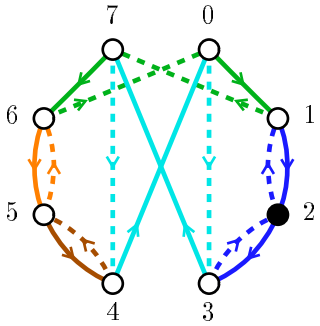
gives as part of derived embedding

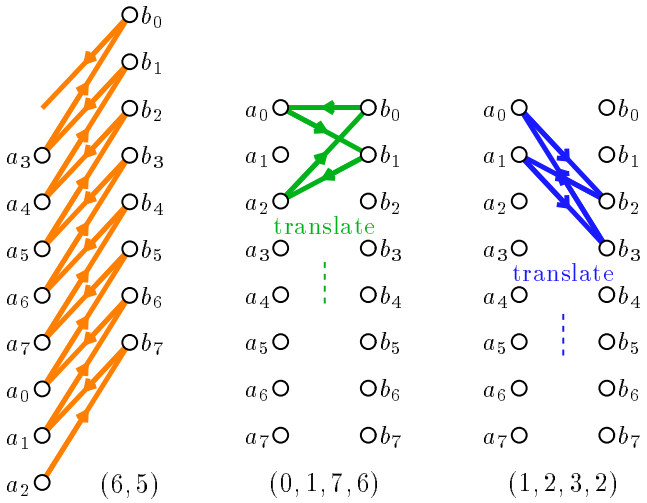


Tracing faces:

Follow edges in transition graph,
switching directed cycles at each vertex,
at twisted vertex also switch directions.

Results: (0,1,7,6), (1,2,3,2), (4,0,3,7), (6,5), (5,4) – give consecutive *slopes* (voltages) of edges in faces.



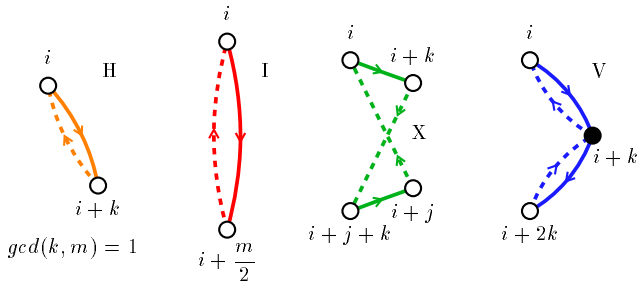


Advantages of transition graphs

- Can be built up from small patterns representing groups of faces of a particular size (H, I, V, X, S, ...).
- Can be used to build whole families of embeddings at once, by making substitutions involving small patterns ($2H \longleftrightarrow V$, $4H \longleftrightarrow 2X$).
- Can be used to build relative (partial) embeddings, then complete with "gadgets" (non-algebraic constructions), when completely algebraic construction is impossible.
- Allow very precise control of emb. structure:
  set up places to add edges;
  set up ways to extend embedding using vertex duplication or special diamond sums.

Building up from small patterns

Easy to build transition graphs from small patterns: specific face sizes.



one ham. cycle $m/2$ 4-cycles m 4-cycles m 4-cycles

Embedding of $K_{8,8}$:

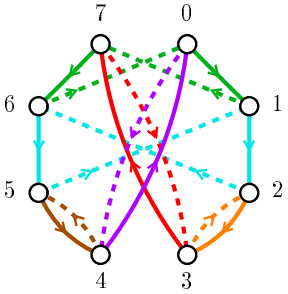
X I

8+8+4+4 = 24 4-cycle faces

H

1 + 1 = 2 ham. cycle faces

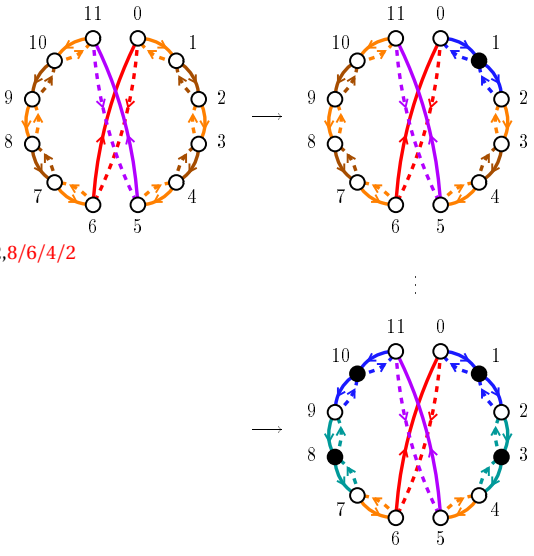
→ min. genus embedding of $K_{8,8,2}$ on S_{12}.



Building families of embeddings

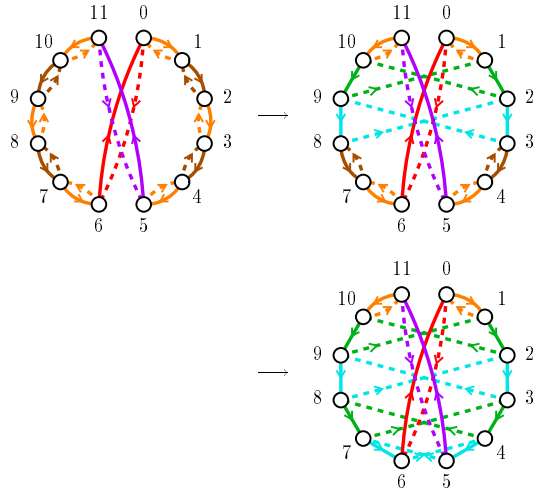
- Switch $2H \to V$ (nonorientable):

  $K_{12,12}$ with 10 ham. cycle faces
  → Ori. min. genus emb. of $K_{12,12,10}$
  which is modified to give
  $K_{12,12}$ with 8, 6, 4, 2 ham. cycle faces
  → Nonori. min. genus emb. of $K_{12,12,8/6/4/2}$



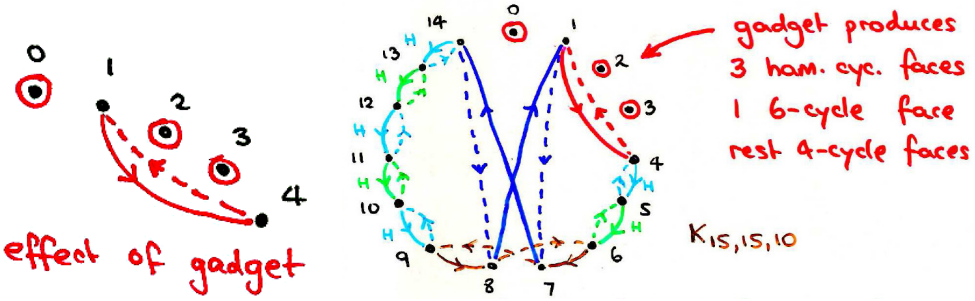
- Switch $4H \to 2X$:

  $K_{12,12}$ with 10 ham. cycle faces
  $\to$ Ori. min. genus emb. of $K_{12,12,10}$
  which is modified to give
  $K_{12,12}$ with 6, 2 ham. cycle faces
  $\to$ Ori. min. genus emb. of $K_{12,12,6/2}$

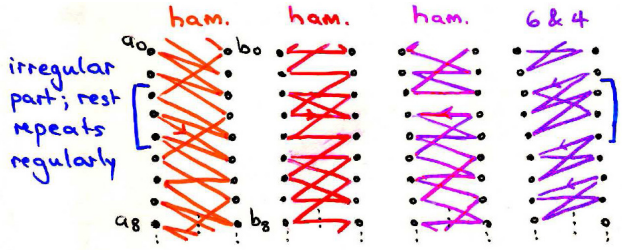


Gadgets

Sometimes there is no purely algebraic way to construct an embedding of $K_{m,m,n}$ using a transition graph. Instead use a *partial transition graph* together with a *gadget*, a set of faces not perfectly symmetric under the action of $\mathbb{Z}_m$, but which easily generalizes.



Detailed faces in gadget:



Special transition graphs (adding edges)

Can also do other things with transition graphs. For example, by controlling the lengths of edges (length of $i \to j$ is $j - i$) we can control which vertices share faces. If we get edges of one type (solid or dashed) with all possible lengths, means vertices in one class share a face with every other vertex in the same class, so can add a complete graph on that side of the bipartition. Used for example to construct orientable minimum genus embeddings of $\overline{K_n} + K_n$ for even n.

Exercise: Find a transition graph that generates a *nonorientable* embedding of $K_{14,14}$ with twelve hamilton cycle faces and all other faces being 4-cycles. [Note: using any V pattern guarantees that you have a nonorientable embedding.]

Now repeat for eleven hamilton cycle faces.

These allow us to get minimum nonorientable genus embeddings of $K_{14,14,12}$ and $K_{14,14,11}$, by adding vertices in the hamilton cycle faces. Can you set up your embeddings so that by using 2H $\longleftrightarrow$ V transformations you can also get minimum genus nonorientable embeddings of $K_{14,14,t}$ for some other values of t? Set up your original embeddings so that you can cover as many t as possible in this way.

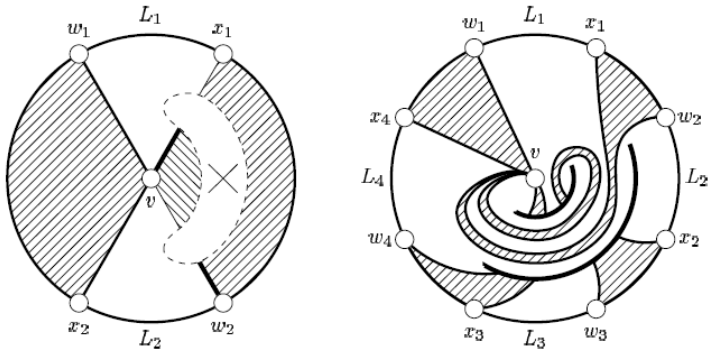
Now (if you are not worn out) repeat for *orientable* embeddings. Besides having different original embeddings, you should use 4H $\longleftrightarrow$ 2X transformations instead of 2H $\longleftrightarrow$ V transformations.

1.6 Surgery

Surgery (cutting and pasting) can be used in many ways. Two very typical ways are for local modification of embeddings and for recursive constructions. Will give illustrations for each.

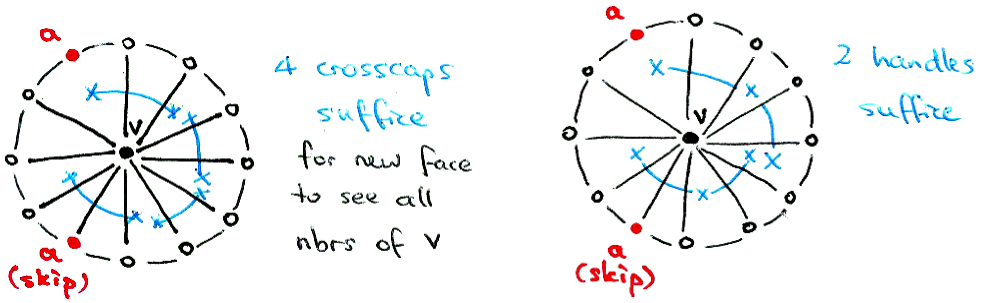
Local modifications

Merging faces around a vertex: Can use a single crosscap to merge two faces around the same vertex into a single face. Similarly, can use a handle to merge three faces around the same vertex into a single face.



By repeating this process we can merge enough faces around a given vertex v into a single face so that we can add into the new face a new vertex v' that is adjacent to all neighbours of v. We call this *duplicating a vertex*. See [ESZ06]; similar ideas also used by other people.

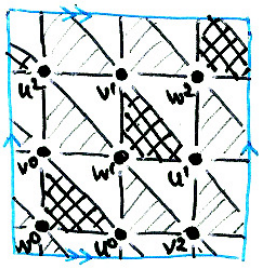
The problem is often that we wish (when constructing minimum genus embeddings) to use only a certain number of crosscaps or handles. We may have to be careful and creative in how we place the crosscaps or handles.



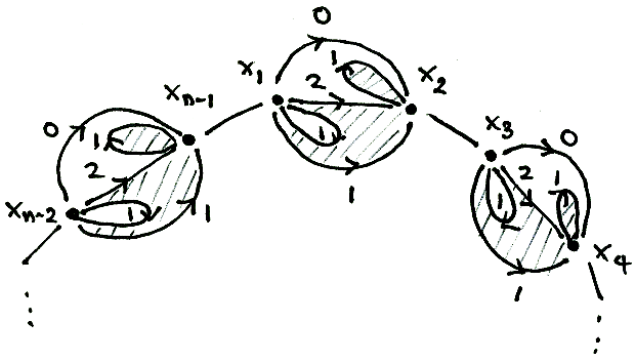
Recursive constructions

'Tripling' for triangulations of complete graphs: Grannell, Griggs and Širáň [GGS98] use 2-face-colourable triangulation of K_n to construct 2-face-colourable triangulation of K_{3n-2}. (Face colouring is important.)

- Take triangulation of K_n, cut out one vertex z, now have K_{n-1} on surface with boundary S.
- Take three copies of S: S^0, S^1, S^2, where v^i on S^i corresponds to v on S, etc.
- For each white triangle $t = (uvw)$ cut out t^0, t^1, t^2 and glue on 2-face-colourable toroidal embedding of $K_{3,3,3}$ with vertex classes $\{u^0, u^1, u^2\}, \{v^0, v^1, v^2\}, \{w^0, w^1, w^2\}$ which has three black triangles $(u^i v^i w^i)$ deleted. Gives all edges of K_{3n-3} except those $x^i y^j$ where $i \neq j$ and xy incident with boundary and black triangle (then no white triangle containing that edge), and edges of form $x^i x^j$ where $i \neq j$.



- Now suppose boundary is $(x_1 x_2 \ldots x_{n-1})$ (where $n-1$ is even) where x_1x_2, x_3x_4, ... are incident with only black triangles. Construct derived embedding from $\mathbb{Z}_3$-voltage graph shown: contains
  cycles $(x_1^i x_2^i \ldots x_{n-1}^i)$ to glue on to boundaries of S^0, S^1, S^2,
  assuming $3 \not| n-1$, hamilton cycle $(x_1^0 x_2^1 x_3^1 x_4^2 \ldots x_{n-1}^0)$ in which to add extra vertex, and all missing edges;



This construction is important: by varying the way the $K_{3,3,3}$ embeddings are glued on, was first construction of large number (c^{n^2}) of nonisomorphic triangular embeddings of given complete graphs K_n [BGGS00].

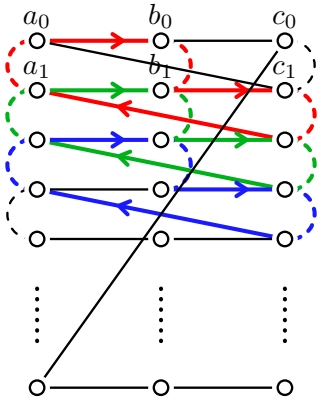
'Doubling' and 'tripling' for hamilton cycle embeddings of complete graphs: Due to Ellingham and Stephens [ES09]/Ellingham and Schroeder [ES14b] Use hamilton cycle embedding of regular complete bipartite/tripartite graph (known) to glue together hamilton cycle embeddings of K_n to get hamilton cycle embedding of K_{2n-2} or K_{3n-3}.

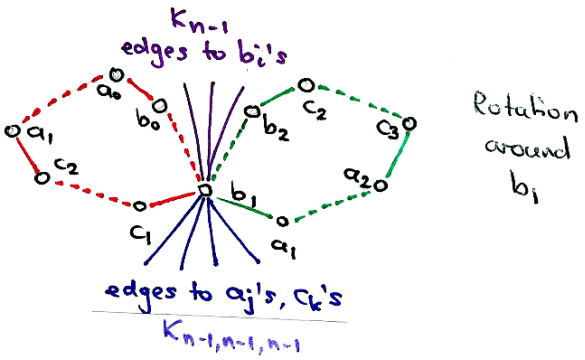
For 'tripling', glue together:

(a) three hamilton cycle embeddings of K_n, each with one vertex deleted, and

(b) one hamilton cycle embedding of $K_{n-1,n-1,n-1}$ with at least one abc-pattern face (which we remove).

Result is hamilton cycle embedding of K_{3n-3}. Rotation around b_1 shown to see how it works.





Exercise: Use adding a crosscap around a vertex to transform the embedding of $K_{4,4}$ on the torus, given in Section 4 above, into an embedding of $K_{4,5}$ on N_3.

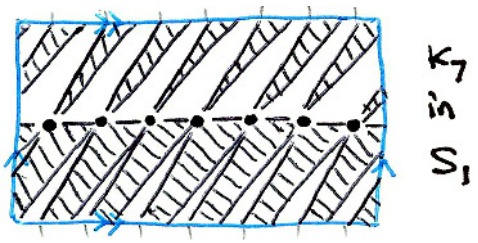
1.7 Connections with design theory

Designs can often be used to help construct embeddings. Often need some kind of extra condition to make sure we get proper rotations.

Biembeddings of Steiner triple systems

If we have a 2-face-colourable triangular embedding of K_n, then each colour class forms a partition of the edges of K_n into triangles. In other words, we have a set of triples chosen from n elements so that every pair occurs in exactly one triple: a *Steiner triple system (STS)*. Altogether this is a *biembedding of Steiner triple systems*.

Example: 2-face-colourable embedding of K_7 on torus, shown below, is a biembedding of the Fano plane (the unique up to isomorphism STS of order 7) with itself.



In general if we just take two arbitrary Steiner triple systems then we do not get an embedding: we have a set of closed walks covering each edge twice, but may not have proper rotations.

If we take two Steiner triple systems T_1 and T_2, not clear when T_1 can be biembedded with something isomorphic to T_2. At least one case known where this cannot be done if we insist that the embedding must be orientable.

Biembeddings of Latin squares

Definition: A *latin square* is an $n \times n$ array of n symbols so that every symbol occurs exactly once in each row and each column.

Suppose we have a 2-face-colourable triangular embedding of a complete tripartite graph $K_{n,n,n}$ with tripartition (A, B, C) where $A = \{a_1, a_2, \ldots, a_n\}$, etc. Take one colour class of faces, then we have a partition of the edges of $K_{n,n,n}$ into triangles. If we interpret a triangle $(a_i b_j c_k)$ as telling us to put symbol k in row i, column j, then we get a latin square for each colour class. Altogether this is a *biembedding of latin squares*. If this exists, the surface is necessarily orientable.

Again, if we take two arbitrary latin squares, it is not clear if we can biembed them. But there is one positive result with very useful consequences.

Definition: A latin square L is *consecutive-row-hamiltonian* if for every two (cyclically) consecutive rows, the permutation we get by mapping symbols in the first row to the symbols in the same column in the second row is a cyclic (hamiltonian!) permutation.

Simple example: Z_n, the addition table of $\mathbb{Z}_n$, is consecutive-row-hamiltonian.

Theorem (Grannell and Griggs [GG08]): Any latin square that is consecutive-row-hamiltonian has a biembedding with something isomorphic to itself (in fact, to itself with all rows shifted up one position).

This was used as part of first construction of n^{an^2} nonisomorphic triangulations of K_n for certain n. Overall construction used ideas related to earlier result giving c^{n^2} such triangulations (mentioned in section on surgery).

Latin squares and hamilton cycle embeddings of complete tripartite graphs

Can also use latin squares to get other embeddings of complete tripartite graphs: ones where all facial walks are hamilton cycles. Need two conditions. First, latin square must be consecutive-entry-hamiltonian (similar to consecutive-row-hamiltonian, and in fact could use that instead). Second, latin square L must have an *orthogonal mate*: another latin square L' such that for every symbol s of L and every symbol s' of L' there is some row and column that contains s in L and s' in L'.

Theorem (Ellingham and Schroeder): An $n \times n$ latin square that is consecutive-entry-hamiltonian and has an orthogonal mate can be used to construct a 2-face-colourable

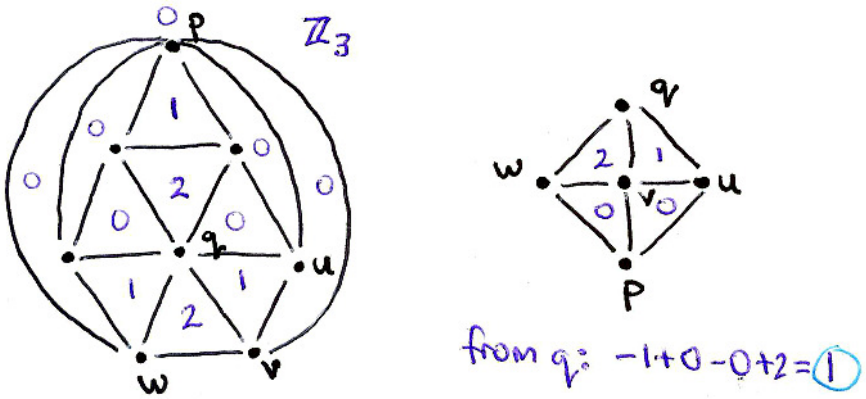
hamilton cycle embedding of $K_{n,n,n}$. Every face has an abc-pattern (useful for tripling construction mentioned in section on surgery). If $n \geq 3$ is not twice a prime then such a latin square exists.

For n odd can again use Z_n, addition table of $\mathbb{Z}_n$. Much trickier for even n.

1.8 Bouchet's covering triangulations

Idea: Lift triangulation of G to triangulation of $G[\overline{K_m}] = G_{(m)}$, graph where we replace each $x \in V(G)$ by m independent vertices (x,i), $i \in \mathbb{Z}_m$, and $(x,i)(y,j) \in E(G_{(m)}) \Leftrightarrow xy \in E(G)$, i.e. each edge is replaced by a copy of $K_{m,m}$. Original paper is [Bo78b].

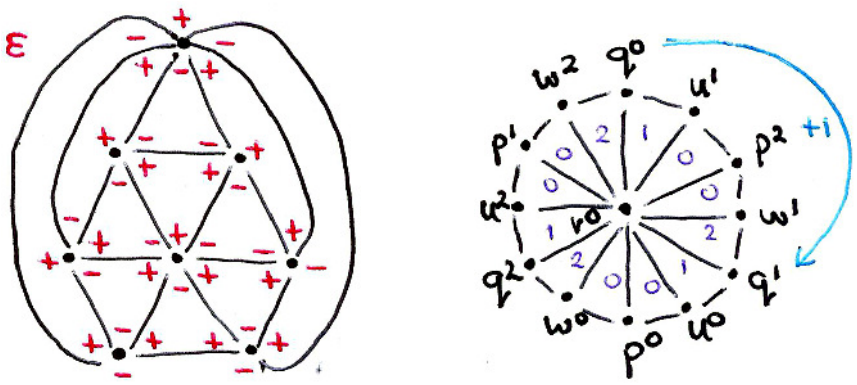
Definition: Suppose G is eulerian (every vertex has even degree) and Ψ is a triangulation of G. Let $T = T(\Psi)$ be the set of triangles of Ψ. An *m-valuation* is a map $\phi : T \to \mathbb{Z}_m$. An m-valuation is *generative* if the alternating sum around every vertex is a generator of $\mathbb{Z}_m$.



Formalization: Make a bit more precise: a corner of the embedding is represented by a (vertex, triangle) pair (x,t). Assign a sign $\varepsilon(x,t) \in \{-1,1\}$ to each corner so that the signs alternate around every vertex. Define

$\overline{\phi}(x) = \sum_{x \in t} \varepsilon(x,t)\phi(t)$

for each vertex x. Then we want every $\overline{\phi}(x)$ to be a generator of $\mathbb{Z}_m$.



Theorem: If ϕ is a generative m-valuation then we have a triangulation of $G_{(m)}$ whose triangles are given by

$$\{((x,i)(y,j)(z,k)) \mid (xyz) \in T,\ i+j+k = \phi(t)\}.$$

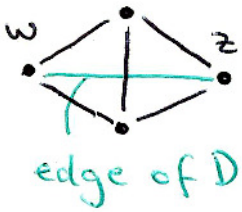
This has the same orientability as the original triangulation.

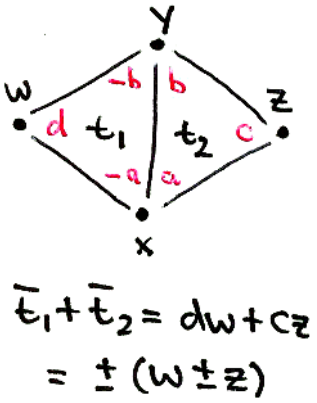
- Clear that we get two triangles containing every $(x,i)(y,j)$, corresponding to the two original triangles (wxy) and (xyz): values of i and j force values of h and k for third vertices (w,h) and (z,k).
- So just need to verify proper rotations. When we follow triangles around a vertex (x,i) from edge $(x,i)(y,j)$ will end up at edge $(x,i)(y,j \pm \overline{\phi}(x))$ after going around x once: since $\overline{\phi}(x)$ generates $\mathbb{Z}_m$, we end up with *all* neighbours of (x,i) after doing this m times.

Finding a generative m-valuation

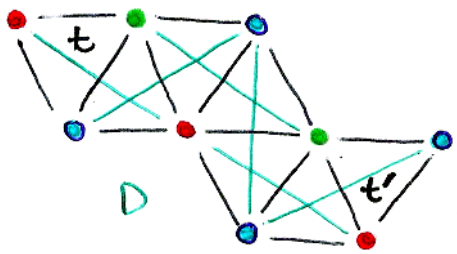
Restate question in more formal algebraic way.

- Consider $\mathbb{Z}_m V$ = formal $\mathbb{Z}_m$-linear combinations of vertices in G, $\mathbb{Z}_m$-module.
- For each triangle t define $\overline{t} = \sum_{x \in t} \varepsilon(x,t)x \in \mathbb{Z}_m V$. Define $\overline{\phi} = \sum_{t \in T} \phi(t)\overline{t}$. ϕ is generative m-valuation if coefficient of $\overline{\phi}$ for vertex x is a generator of $\mathbb{Z}_m$ for all x: in that case say that $\overline{\phi}$ is *generative* element of $\mathbb{Z}_m V$. This coefficient is just what we called $\overline{\phi}(x)$ before: formal $\mathbb{Z}_m$-linear combinations are equivalent to $\mathbb{Z}_m$-valued functions.
- Define submodule $\overline{T}$ generated by $\{\overline{t} \mid t \in T\}$. Want to know if any generative element in $\overline{T}$.
- Depends on structure of *diagonal graph* $D = D(\Psi)$: $V(D) = V(G)$, join w and z if they are in adjacent triangles (wxy) and (xyz).





  - If $wz \in E(D)$ then one of $w+z$, $w-z$ is in $\overline{T}$: call it $\alpha(w,z)$.
  - If u and v are in the same component of D then one of $u+v$, $u-v$ is in $\overline{T}$: again call it $\alpha(u,v)$. (Use induction on previous statement.)
  - So if could partition each component of D into pairs of vertices (u_i, v_i), add up all $\alpha(u_i, v_i)$ and all coefficients ± 1, so have a generative element.

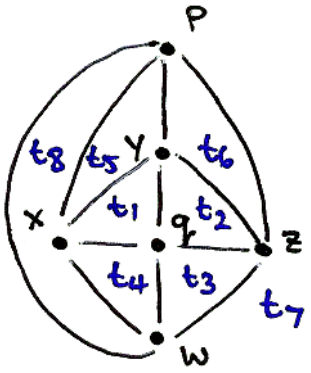


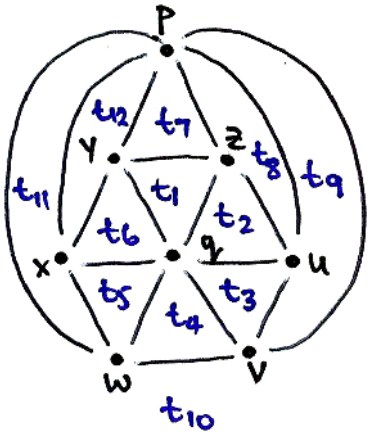
- What do components of D look like? For a given triangle t and $x \in T$, for any other triangle t' there is $x' \in t'$ such that x and x' are in the same component of D. So D has at most three components, and each component contains a fixed number of vertices of each triangle.

  For example, in the octahedron (as shown) there are three components of D with even vertex sets $\{p,q\}$, $\{w,y\}$, $\{x,z\}$. Take

$$\begin{aligned}(\overline{t_1}+\overline{t_5})+(\overline{t_1}+\overline{t_4})+(\overline{t_1}+\overline{t_2}) &= 3t_1+t_2+t_4+t_5\\ &= \pm(p\pm q)\pm(w\pm y)\pm(x\pm z)\end{aligned}$$

  which is generative for any m (even or odd).





Theorem: Suppose m is odd and Ψ is a triangulation of eulerian G. Then Ψ has a generative m-valuation and hence a triangular embedding of $G_{(m)}$ of the same orientability as Ψ.

Proof:

- Fix a triangle $t=(xyz)$. Choose ε values so $\varepsilon(x,t)=\varepsilon(y.t)=\varepsilon(z,t)=1$.
- Partition each component of D into pairs of vertices, as follows:
  - if there is a leftover vertex make sure it is a vertex of t;
  - if a vertex of t is in one of the pairs (u_i,v_i), make sure it is u_i (so its coefficient is definitely 1, not -1).

    If all components of D are even, just add up all $\alpha(u_i,v_i)$ as mentioned above: all coefficients are ± 1.
- If some component of D is odd then adding up all $\alpha(u_i,v_i)$ will leave out some element(s) of t. So add up all $\alpha(u_i,v_i)$ and add $\overline{t}=x+y+z$. Now all coefficients ± 1 except possibly coefficients of 2 for x, y or z: since m is odd, still generative. ■

Example: In $\overline{K_2}+C_6$ as shown, D has three components with vertex sets $\{p,q\}$, $\{u,w,y\}$ and $\{v,x,z\}$. Assuming all ε values of t_1 are $+1$, we take $\overline{t_1}+\alpha(q,p)+\alpha(u,w)+\alpha(v,x)$ which has coefficient 2 for q and coefficient ± 1 for everything else.

Note: As mentioned earlier, if all components of D are even order then works for any m; Bouchet gives other conditions that will guarantee this.

Folded coverings

If we want to extend Theorem above to even m, will be enough to do it for $m=2$, then can use induction for powers of 2 and combine with result for odd m. But it can be shown that it is not always possible to get a generative 2-valuation.

Instead, need to use *folded coverings* [Bo82]. Original coverings have property that two triangles containing given edge $(x,i)(y,j)$ correspond to the two distinct triangles containing xy in G. But for folded covering, may have *fold* on edge $(x,i)(y,j)$: both triangles containing this edge correspond to same original triangle (xyz).

Theorem: Suppose Ψ is a triangulation of eulerian G. Then there is a triangular embedding of $G_{2)}$ of the same orientability as Ψ, obtained by a folded covering.

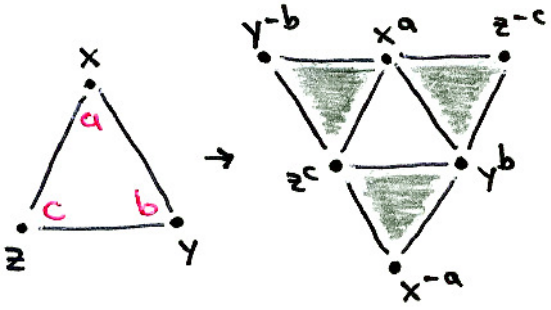
Proof: Assign $\varepsilon(x,t)$ values as previously (± 1 values at corners, alternating around each vertex).

- For each $x \in V(G)$ let $(x,-1)$ and $(x,1)$ be corresponding vertices in $G_{(2)}$.
- Given a triangle $t=(xyz)$ in Ψ with $a=\varepsilon(x,t)$, $b=\varepsilon(y,t)$, $c=\varepsilon(z,t)$, replace by four triangles

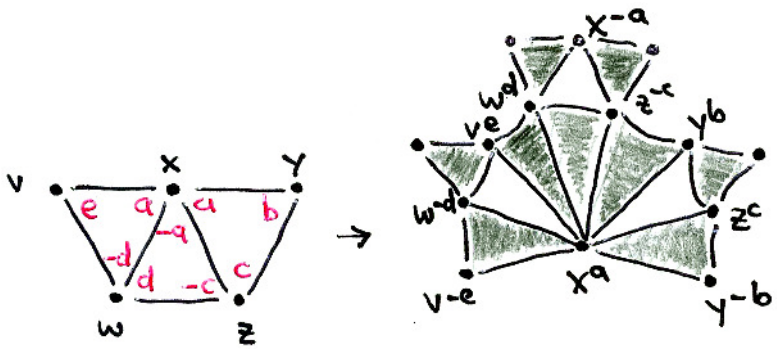
  $((x,a)(y,b)(z,c))$ (primary triangle),

  $((x,-a)(y,b)(z,c))$, $((x,a)(y,-b)(z,c))$, $((x,a)(y,b)(z,-c))$ (three secondary triangles).

  Note that each edge $(x,a)(y,b)$, $(x,a)(z,c)$, $(y,b)(z,c)$ appears in two triangles coming from (xyz) so each of these edges is a fold.



- Each edge occurs in two triangles: suppose we also have original triangle $t'=(wxy)$. Then $(x,a)(y,b)$ occurs in two triangles from $t=(xyz)$; $(x,-a)(y,-b)$ occurs in two triangles from $t'=(wxy)$ (also a fold) because $\varepsilon(x,t')=-\varepsilon(x,t)=-a$ and $\varepsilon(y,t')=-\varepsilon(y,t)=-b$; $(x,a)(y,-b)$ and $(x,-a)(y,b)$ each appear in one triangle from $t=(xyz)$ and one triangle from $t'=(wxy)$ (so not folds).
- Can follow triangles around each vertex $(x,\pm 1)$: close up because original degree of x was even, so have proper rotation.



- Map local orientation of triangles in Ψ to new triangulation: use same orientation for primary triangles, reverse for secondary triangles. Consistent if and only if original orientation consistent. ■

Other important results by Bouchet and coauthors

Theorem: If Ψ is a triangulation of an eulerian complete multipartite graph then $G_{(m)}$ has a triangulation of the same orientability as Ψ, obtained using a generative m-valuation, for all $m \geq 2$. Proof shows that we can avoid odd order components of diagonal graph when m is even.

Theorem: If p is an odd prime and Ψ is a triangulation of a graph G such that Ψ^* (the dual of Ψ) has a nowhere zero p-flow then there is a triangular embedding of $G_{(p)}$ of the same orientability as Ψ.

Note: Since all 2-connected graphs have nowhere zero 6-flows (Seymour), can always do this for $p \geq 7$. If 5-flow conjecture is true, would always work for $p = 5$, too. In special cases can work for $p = 3$ or 5 (e.g., see below).

Theorem: If Ψ is a triangulation of a 4-colourable graph $G \ncong K_4$, then we get a triangular embedding of $G_{(m)}$ of the same orientability as Ψ for $m = 3$ and hence (by repetition, and using the fact that a 4-face-colourable graph has a nowhere zero 4-flow) for all odd m.

Non-triangular embeddings

Bouchet's constructions are for triangulations. But can use, perhaps in modified form, if convert other embeddings into triangulations by adding extra edges or vertices. A couple of examples:

1. Lifting embeddings where all faces have even lengths, paper by Bouchet. First add a new vertex inside each face so we have an Eulerian triangulation. Now find an m-valuation ϕ so that values/coefficients of $\overline{\phi}$ are generators of $\mathbb{Z}_m$ for original vertices, but are 0 for new vertices.
2. In some cases it makes sense to just directly apply Bouchet's results after converting to a triangulation. For example, Ellingham and Schroeder [ES12] used Bouchet's results to help construct hamilton cycle embeddings of regular complete tripartite graphs:

   hamilton cycle embedding of $K_{t,t,t}$

   $\rightarrow$ triangulation of $K_{2t,t,t,t}$ (add vertex in each face)

   and apply Bouchet lifting to get

   triangulation of $K_{2mt,mt,mt,mt}$

→ hamilton cycle embedding of $K_{mt,mt,mt}$ (delete first vertex class).

1.9 References

[Ar92] Dan Archdeacon, The medial graph and voltage-current duality, Discrete Math. 104 (1992) 111-141.

[BGGS00] C. P. Bonnington M. J. Grannell, T. S. Griggs and J. Širáň, Exponential families of non-isomorphic triangulations of complete graphs, J. Combin. Theory Ser. B 78 (2000) 169-184.

[BL] C.Paul Bonnington and Charles H.C. Little, The Foundations of Topological Graph Theory, Springer, 1995.

[Bo78a] A. Bouchet, Orientable and nonorientable genus of the complete bipartite graph, J. Combin. Theory Ser. B 24 (1978) 24-33.

[Bo78b] A. Bouchet, Triangular imbeddings into surfaces of a join of equicardinal independent sets following an Eulerian graph. Theory and applications of graphs (Proc. Internat. Conf., Western Mich. Univ., Kalamazoo, Mich., 1976), pp. 86-115, Lecture Notes in Math., 642, Springer, Berlin, 1978.

[Bo82] A. Bouchet, Constructions of covering triangulations with folds. J. Graph Theory 6 (1982) 57–74.

[ES12] M. N. Ellingham and Justin Z. Schroeder, Nonorientable hamilton cycle embeddings of complete tripartite graphs, Discrete Math. 312 (2012) 1911-1917.

[ES14a] M. N. Ellingham and Justin Z. Schroeder, Orientable hamilton cycle embeddings of complete tripartite graphs I: latin square constructions, J. Combin. Designs 22 (2014) 71-94.

[ES14b] M. N. Ellingham and Justin Z. Schroeder, Orientable hamilton cycle embeddings of complete tripartite graphs II: voltage graph constructions and applications, J. Graph Theory, to appear (available online).

[ES09] M. N. Ellingham and D. Christopher Stephens, The orientable genus of some joins of complete graphs with large edgeless graphs, Discrete Math. 309 (2009) 1190-1198.

[ESZ06] M. N. Ellingham, Chris Stephens and Xiaoya Zha, The nonorientable genus of complete tripartite graphs, J. Combin. Theory Ser. B 96 (2006) 529-559.

[GG08] M. J. Grannell and T. S. Griggs A lower bound for the number of triangular embeddings of some complete graphs and complete regular J. Combin. Theory Ser. B 98 (2008) 637-650.

[GGS98] M. J. Grannell, T. S. Griggs and J. Širáň, Face 2-colourable triangular embeddings of complete graphs, J. Combin. Theory Ser. B 74 (1998) 8-19.

[GT] J. L. Gross and T. W. Tucker, Topological Graph Theory (reprint edition), Dover, Mineola, NY, 2001.

[KSZ04] Ken-ichi Kawarabayashi, Chris Stephens and Xiaoya Zha, Orientable and nonorientable genera of some complete tripartite graphs, SIAM J. Discrete Math. 18 (2004) 479-487.

[MMP86] Z. Magajna, B. Mohar, and T. Pisanski, Minimal ordered triangulations of surfaces, J. Graph Theory 10 (1986) 451-460.

[MPP85] B. Mohar, T. D. Parsons, and T. Pisanski, The genus of nearly complete bipartite graphs, Ars Combin. 20 (1985) 173-183.

[MT] Bojan Mohar and Carsten Thomassen, Graphs on Surfaces, Johns Hopkins University Press, Baltimore, 2001.

[Ri] G. Ringel, Map Color Theorem, Springer, Berlin, 1974.

Chapter 2

Combinatorial Designs

Mariusz Meszka
AGH University of Science and Technology, Poland

SUMMARY

The roots of combinatorial design theory, date from the 18th and 19th centuries, may be found in statistical theory of experiments, geometry and recreational mathematics. Design theory rapidly developed in the second half of the twentieth century to an independent branch of combinatorics. It has deep interactions with graph theory, algebra, geometry and number theory, together with a wide range of applications in many other disciplines. Most of the problems are simple enough to explain even to non-mathematicians, yet the solutions usually involve innovative techniques as well as advanced tools and methods of other areas of mathematics. The most fundamental problems still remain unsolved.

2.1 Balanced incomplete block designs

A *design* (or *combinatorial design*, or *block design*) is a pair $(V, \mathscr{B})$ such that V is a finite set and $\mathscr{B}$ is a collection of nonempty subsets of V. Elements in V are called *points* while subsets in $\mathscr{B}$ are called *blocks*.

One of the most important classes of designs are balanced incomplete block designs.

Definition 1. A *balanced incomplete block design* (BIBD) is a pair $(V, \mathscr{B})$ where $|V| = v$ and $\mathscr{B}$ is a collection of b blocks, each of cardinality k, such that each element of V is contained in exactly r blocks and any 2-element subset of V is contained in exactly λ blocks. The numbers v, b, r, k an λ are *parameters* of the BIBD.

Since $r = \frac{\lambda(v-1)}{k-1}$ and $b = \frac{\lambda v(v-1)}{k(k-1)}$ must be integers, the following are obvious arithmetic necessary conditions for the existence of a BIBD(v,b,r,k,λ):
(1) $\lambda(v-1) \equiv 0 \pmod{k-1}$,
(2) $\lambda v(v-1) \equiv 0 \pmod{k(k-1)}$.
Parameter sets that satisfy (1) and (2) are called *admissible*.

The five parameters: v, b, r, k, λ are not independent; three of them: v, k and λ uniquely determine the remaining two as $r = \frac{\lambda(v-1)}{k-1}$ and $b = \frac{vr}{k}$. Hence we often write (v,k,λ)-*design* (or $(v,k,\lambda)-$BIBD) to denote a BIBD(v,b,r,k,λ).

Example 1. A $(7,3,1)-$BIBD (the "Fano plane"):
$V = \{0,1,\ldots,6\}$,
$\mathscr{B} = \{\{0,1,2\}, \{0,3,4\}, \{0,5,6\}, \{1,3,5\}, \{1,3,6\}, \{2,3,6\}, \{2,4,5\}\}$.

Example 2. A $(11,5,2)-$BIBD:
$V = \{0,1,\ldots,10\}$,
$\mathscr{B} = \{\{0,1,2,6,9\}, \{0,1,5,8,10\}, \{0,2,3,4,8\}, \{0,3,5,6,7\}, \{0,4,7,9,10\}, \{1,2,3,7,10\}, \{1,3,4,5,9\}, \{1,4,6,7,8\}, \{2,4,5,6,10\}, \{2,5,7,8,9\}, \{3,6,8,9,10\}\}$.

The necessary conditions are also sufficient for the existence of a (v,k,1)-BIBD with small k:
when $k = 2$, $v \geq 2$,
when $k = 3$, $v \equiv 1,3 \pmod 6$,
when $k = 4$, $v \equiv 1,4 \pmod{12}$,
when $k = 5$, $v \equiv 1,5 \pmod{20}$.
For $k = 6$, a $(v,6,1)-$BIBD exists if $v \equiv 1,6 \pmod{15}$ and $v \neq 16,21,36,46$; $51,61,81,166,226,231,256,261,286,316,321,346,351,376,406,411,436,441,471,501,561,591,616,646,651,676,771,796,801$. In the case of the orders $v = 16,21,36,46$ it is proven that a $(v,6,1)-$BIBD does not exists, in the case of remaining 29 orders the existence problem is still unsettled.

A convenient way to represent a BIBD, other than a list of its blocks, is an incidence matrix. The *incidence matrix* of a $(v,k,\lambda)-$BIBD $(V,\mathscr{B})$, where $V = \{x_i : 1 \leq i \leq v\}$ and $\mathscr{B} = \{B_j : 1 \leq j \leq b\}$, is a $v \times b$ matrix $A = (a_{ij})$, in which $a_{ij} = 1$ when $x_i \in B_j$ and $a_{ij} = 0$ otherwise.

Lemma 1. *If A is an incidence matrix of a $(v,k,\lambda)-$*BIBD*, then $AA^T = (r-\lambda)I + \lambda J$, where I is a $v \times v$ identity matrix and J is a $v \times v$ all ones matrix.*

Theorem 2 (Fisher's inequality)**.** *If a $(v,k,\lambda)-$*BIBD *exists with $2 \leq k < v$, then $b \geq v$.*

This result, for instance, guarantees that a $(21,6,1)-$BIBD cannot exist, since $b = 14 < 21 = v$, even though the above arithmetic necessary conditions are satisfied.

A BIBD is called *symmetric* if $v = b$ (and $r = k$). The most fundamental necessary condition for the existence of symmetric designs is due to Bruck, Ryser and Chowla.

Theorem 3 (Bruck-Ryser-Chowla)**.** *Let v, k and λ be integers satisfying $\lambda(v-1) = k(k-1)$ and for which there exists a symmetric $(v,k,\lambda)-$*BIBD.
(1) *If v is even, then $n = k-\lambda$ is a square.*
(2) *If v is odd, then the equation $z^2 = nx^2 + (-1)^{(v-1)/2}\lambda y^2$ has a solution in integers x, y, z not all zero.*

The *dual* of D is a design $D^* = (\mathscr{B}, V)$, where $\mathscr{B}$ corresponds to a set of elements and V to a set of blocks, such that $B \in \mathscr{B}$ is an element contained in $v \in V$ if and only if v is contained in B in D. Thus, if M is an incidence matrix of D, then M^T is an incidence matrix of D^*.

Remark. *The dual of a* BIBD *is a* BIBD *if and only if the* BIBD *is symmetric.*

Also, the parameters of a symmetric design and its dual are the same, yet they are not necessarily isomorphic.

All necessary conditions specified above (taken together) are still not sufficient for the existence, for instance, of a symmetric $(111,111,11,11,1)-$BIBD. One can easily check the set of parameters satisfies all conditions (including Fisher's inequality and Bruck-Ryser-Chowla theorem) but such design does not exist, what was proven by a detailed structural analysis together with an exhaustive computational search. The general existence question for BIBD's remains crucial open problem for infinitely many sets of parameters.

Definition 2. Two designs, $(V_1, \mathscr{B}_1)$ and $(V_2, \mathscr{B}_2)$, are *isomorphic* if there exists a bijection $\alpha : V_1 \mapsto V_2$ such that for any $B_1 \in \mathscr{B}_1$ there exists $B_2 \in \mathscr{B}_2$, where $B_2 = \{\alpha(x_i) : x_i \in B_1\}$.

An *automorphism* is an isomorphism from a design to itself. The set of all automorphisms of a design forms a group called the *full automorphism group*. An *automorphism group* of a design is any subgroup of its full automorphism group. In particular, a $(v,k,\lambda)-$BIBD is *cyclic* if it admits a cyclic group of order v as its automorphism group.

Specifying an automorphism group allows sometimes to construct a design in much easier way. Then it is enough to select a set of *base* blocks which are representatives of each orbit of blocks under the prescribed automorphism group. All remaining blocks are obtained by action of the group on these base blocks.

Let G be a group of order v. A k-element subset D of G is a (v,k,λ)-*difference set* if every non-zero element of G has exactly λ representations as a difference $d - d'$ with elements from D.

Theorem 4. *A set $D = \{d_1, d_2, \ldots, d_k\}$ of k residues modulo v is a (v,k,λ)-difference set if and only if the sets $B_i = \{d_1 + i, d_2 + i, \ldots, d_k + i\} \bmod v$, $i = 0,1,\ldots,v-1$ form blocks of a cyclic $(v,k,\lambda)-$*BIBD.

Example 3. $\{0,1,3,9\}$ is a $(13,4,1)$-difference set in the group $\mathbb{Z}_{13}$.
Thus $\{0,1,3,9\}$ is the base block of a cyclic $(13,4,1)-$BIBD.

The concept of a difference set may be extended to a larger number of sets. Let G be a

group of order v. A collection $\mathscr{D} = \{D_1, D_2, \ldots D_s\}$ of k-element subsets of G, where $D_i = \{d_1^i, d_2^i, \ldots d_k^i\}$, $i = 1,2,\ldots s$, forms a (v,k,λ)-*difference family* if every non-zero element of G occurs exactly λ times as a difference $d_i^p - d_j^p$.

Theorem 5. *If a set $\mathscr{D} = \{D_1, D_2, \ldots, D_s\}$ is a (v,k,λ)-difference family over the cyclic group G, then $Orb_G(D_1) \cup Orb_G(D_2) \cup \ldots \cup Orb_G(D_s)$ is the collection of blocks of a cyclic $(v,k,\lambda)-$* BIBD.

Example 4. $\{\{0,2,10,15,19,20\}, \{0,3,7,9,10,16\}\}$ is a $(21,6,3)$-difference family in the group $\mathbb{Z}_{21}$.

Let G be a group of order $v-1$. A collection $\mathscr{D} = \{D_1, D_2, \ldots D_s\}$ of k-element subsets of $G \cup \{\infty\}$, is a *1-rotational* (v,k,λ)-*difference family* if every non-zero element of $G \cup \{\infty\}$ occurs exactly λ times as a difference $d_i^p - d_j^p$.

Theorem 6. *If a set $\mathscr{D} = \{D_1, D_2, \ldots, D_s\}$ is a 1-rotational (v,k,λ)-difference family over the group G, then $Orb_G(D_1) \cup Orb_G(D_2) \cup \ldots \cup Orb_G(D_s)$ is the collection of blocks of a $(v,k,\lambda) -$* BIBD *admitting an automorphism group fixing one point and acting sharply transitively on the other points.*

Example 5. $\{\{0,1,3\}, \{0,1,5\}\}, \{0,2,5\}\}, \{0,4,\infty\}\}$ is a 1-rotational $(12,3,2)$-difference family.

The concept of a difference family has been generalized by Bose to form a basis of a method that is called the *method of pure and mixed differences*. Let G be an additive abelian group and let T be a t-element set. Consider the set $V = G \times T$. For any two elements $(x,i) \neq (y,j)$ of V, the differences arising from this pair may be of two kinds:
(1) if $i = j$ then $\pm(x-y)$ is a *pure* difference of class i
(2) if $i \neq j$ then $\pm(x-y)$ is a *mixed* difference of class $i\,j$.
A pure difference of any class may equal to any nonzero element of G while a mixed difference may equal to any element of G.

Suppose that there exists a collection of k-element sets $\mathscr{D} = \{D_1, D_2, \ldots, D_s\}$ such that every nonzero element of G occurs exactly λ times as a pure difference of class i for each $i \in T$, and moreover every element of G occurs exactly λ times as a mixed difference of class $i\,j$ for all $i, j \in T$, $i \neq j$. Then the sets in $\mathscr{D}$ form a *basis* of a $(v,k,\lambda) -$ BIBD $(V, \mathscr{B})$, where $\mathscr{B} = \{D_i + g : g \in G, i = 1,2,\ldots s\}$.

Example 6. Let $G = \mathbb{Z}_5$ and $T = \{1,2\}$.
$\mathscr{D} = \{\{0_1, 2_1, 3_1, 3_2\}, \{0_1, 2_2, 3_2, 4_2\}\}, \{0_1, 1_1, 0_2, 2_2\}\}$ is a basis for a $(10,4,2) -$ BIBD.

Example 7. Let $G = \mathbb{Z}_3$ and $T = \{1,2,3\}$.
$\mathscr{D} = \{\{0_1, 1_1, 0_2\}, \{0_2, 1_2, 0_3\}, \{0_1, 0_3, 1_3\}\}, \{0_1, 1_2, 2_3\}\}$ is a basis for a $(9,3,1) -$ BIBD.

The above construction may be extended by adding one fixed point.

Example 8. Let $V = (\mathbb{Z}_7 \times \{1,2\}) \cup \{\infty\}$.
$\mathscr{D} = \{\{0_1, 1_1, 3_1\}, \{0_1, 0_2, 1_2\}, \{0_1, 2_2, 4_2\}\}, \{0_1, 3_2, 6_2\}, \{0_1, 4_2, \infty\}\}$ is a basis for a $(15,3,1) -$ BIBD.

A *complement* of a design $(V, \mathscr{B})$ is a design $(V, \overline{\mathscr{B}})$, where $\overline{\mathscr{B}} = \{V \setminus B : B \in \mathscr{B}\}$. Thus a complement of a BIBD(v,b,r,k,λ) is a BIBD$(v, b, b-r, v-k, b-2r+\lambda)$. A *supplement* of a BIBD(v,b,r,k,λ) is a BIBD obtained by taking all k-subsets which are not in $\mathscr{B}$ as blocks; in this way we get a BIBD$(v, \binom{v}{k} - b, \binom{v-1}{k-1} - r, k, \binom{v-2}{k-2} - \lambda)$.

A design $(V', \mathscr{B}')$ is a subdesign of $(V, \mathscr{B})$ if $V' \subset V$ and $\mathscr{B}' \subset \mathscr{B}$.

Given a design $D = (V, \mathscr{B})$, a *block intersection graph* $G(D)$ is a graph with the vertex set $\mathscr{B}$ and the edge set $\{\{B_i, B_j\} : B_i \cap B_j \neq \emptyset\}$. In particular, for a $(v,k,1)-$BIBD, $G(D)$ is strongly regular.

Exercise 1.
(1) Construct a $(6,3,2)-$BIBD.
(2) Construct a $(13,4,1)-$BIBD.

Exercise 2.
Find an isomorphism for the Fano plane given in Example 1 and its dual.

Exercise 3.
Prove that Fano plane is unique up to automorphism. Determine the order of its full automorphism group.

Exercise 4.
Find a $(41,5,1)$-difference family in the group $\mathbb{Z}_{41}$.

Exercise 5.
Construct a cyclic $(21,3,1)-$BIBD.

Exercise 6.
Given a BIBD$(v,b,r,k,1)$, determine the parameters (i.e., order, size, degree, clique number, the number of common neighbors for each pair of adjacent vertices and for each pair of nonadjacent vertices) of its block intersection graph.

2.2 Latin squares

Definition 3. A *latin square* of *order* n (or *side* n) is an $n \times n$ array in which each cell contains a single symbol from an n-element set S, such that each symbol occurs exactly once in each row and exactly once in each column.

The nature of symbols in S is of no importance so usually we take $S := \{1, 2, \ldots, n\}$.

Definition 4. A *quasigroup* is an algebraic structure $(Q, \circ)$, where Q is a set and $\circ$ is a binary operation on Q such that the equations $a \circ x = b$ and $y \circ a = b$ have unique solutions for every pair of elements a, b in Q. If Q is finite, then $|Q| = n$ is the *order* of the quasigroup.

A latin square can be viewed as a multiplication table of a quasigroup with the headline and sideline removed. Thus latin squares and quasigroups are equivalent combinatorial objects and we may use these two terms interchangeably.

Example 9. Latin square of order 4 and its corresponding quasigroup of order 4.

1	2	4	3
3	4	2	1
4	1	3	2
2	3	1	4

$\circ$	1	2	3	4
1	1	2	4	3
2	3	4	2	1
3	4	1	3	2
4	2	3	1	4

A latin square L of side n is *commutative* (or *symmetric*) if $L(i,j) = L(j,i)$ for all $1 \leq i,j \leq n$. L is *idempotent* if $L(i,i) = i$ for all $1 \leq i \leq n$. A latin square L' of even order $n = 2k$ is *half-idempotent* if $L'(i,i) = i$ and $L'(k+i,k+i) = i$ for all $1 \leq i \leq k$.

The existence of a latin square of order n is equivalent to the existence of a one-factorization of the complete bipartite graph $K_{n,n}$. Moreover, the existence of a commutative idempotent latin square of order n is equivalent to the existence of a one-factorization of the complete graph K_n.

A latin square is in *standard form* (or *normalized*) if both its first column and first row contain consecutive symbols in an increasing order.

Two latin squares, L and L', of order n are *isotopic* (or *equivalent*) if there are three bijections from the rows, columns and symbols of L to the rows, columns and symbols, respectively, of L', that map L to L'. Latin squares L and L' are *isomorphic* if there exists a bijection $\varphi : S \mapsto S$ such that $\varphi(L(i,j)) = L'(\varphi(i),\varphi(j))$ for every $i,j \in S$, where S is not only the set of symbols of each square but also the indexing set for the rows and columns of each square.

Latin squares are completely enumerated for small orders.

n	number of non-isomorphic latin squares	number of distinct latin squares
2	1	1
3	1	2
4	2	24
5	2	1,334
6	17	1,128,960
7	324	12,198,297,600
8	842,227	2,697,818,265,354,240
9		15,224,734,061,278,915,461,120
10		2,750,892,211,809,148,994,633,229,926,400
11		19,464,657,391,668,924,966,616,671,344,752,852,992,000

Two latin squares, L and L', of order n are *orthogonal* if the n^2 ordered pairs $(L(i,j), L'(i,j))$ are all distinct. A set of latin squares $L_1, L_2, \ldots, L_m$ is *mutually orthogonal* (or a set of MOLS(n)) if for every $1 \leq i < j \leq m$, L_i and L_j are orthogonal.

Example 10. A set of three MOLS(4):

1	2	3	4
4	3	2	1
2	1	4	3
3	4	1	2

1	2	3	4
3	4	1	2
4	3	2	1
2	1	4	3

1	2	3	4
2	1	4	3
3	4	1	2
4	3	2	1

In any latin square belonging to some set of MOLS(n), relabeling symbols does not affect to the orthogonality.

Theorem 7. *A pair of orthogonal latin squares of order n exists for all n other than* 2 *and* 6 *(for which no such pair exists).*

Construction of a pair of orthogonal latin squares of odd order n.
Let $S=\mathbb{Z}_n$. Then $L_1(i,j)=(i+j) \bmod n$ and $L_2(i,j)=(i-j) \bmod n$.

Let $N(n)$ denote the largest number of latin squares in a set of MOLS(n).

Remark. *For every n, $1 \leq N(n) \leq n-1$.*

Theorem 8. *If $q=p^k$ is a prime power, then $N(q)=q-1$.*

Construction of a set of n-1 MOLS **of order $q=p^k$, where p is a prime.**
Let $\mathbb{F}_q$ be a finite field of order q. Let $\alpha_0,\alpha_1,\ldots,\alpha_{q-1}$ be elements of $\mathbb{F}_q$, where α_0 is a zero element. For each nonzero element α_r ($r \neq 0$) in $\mathbb{F}_q$, define a latin square $L_r(i,j)=\alpha_r \times \alpha_i + \alpha_j$.

Determining the value of $N(n)$ remains one of the most foremost problems in combinatorics. For instance, it is known that $N(3) \geq 3$ for all $n \neq 2,3,6$ and possibly 10.

Definition 5. A *partial latin square* of order n is an $n \times n$ array in which some cells are empty and some are filled with elements of S, such that each element of S appears in every row and every column at most once.

Theorem 9. *Any partial latin square of order n which has at most $n-1$ cells occupied can be completed to a latin square.*

Deciding whether a partial latin square can be completed is an NP-complete problem, even if there are no more than 3 unfilled cells in any row or column.

Definition 6. A *latin rectangle* of size $m \times n$ ($m \leq n$) is an $m \times n$ array with entries from a set S of cardinality n such that every row is a permutation of S and every column contains no repetition.

Theorem 10. *If L is an $m \times n$ latin rectangle, then one can append $n-m$ further rows to L so that the resulting array in a latin square.*

Definition 7. Let a, b and n be positive integers with $a \times b = n$. Let an $n \times n$ array be partitioned into disjoint $a \times b$ *regions*. An (a,b)-*Sudoku latin square* is a latin square on the set $\{1,2,\ldots,n\}$ where each region contains all of the symbols. An (a,b)-*Sudoku critical set* of size k is a partial latin square P with k nonempty cells that may be completed in exactly one way to an (a,b)-Sudoku latin square, but removal of any of the filled cells from P destroys the uniqueness of a completion.

Example 11. A $(3,3)$-Sudoku critical set of size 17:

4 2		1
8 1	5 9	4 7 3
3 5	4 1 8 6	2

(3,3)-Sudoku critical sets are known for all sizes from 17 to 35. For instance, the existence of a (3,3)-Sudoku critical set is still unsettled for the size 16. The number of distinct (n,n)-Sudoku latin squares for $n=1$, 2 and 3 is 1, 288 and 6,670,903,752,021,072,936,960, respectively. The number of inequivalent (n,n)-Sudoku latin squares for $n=1$, 2 and 3 is 1, 2 and 5,472,730,538, respectively.

Exercise 7.
(1) Find an idempotent commutative latin square of order 5.
(2) Find a half-idempotent commutative latin square of order 6.

Exercise 8.
Construct a set of two MOLS(3).

Exercise 9.
Complete a Sudoku critical set from the Example 11.

2.3 Pairwise balanced designs and group divisible designs

Relaxing some of conditions in the definition of BIBD leads to other classes of designs. One of them concerns the case when all blocks do not have to have the same size.

Definition 8. Let λ be a positive integer and K be a set of positive integers. A *pairwise balanced design*, PBD(v,K,λ), of order v with block sizes from K is a pair $(V,\mathscr{B})$ where V is a set of cardinality v and $\mathscr{B}$ is a collection of subsets of V called *blocks* such that each block $B\in\mathscr{B}$ has size $|B|\in K$ and every pair of distinct elements of V occurs in exactly λ blocks.

Example 12. A PBD$(6,\{3,4\},3)$:
$V=\{1,2,3,4,5,6\}$,
$\mathscr{B}=\{\{1,2,3,4\},\{1,3,4,5\},\{1,4,5,6\},\{2,3,4,6\},\{2,4,5,6\},\{1,2,5\},\{1,2,6\},\{1,3,6\},$
$\{2,3,5\},\{3,5,6\}\}$.

If a PBD(v,K,λ) has b_i blocks of size k_i for each $k_i\in K$, then $\lambda\binom{v}{2}=\sum_i b_i\binom{k_i}{2}$.

For a set of positive integers K, let $\alpha(K)=\gcd\{k-1: k\in K\}$ and $\beta(K)=\gcd\{k(k-1): k\in K\}$. Then the necessary conditions for the existence of a PBD(v,K,λ) are:
(1) $\lambda(v-1)\equiv 0 \pmod{\alpha(K)}$, and
(2) $\lambda v(v-1)\equiv 0 \pmod{\beta(K)}$.

Remark. *Let $K\neq\{v\}$. If there exists a* PBD$(v,K,1)$*, then $v\geq l(s-1)+1$, where l and s are the largest and the smallest sizes, respectively, of blocks in a* PBD.

Definition 9. Let K and G be sets of positive integers and λ be a positive integer. A *group divisible design* of order v and index λ, GDD(v,K,G,λ), is a triple $(V,\mathscr{B},\mathscr{G})$ where V is a finite set of cardinality v, $\mathscr{G}$ is a partition of V into *groups* whose sizes belong to G, and $\mathscr{B}$ is a collection of subsets of V called *blocks* such that each $B\in\mathscr{B}$ has $|B|\in K$ and every pair of distinct elements of V is contained in exactly λ blocks or in one group, but not both. Moreover, $|\mathscr{G}|\geq 2$.

Given a GDD(v,K,G,λ) with a_i groups of size g_i, $i=1,2,\ldots,s$ (so that $\sum_{i=1}^{s}a_ig_i=v$), we use exponential notation $g_1^{a_1}g_2^{a_2}\ldots g_s^{a_s}$ for the *group type*. If $K=\{k\}$ and $\lambda=1$, then we write $k-$GDD.

Example 13. A GDD$(10,\{3,4\},\{1,3\},1)$ of type $1^1 3^3$:
$V=\{1,2,\ldots,10\}$,
$\mathscr{G}=\{\{1,2,3\},\{4,5,6\},\{7,8,9\},\{10\}\}$,
$\mathscr{B}=\{\{1,4,7,10\},\{2,5,8,10\},\{3,6,9,10\},\{1,5,9\},\{2,6,7\},\{3,4,8\},\{1,6,8\},\{2,4,9\},$
$\{3,5,7\}\}$.

A GDD is *uniform* if $K=\{k\}$ and all its groups have the same size m, that is, if it is of type m^u for some positive integer u. The necessary conditions for the existence of a uniform GDD(v,k,m,λ) of type m^u are:
(1) $u\geq k$,
(2) $\lambda(u-1)m\equiv 0 \pmod{k-1}$,
(3) $\lambda u(u-1)m^2\equiv 0 \pmod{k(k-1)}$.

Definition 10. A *transversal design*, TD(k,m), is a uniform $k-$GDD of type m^k.

In other words, a GDD is a transversal design if and only if each block meets every group in exactly one point.

Theorem 11. *A transversal design* TD(k,m) *exists if and only if there exists a set of* $k-2$ MOLS(m).

A GDD$(v,K,G,1)$ may be viewed as a PBD$(v,K\cup G,1)$ by considering all groups of the GDD to be blocks of the PBD, together with blocks of the GDD.

Lemma 12. *If there exists a group divisible design* $(V,\mathscr{B},\mathscr{G})$ *with* $\lambda=1$*, then there exists a pairwise balanced design* $(V,\mathscr{C})$*, where* $\mathscr{C}=\mathscr{B}\cup\{G\in\mathscr{G}:|G|\geq 2\}$.

Moreover, a GDD$(v,K,G,1)$ can be used to built a PBD$(v+1,K\cup\{g+1:\ g\in G\},1)$ by adjoining a new point to each group to form new blocks. Conversely, a GDD may be obtained from a PBD by deleting a point.

Lemma 13. *Suppose there exists a group divisible design* $(V,\mathscr{B},\mathscr{G})$*,* $\lambda=1$ *and* $\infty\notin V$*. Define* $W=V\cup\{\infty\}$ *and* $\mathscr{C}=\mathscr{B}\cup\{G\cup\{\infty\}:\ G\in\mathscr{G}\}$*. Then* $(W,\mathscr{C})$ *is a pairwise balanced design.*

Certain transversal designs may be obtained using some recursive constructions.

TD$(4,m)\to$ TD$(4,3m)$ **construction.**
Let $(V,\mathscr{B},\mathscr{G})$ be a TD$(4,m)$ and let $W=\{1,2,3\}$. Let $V'=V\times W$ and define a collection $\mathscr{G}'$ of groups and a collection $\mathscr{B}'$ of blocks as follows:
(1) $\mathscr{G}'=\{G\times W:\ G\in\mathscr{G}\}$
(2) for each $B\in\mathscr{B}$, let $(B\times W,\{\{a\}\times W:\ a\in B\},W(B)\}$ be a TD(4,3) and place the 9 blocks belonging to $W(B)$ in $\mathscr{B}'$.
Then $(V',\mathscr{B}',\mathscr{G}')$ is a TD$(4,3m)$.

TD$(4,m)$ **with a parallel class** $\to$ TD$(4,3m+1)$ **construction.**
Let $(V,\mathscr{B},\mathscr{G})$ be a TD$(4,m)$ and let Π be a parallel class of blocks. Let $W=\{1,2,3\}$ and set $V_1=\{\infty_1,\infty_2,\infty_3,\infty_4\}$. Let $V'=V\times W\cup V_1$. Define a collection $\mathscr{G}'$ of groups and a collection $\mathscr{B}'$ of blocks as follows:
(1) $\mathscr{G}'=\{(G_i\times W)\cup\{\infty_i\}:\ G_i\in\mathscr{G}\}$
(2) for each block $B\in\Pi$, let $((B\times W)\cup V_1,\{(\{a\}\times W)\cup\{\infty_i\}:\ a\in B\cap G_i,\ i\in W\},W(B)\}$ be a TD$(4,4)$ with a requirement that $\{\infty_1,\infty_2,\infty_3,\infty_4\}$ is a block; place 15 blocks of $W(B)\setminus\{\infty_1,\infty_2,\infty_3,\infty_4\}$ in $\mathscr{B}'$

(3) for each $B \in \mathscr{B} \setminus \Pi$, let $(B \times W, \{\{a\} \times W : a \in B\}, W(B)\}$ be a TD$(4,3)$ and place the 9 blocks belonging to $W(B)$ in $\mathscr{B}'$
(4) place $\{\infty_1, \infty_2, \infty_3, \infty_4\}$ in $\mathscr{B}'$.
Then $(V', \mathscr{B}', \mathscr{G}')$ is a TD$(4, 3m+1)$.

Exercise 10.
(1) Construct a PBD$(10, \{3,4\}, 1)$.
(2) Construct a PBD$(12, \{3,4\}, 1)$.
(3) Construct a PBD$(11, \{3,5\}, 1)$.

Exercise 11.
Show that a PBD$(8, \{3,4\}, 1)$ does not exist.

Exercise 12.
(1) Construct a $3-$GDD of type 3^5.
(2) Construct a $4-$GDD of type 3^4.

Exercise 13.
Construct a TD$(4,13)$.

2.4 Steiner triple systems

The first class of intensively studied designs were BIBD's with block size 3 and $\lambda = 1$.

Definition 11. A *Steiner triple system*, STS(v), of order v is a $(v,3,1)-$BIBD. Blocks of an STS(v) are often called *triples*.

The arithmetic necessary conditions for the existence of an STS(v) reduce to $v \equiv 1,3 \pmod 6$. This is also a sufficient condition, what was proven in 1847 by Kirkman. One of the simplest known direct constructions is due to Bose and Skolem.

Bose construction (for STS(v) when $v \equiv 3 \pmod 6$).
Let $v = 6k+3$ and let $(Q, \circ)$ be an idempotent commutative quasigroup of order $2k+1$, where $Q = \{0, 1, \ldots, 2k\}$. Let $V = Q \times \{1,2,3\}$, and define $\mathscr{B}$ to contain the following two types of triples:
(1) for $0 \le i \le 2k$, $\{(i,1),(i,2),(i,3)\} \in \mathscr{B}$,
(2) for $0 \le i < j \le 2k$, $\{(i,1),(j,1),(i \circ j,2)\} \in \mathscr{B}$, $\{(i,2),(j,2),(i \circ j,3)\} \in \mathscr{B}$,
$\{(i,3),(j,3),(i \circ j,1)\} \in \mathscr{B}$.

Skolem construction (for STS(v) when $v \equiv 1 \pmod 6$).
Let $v = 6k+1$ and let $(Q, \circ)$ be a half-idempotent commutative quasigroup of order $2k$, where $Q = \{0, 1, \ldots, 2k-1\}$. Let $V = (Q \times \{1,2,3\}) \cup \{\infty\}$, and define $\mathscr{B}$ as follows:
(1) for $0 \le i \le k-1$, $\{(i,1),(i,2),(i,3)\} \in \mathscr{B}$,
(2) for $0 \le i \le k-1$, $\{\infty,(k+i,1),(i,2)\} \in \mathscr{B}$, $\{\infty,(k+i,2),(i,3)\} \in \mathscr{B}$,
$\{\infty,(k+i,3),(i,1)\} \in \mathscr{B}$,
(3) for $0 \le i < j \le 2k-1$, $\{(i,1),(j,1),(i \circ j,2)\} \in \mathscr{B}$, $\{(i,2),(j,2),(i \circ j,3)\} \in \mathscr{B}$,
$\{(i,3),(j,3),(i \circ j,1)\} \in \mathscr{B}$.

An STS(v) is *cyclic* if it admits an automorphism which is a single cycle of length v. Then all triples may be represented by base triples, one for each orbit of triples under a cyclic automorphism. The existence of cyclic Steiner triple systems may be proven by

solving two problems posed by Heffter in 1896. An ordered 3-element subset $\{a,b,c\}$ of the set $\{1,2,\ldots,(v-1)/2\}$ is called a *difference triple* if either $a+b=c$ or $a+b+c=v$.

Heffter's difference problems.
(1) Let $v=6k+1$. Is it possible to partition the set $\{1,2,\ldots,3k\}$ into k difference triples?
(2) Let $v=6k+3$. Is it possible to partition the set $\{1,2,\ldots,3k+1\}\setminus\{2k+1\}$ into k difference triples?

In 1939, Peltesohn solved both Heffter's difference problems in the affirmative except for $v=9$ (for which no solution exists).

Example 14. A solution to the second Heffer's difference problem for $v=27$ is:
$\{\{1,2,3\},\{4,10,13\},\{5,6,11\},\{7,8,12\}\}$.
The base blocks corresponding to the difference triples are:
$\{0,1,3\},\{0,4,14\},\{0,5,11\},\{0,7,15\}$.

Given a solution to the first Heffter's difference problem, i.e. the collection of k ordered triples, each triple $\{a,b,c\}$ forms the base triple $\{0,a_i,a_i+b_i\}$ of a cyclic STS($6k+1$). Similarly, given a solution to the second Heffter's difference problem, each triple $\{a,b,c\}$ forms the base triple $\{0,a_i,a_i+b_i\}$ of a cyclic STS($6k+3$); one more base triple (for *short orbit*) is $\{0,2k+1,4k+2\}$.

Solutions to both Heffter's difference problems may be reduced to finding certain integer sequences.

A *Skolem sequence* of order n is a sequence $S=(s_1,s_2,\ldots,s_{2n})$ of $2n$ integers satisfying:
(1) for every $k\in\{1,2,\ldots,n\}$ there exist exactly two elements $s_i,\ s_j\in S$ such that $s_i=s_j=k$,
(2) if $s_i=s_j=k$ with $i<j$, then $j-i=k$.

Example 15. A Skolem sequence of order 5:
$S=(2,4,2,3,5,4,3,1,1,5)$.

A Skolem sequence of order n exists if and only if $n\equiv 0,1 \pmod 4$. Given a Skolem sequence S of order n, the collection of triples $\{\{k,n+i,n+j\}:\ s_i=s_j=k, k=1,2,\ldots,n\}$ is a solution to the first Heffter's problem.

When $n\equiv 2,3 \pmod 4$ we use an extension of a Skolem sequence. A *hooked Skolem sequence* of order n is a sequence $HS=(s_1,s_2,\ldots,s_{2n+1})$ of $2n+1$ integers satisfying:
(1) for every $k\in\{1,2,\ldots,n\}$ there exist exactly two elements $s_i,\ s_j\in S$ such that $s_i=s_j=k$,
(2) if $s_i=s_j=k$ with $i<j$, then $j-i=k$,
(3) $s_{2n}=0$.

Example 16. A hooked Skolem sequence of order 6:
$S=(6,3,5,2,3,2,6,5,4,1,1,0,4)$.

A hooked Skolem sequence of order n exists if and only if $n\equiv 2,3 \pmod 4$. Given a hooked Skolem sequence S of order n, the collection of triples $\{\{k,n+i,n+j\}:\ s_i=s_j=k, k=1,2,\ldots,n\}$ is a solution to the first Heffter's problem.

Extensions of Skolem and hooked Skolem sequences, called *split* and *split hooked Skolem sequences*, with zero on the position $n+1$ and two zeros on the positions $n+1$, $2n+1$, respectively, can be used is similar way in order to obtain solutions to the second

Heffter's difference problem.

The number of pairwise nonisomorphic Steiner triple systems increases rapidly with v. While STS(7) and STS(9) are unique (up to isomorphism), there are two STS(13)'s, 80 STS(15)'s and 11,084,874,829 STS(19)'s.

The existence of Steiner triple systems for each admissible order $v \equiv 1,3 \pmod 6$ may be also proven by applying two recursive constructions.

$v \to 2v+1$ **construction.**
Let $(V, \mathscr{B})$ be an STS(v) and let $(X, \mathscr{F})$ be a one-factorization of the complete graph of order $v+1$ on the set of vertices X. Let $\mathscr{C} = \{\{v_i, x, y\} : v_i \in V, \{x,y\} \in F_i \in \mathscr{F}\}$. Then $(V \cup X, \mathscr{B} \cup \mathscr{C})$ is an STS($2v+1$) with a subsystem STS(v).

The second construction uses the existence of one-factorizations in some circulant graphs, determined by Stern-Lenz Lemma.

Lemma 14. *A circulant graph* $C(n; d_1, d_2, \ldots, d_s)$ *has a 1-factorization if and only if* $n/\gcd(d_i, n)$ *is even for at least one generator* d_i.

$v \to 2v+7$ **construction.**
Let $(V, \mathscr{B})$ be an STS(v). Let $(X, \mathscr{F})$ be a collection of edge-disjoint one-factors in the complete graph K_{v+7} on the set X, and moreover let T be a set of $v+7$ triples, which together with one-factors in $\mathscr{F}$ form a partition of the edge set of K_{v+7}. Let $\mathscr{C} = \{\{v_i, x, y\} : v_i \in V, \{x,y\} \in F_i \in \mathscr{F}\}$. Then $(V \cup X, \mathscr{B} \cup \mathscr{C} \cup T)$ is an STS($2v+7$) with a subsystem STS(v).

Another well studied class of Steiner triple systems are projective triple systems. Let W_m be an $(m+1)$-dimensional vector space over $\mathbb{F}_2$. Let $\oplus$ be the operation of vector addition in W_m. Any two nonzero $(m+1)$-vectors $\mathbf{x}$ and $\mathbf{y}$ determine uniquely a third vector $\mathbf{x} \oplus \mathbf{y}$ in W_m, where addition is performed modulo 2 componentwise. Let every nonzero vector in W_{m+1} be represented by a point in a set V of cardinality $2^{m+1}-1$. Every two distinct points, corresponding to $\mathbf{x}$ and $\mathbf{y}$, define a unique triple formed by $\{\mathbf{x}, \mathbf{y}, \mathbf{x} \oplus \mathbf{y}\}$. The STS($2^{m+1}-1$) produced in this way is called a *projective triple system* and it is often denoted by PG(m,2) (just consider the triples as lines in the projective space over GF(2)). To simplify notation, let every point in V be labeled by an integer whose binary representation is determined by the coordinates of its corresponding vector. Thus $V(\mathrm{PG}(m,2)) = \{1, 2, \ldots, 2^{m+1}-1\}$.

A *partial triple system* PTS(v) is a pair $(V, \mathscr{B})$, where $|V| = v$ and $\mathscr{B}$ is a collection of 3-element subsets of V such that each unordered pair of elements of V occurs in at most one triple of $\mathscr{B}$. Let $(V, \mathscr{B})$ be a PTS(v) and $(W, \mathscr{D})$ be an STS(w) for which $V \subseteq W$ and $\mathscr{B} \subseteq \mathscr{D}$. Then $(W, \mathscr{D})$ is an *embedding* of $(V, \mathscr{B})$.

Theorem 15. *Any partial triple system* PTS(v) *can be embedded in an* STS(w) *if* $w = 1,3 \pmod 6$ *and* $w \geq 2v+1$.

Theorem 16. *Let* $v, w \equiv 1,3 \pmod 6$ *and* $v \geq 2w+1$. *Then there exists an* STS(v) *containing an* STS(w) *as a subsystem.*

Exercise 14.
Apply Skolem construction to get an STS(13).

Exercise 15.
Show that a cyclic STS(9) does not exist.

Exercise 16.
Find a solution to the Heffer's difference problems when:
(1) v=19
(2) v=21.

Exercise 17.
Construct an embedding of an STS(7) into an STS(27).

2.5 Resolvable designs

A *parallel class* in a design $(V, \mathscr{B})$ is a set of blocks that partition the set V. A *partial parallel class* is a set of blocks that contain no point of the design more than once.

Definition 12. A design $(V, \mathscr{B})$ is *resolvable* if all its blocks can be partitioned into parallel classes.

Example 17. A $(9,3,1)-$BIBD is resolvable; parallel classes are R_1, R_2, R_3, R_4:
$V = \{0,1,\ldots,9\}$,
$R_1 = \{\{0,1,2\}, \{3,4,5\}, \{6,7,8\}\}$,
$R_2 = \{\{0,3,6\}, \{1,4,7\}, \{2,5,8\}\}$,
$R_3 = \{\{0,4,8\}, \{1,5,6\}, \{2,3,7\}\}$,
$R_4 = \{\{0,5,7\}, \{1,3,8\}, \{2,4,6\}\}$.

Definition 13. A *Kirkman triple system*, KTS(v), of order v is a resolvable STS(v) together with a resolution of its blocks.

Distinct resolutions of a given STS(v) may form nonisomorphic KTS's.

Example 18. KTS(15), $V = \{1,2,\ldots,15\}$,
$R_1 = \{\{1,2,3\}, \{4,8,12\}, \{5,11,14\}, \{6,9,15\}, \{7,10,13\}\}$,
$R_2 = \{\{1,4,5\}, \{2,12,14\}, \{3,9,10\}, \{6,11,13\}, \{7,8,15\}\}$,
$R_3 = \{\{1,6,7\}, \{2,13,15\}, \{3,8,11\}, \{4,10,14\}, \{5,9,12\}\}$,
$R_4 = \{\{1,8,9\}, \{2,4,6\}, \{3,13,14\}, \{5,10,15\}, \{7,11,12\}\}$,
$R_5 = \{\{1,10,11\}, \{2,5,7\}, \{3,12,15\}, \{4,9,13\}, \{6,8,14\}\}$,
$R_6 = \{\{1,12,13\}, \{2,8,10\}, \{3,5,6\}, \{4,11,15\}, \{7,9,14\}\}$,
$R_7 = \{\{1,14,15\}, \{2,9,11\}, \{3,4,7\}, \{5,8,13\}, \{6,10,12\}\}$.

The existence problem for Kirkman triple systems was completely solved by Ray-Chaudhuri and Wilson in 1971, more than 120 years after the problem was posed by Kirkman.

Theorem 17. *A Kirkman triple system of order v exists if and only if $v \equiv 3 \pmod 6$.*

A proof of sufficiency bases on two important facts:

Lemma 18. *For each $v \equiv 1 \pmod 3$, there exists a $(v, \{4,7,10,19\}, 1)-$PBD.*

Lemma 19. *If there exists a $(v, K, 1)-$PBD, $v \equiv 1 \pmod 3$, and for each $k_i \in K$ there exists a* KTS($2k_i+1$)*, then there exists a* KTS($6n+3$).

Construction of a Kirkman triple system.
Let $v = 6n+3$ and let $W = V \times \{1,2\} \cup \{\infty\}$ where $|V| = 3n+1$.
Let $(V, \mathscr{B})$ be a $(3n+1, \{4,7,10,19\}, 1)-$PBD.

For each block $B \in \mathscr{B}$, put on the set $B \times \{1,2\} \cup \{\infty\}$ a copy of a KTS($2|B|+1$) with a resolution $\mathscr{R}_B$ in such a way that $\{x_1, x_2, \infty\}$ is a triple for each $x \in B$.
Let R_{Bx} be a parallel class of $\mathscr{R}_B$ containing the triple $\{x_1, x_2, \infty\}$. Then $R_x = \bigcup_{B \in \mathscr{B}} R_{Bx}$ is a parallel class on W and $\mathscr{R} = \{R_x : x \in V\}$ is a resolution of a KTS(v).

The necessary conditions are also sufficient for the existence of a resolvable (v,k,1)-BIBD when k is small, namely:
if $k = 2$, $v \equiv 0 \pmod 2$,
if $k = 3$, $v \equiv 3 \pmod 6$,
if $k = 4$, $v \equiv 4 \pmod{12}$.
For $k = 5$, a resolvable $(v,5,1)-$BIBD exists if $v \equiv 5 \pmod{20}$ and $v \neq 45, 345, 465, 645$, in which cases the existence problem remains open.

Theorem 20. *A resolvable transversal design* TD(k,m) *exists if and only if there exists a set of* $k-1$ MOLS(m).

Corollary 21. *A resolvable transversal design* TD(k,m) *exists if and only if there exists transversal design* TD($k+1,m$).

When $v \equiv 1 \pmod 6$, the maximum number of pairwise disjoint triples is $\frac{v-1}{3}$. Then the maximum partial parallel class has to miss one point.

Definition 14. A *Hanani triple system*, HTS(v), of order v is an STS(v) with a partition of its blocks into $(v-1)/2$ almost parallel classes and a single partial parallel class with $(v-1)/6$ triples.

Theorem 22. *A Hanani triple system of order* v *exists if and only if* $v \equiv 1 \pmod 6$ *and* $v \notin \{7, 13\}$.

Exercise 18.
Construct a resolvable $(16,4,1)-$BIBD.

Exercise 19.
Construct a resolvable TD(5,7).

Exercise 20.
Show that an HTS(7) does not exist.

2.6 Other classes of designs

2.6.1 Affine and projective planes

A *finite incidence structure* (or *finite geometry*), $P = (\mathscr{P}, \mathscr{L}, I)$ is made of a finite set of points $\mathscr{P}$, a finite set of lines $\mathscr{L}$, and an *incidence* relation I between them.

Definition 15. A *finite affine plane* is a finite incidence structure such that the following axioms are satisfied:
(A1) any two distinct points are incident with exactly one line,
(A2) for any point P outside a line l there is exactly one line through P that has no point in common with l,

(A3) there exist three points not on a common line.

For a finite affine plane A, there is a positive integer n such that any line of A has exactly n points. This number is the *order* of A. A finite affine plane of order n has n^2 points, n^2+n lines, and $n+1$ lines through each point.

Lemma 23. *An affine plane or order n is a* BIBD$(n^2, n^2+n, n, n+1, 1)$. *Conversely,* BIBD$(n^2, n^2+n, n, n+1, 1)$ *is an affine plane of order n.*

Remark. *An affine plane is resolvable.*

Theorem 24. An affine plane of order n exists if n is a prime power.

Construction of an affine plane of a prime power order.
Let $q = p^k$ be a prime power. Let $V = \mathbb{F}_q \times \mathbb{F}_q$.
For any $a, b \in F_q$, define a block $L_{a,b} = \{(x,y) \in V : y = ax+b\}$.
For any $c \in F_q$, define $L_{\infty,c} = \{(c,y) \in V : y \in \mathbb{F}_q\}$.
Finally, define $\mathscr{L} = \{L_{a,b} : a, b \in \mathbb{F}_q\} \cup \{L_{\infty,c} : c \in \mathbb{F}_q\}$.
$(V, \mathscr{L})$ is a $(q^2, q, 1) -$ BIBD.

Remark. The existence of an affine plane of order n is equivalent to the existence a set of $n-1$ MOLS(n).

Definition 16. A *finite projective plane* is a finite incidence structure such that the following axioms are satisfied:
(P1) any two distinct points are incident with exactly one line,
(P2) any two distinct lines are incident with exactly one point,
(P3) there exist four points no three of which are on the same line.

For a finite projective plane P, there is a positive integer n such that any line of P has exactly $n+1$ points. This number is the *order* of P. A finite projective plane of order n has n^2+n+1 points, n^2+n+1 lines, and $n+1$ lines through each point.

Lemma 25. *A projective plane or order n is a* BIBD$(n^2+n+1, n^2+n+1, n+1, n+1, 1)$. BIBD$(n^2+n+1, n^2+n+1, n+1, n+1, 1)$ *is a projective plane of order n.*

Remark. *A projective plane of order n exists if and only if an affine plane of order n exists.*

Exercise 21.
Construct an affine plane of order 4.

2.6.2 Cycle systems

Definition 17. A *k-cycle system* of order n is a pair $(X, \mathscr{C})$ where $\mathscr{C}$ is a collection of edge-disjoint k-cycles which partition the edge set of K_n with $V(K_n) = X$.

Example 19. A 4-cycle system $(X, \mathscr{C})$ of order 9:
$V = \{0, 1, \ldots, 8\}$, $\mathscr{C} = \{(0,1,5,2), (0,3,8,7), (0,4,1,8), (0,5,4,6), (1,2,6,3), (1,6,5,7),$

$(2,3,7,4), (2,7,6,8), (3,4,8,5)\}$.

Theorem 26. *A k-cycle system of order n exists if and only if:*
(1) $n \geq k \geq 3$,
(2) *n is odd,*
(3) $2k|n(n-1)$.

A k-cycle system $(X, \mathscr{C})$ of order n is *resolvable* if the k-cycles belonging to C can be partitioned into parallel classes.

Example 20. A resolvable 5-cycle system $(X, \mathscr{C})$ of order 15:
$V = \{0, 1, \ldots, 14\}$,
$R_1 = \{(0,1,3,6,9), (2,7,8,10,13), (4,11,5,14,12)\}$,
$R_1 = \{(0,2,5,8,6), (1,12,9,7,13), (3,10,4,14,11)\}$,
$R_1 = \{(0,3,13,4,5), (1,8,2,14,9), (6,10,7,12,11)\}$,
$R_2 = \{(0,4,2,1,10), (3,7,11,9,8), (5,12,6,14,13)\}$,
$R_1 = \{(0,7,1,14,8), (2,6,4,3,12), (5,10,11,13,9)\}$,
$R_1 = \{(0,11,8,13,12), (1,4,7,5,6), (2,9,3,14,10)\}$,
$R_3 = \{(0,13,6,7,14), (1,5,3,2,11), (4,8,12,10,9)\}$.

Theorem 27. *A resolvable k-cycle system of order n exists if and only if:*
(1) $n \geq k \geq 3$,
(2) *n is odd,*
(3) $k|n$.

Theorem 28. *Let n be odd, $3 \leq m_1, m_2, \ldots, m_t \leq n$ and $m_1 + m_2 + \ldots + m_t = n(n-1)/2$. Then there exists a decomposition of K_n into t cycles of lengths $m_1, m_2, \ldots, m_t$.*

Oberwolfach Problem. *Let n be odd, $3 \leq m_1, m_2, \ldots, m_t \leq n$ and $m_1 + m_2 + \ldots + m_t = n$. Does the complete graph K_n have a 2-factorization in which every 2-factor consists of cycles of lengths $m_1, m_2, \ldots, m_t$?*

The Oberwolfach problem has an affirmative solution for $n \leq 40$ and every admissible collection of cycles lengths, with the exception of two cases:
(1) $m_1 = 4$, $m_2 = 5$
(2) $m_1 = m_2 = 3$, $m_3 = 5$.

Exercise 22.
Construct a 5-cycle system of order 11.

Exercise 23.
Construct a resolvable 5-cycle system of order 25.

2.6.3 G-designs

Definition 18. A *G-design* of order v and index λ (or *$(\lambda K_n, G)$-design*) is a G-decomposition of a complete λ-multigraph λK_n. A $(\lambda K_n, G)$-design is *balanced* if each vertex of λK_n occurs in the same number of copies of G.

Theorem 29. *There exists a $(\lambda K_n, K_{1,k})$-design if and only if $n \geq k+1$ and $\lambda n(n-1) \equiv 0 \pmod{2k}$.*

Theorem 30. *There exists a $(\lambda K_n, P_k)$-design if and only if $n \geq k$ and $\lambda n(n-1) \equiv 0 \pmod{(2k-2)}$.*

Example 21. A (K_6, P_4)-design:
$V = \{0,1,2,3,4,5\}$,
$\mathscr{P} = \{(0,1,2,4), (0,2,3,5), (0,3,4,1), (0,4,5,2), (0,5,1,3)\}$.

Conjecture. *There exists a (K_{2n+1}, T)-design for each tree T with n edges.*

Exercise 24.
Construct a $(2K_7, K_{1,6})$-design.

2.6.4 t-designs

Definition 19. A *$t-(v,k,\lambda)$-design* is a pair $(V, \mathscr{B})$ where $|V| = v$ and $\mathscr{B}$ is a collection of k-element subsets of V (*blocks*) with the property that each t-element subset of V is contained in exactly λ blocks.

An ordered quadruple of positive integers (λ, t, k, v) is called *admissible* if $\lambda_s = \lambda(\binom{v-s}{t-s})/(\binom{k-s}{t-s})$ is an integer for each $0 \leq s < t$.

A *Steiner quadruple system* of order v (SQS(v)) is a $3-(v,4,1)$-design.

Example 22. A cyclic SQS(10):
$V = \mathbb{Z}_{10}$. The base blocks are: $B_1 = \{0,1,3,4\}$, $B_2 = \{0,1,2,6\}$, $B_3 = \{0,2,4,7\}$.

Theorem 31. *An* SQS(v) *exists if and only if $v \equiv 2,4 \pmod 6$.*

Exercise 25.
Construct an SQS(8).

2.6.5 Room squares

Definition 20. Let S be a set of $n+1$ elements (*symbols*). A *Room square* of side n is an $n \times n$ array, R, that satisfies the following properties:
(1) every cell of R is either empty or contains an unordered pair of symbols from S,
(2) every symbol of S occurs exactly once in each row and exactly once in each column of R,
(3) every unordered pair of symbols occurs in precisely one cell in R.

Thus each row and each column of R contain $\frac{n-1}{2}$ empty cells.

Example 23. A room square of side 9:
$S = \{0, 1, \ldots, 9\}$,

01		49	37	28		56		
89	02				57	34		16
	58	03		69	24		17	
	36	78	04		19		25	
	79		12	05	38		46	
45					06	18	39	27
		26	59	13		07		48
67	14					29	08	35
23		15	68	47				09

Theorem 32. *A room square of side n exists if and only if n is odd and $n \notin \{3,5\}$.*

For odd n, two 1-factorizations of the complete graph K_{n+1}, $\mathscr{F} = \{F_1, F_2, \ldots, F_n\}$ and $\mathscr{G} = \{G_1, G_2, \ldots, G_n\}$ are *orthogonal* if $|F_i \cap G_i| \leq 1$ for all $1 \leq i, j \leq n$. The existence of a Room square of side n is equivalent to the existence of two orthogonal 1-factorizations of K_{n+1}.

Exercise 26.
Show that a Room square of side 5 does not exist.

Exercise 27.
Construct a Room square of side 7.

2.6.6 Hadamard matrices and designs

In 1893, Hadamard addressed the problem of the maximum absolute value of the determinant of an $n \times n$ complex matrix H with all its entries on a unit circle. That maximum value is $\sqrt{n^n}$. Among real matrices, this value is attained if and only if H has every entry either 1 or -1, and satisfies $HH^T = nI$. This condition means that any two distinct rows of $H(n)$ are orthogonal.

Definition 21. An $n \times n$ (± 1)-matrix $H(n)$ is a *Hadamard matrix* of side n if $HH^T = nI$.

Notice that we may multiply all entries in any row (and column) by -1 and the result is again a Hadamard matrix. By a sequence of such multiplications, a Hadamard matrix may be transformed into another Hadamard matrix, in which every entry in the first row or in the first column is 1. Such a Hadamard matrix is called *standardized*.

Example 24. $H(4)$:

$$\begin{bmatrix} + & + & + & + \\ + & + & - & - \\ + & - & + & - \\ + & - & - & + \end{bmatrix}$$

Necessary condition for the existence of an $H(n)$ is $n \equiv 0 \pmod 4$ or $n = 1,2$. The famous conjecture, stated by Hadamard in 1893, claims that the above condition is also sufficient. The smallest order for which the conjecture remains open is 668.

Definition 22. A *Hadamard design* is a symmetric $(4m-1, 2m-1, m-1)-$BIBD.

The existence of a Hadamard design of order $4m-1$ is equivalent to the existence of a Hadamard matrix of side $4m$.

Example 25. $(7,3,1)-$BIBD and its corresponding $H(8)$.

1	1	0	1	0	0	0
0	1	1	0	1	0	0
0	0	1	1	0	1	0
0	0	0	1	1	0	1
1	0	0	0	1	1	0
0	1	0	0	0	1	1
1	0	1	0	0	0	1

$$\begin{bmatrix} + & + & + & + & + & + & + & + \\ + & + & + & - & + & - & - & - \\ + & - & + & + & - & + & - & - \\ + & - & - & + & + & - & + & - \\ + & - & - & - & + & + & - & + \\ + & + & - & - & - & + & + & - \\ + & - & + & - & - & - & + & + \\ + & + & - & + & - & - & - & + \end{bmatrix}$$

Exercise 28.
Construct a Hadamard matrix $H(12)$.

2.7 References

[1] C.J. Colbourn, J.H. Dinitz (eds.), *Handbook of Combinatorial Designs, Second Edition*, Chapman & Hall/CRC, 2006.

[2] C.J. Colbourn, A. Rosa, *Triple Systems*, Clarendon Press, 1999.

[3] C.C. Lindner, C.A. Rodger, *Design Theory, Second Edition*, Chapman & Hall/CRC, 2009.

[4] D.R. Stinson, *Combinatorial Designs, Constructions and Analysis*, Springer, 2004.

[5] W.D. Wallis, *Introduction to Combinatorial Designs*, Chapman & Hall/CRC, 2007.

Chapter 3

Some Topics in the Theory of Finite Groups

Primož Moravec
University of Ljubljana, Slovenia

SUMMARY

The theory of finite groups plays a central role in group theory and has several applications in other branches of mathematics, including discrete mathematics and cryptography. The theory culminated with the classification of finite simple groups in 1983, and has developed afterwards into several different directions such as the theory of groups of prime power order, invariant theory, and many others. This mini course will address some topics of the above theory. These will include advanced applications of Sylow's theory, techniques of building new groups from old, basic theory of finite p-groups, and problems regarding enumeration of finite groups.

3.1 Introduction

These notes form a background material for a short course on group theory that was given at *2014 PhD Summer School in Discrete Mathematics and SYGN, Rogla, Slovenia*. Since the summer school was aimed primarily at PhD students who are working in the latter area and may not necessarily be experts in group theory, the notes give a fairly general introduction to three main topics: Finite Simple Groups, Extension Theory of Groups, and Nilpotent groups and Finite p-groups. The choice of the first two topics is clear from the point of view of classifying all finite groups. It turns out that the knowledge of all finite simple groups, together with knowing how to "glue" two groups together to produce new ones, in principle provides a way of constructing *all* finite groups. The first problem, classification of finite simple groups (CFSG), has been resolved satisfactory, and one can operate with a full list of these groups. In these notes we will only touch this vast area by showing simplicity of alternating groups and projective special linear groups. We will sketch the classification, but ommit almost all further details. We will move on to extension theory which tells us how to construct new groups from old. The extension problem of classifying all possible extensions of one group by another appears to be hard (impossible?) to solve in general. We will only study a very special case of it.

There are two main reasons why to deal with finite p-groups, i.e., groups whose orders are powers of a prime p. The first is clear to an undergraduate student: finite p-groups appear as Sylow p-subgroups of finite groups. The second is more delicate and motivated by a vague statement "*Almost all finite groups are p-groups.*" We will not make any attempt of making this statement more precise, but rather develop some basic theory of these groups and indicate their complexity within the universe of all finite groups.

In addition to the above, we include preliminaries that will be needed in subsequent sections. We collect some basic properties of groups with focus on finite groups. We also exhibit as many examples as possible in order to illustrate and motivate the theory. A general experience is that most of the students only know some standard types of groups, such as abelian groups, dihedral groups, symmetric and alternating groups,... Other groups which do not have clean descriptions are usually put aside. In order to avoid this, I use GAP (Groups, Algorithms, and Programming), a computational system designed for constructing and manipulating with groups. GAP is applied in exploring properties of groups, and even providing proofs of statements. Examples with full GAP code are be given, but I have decided to leave out all explanations of the syntax and programming rules. There are two reasons for this. One is that the reader will mostly find it easy to figure out what a given line of GAP code does, since the syntax is very much self-explanatory. The second one is that there is an extensive manual of GAP, together with tons of tutorials and self-study material available at GAP's web page [5]. We encourage the reader to download GAP (it's open source) and try out all of the examples in these

notes.

I have closely followed Robinson's book *A course in the theory of groups* [8] and Cameron's lecture notes on finite groups [4], thus I claim very little originality as far as for the exposition goes.

3.2 Basic notions and examples

In this section we collect some basic properties of groups and important examples the reader should be familiar with in order to read these notes. Most of the proofs in this section will be omitted. We will also show how to use GAP in performing explicit calculations with groups. Concrete examples of computations will be presented.

A convention about the notations. All (or most) of the functions we consider will be acting from the right. This means that if $f: X \to Y$ is a function and $x \in X$, then the image of x under f will (usually) be denoted by x^f or xf.

The main sources of the material covered here are [6] and [8].

3.2.1 Groups

A non-empty set G equipped with a binary operation $\circ$ is a *group* if the following hold:

- Associativity: $(a \circ b) \circ c = a \circ (b \circ c)$ for all $a, b, c \in G$;
- Identity element: there exists $e \in G$ such that $e \circ a = a \circ e = a$ for all $a \in G$;
- Inverse: For every $a \in G$ there exists $a' \in G$ such that $a \circ a' = a' \circ a = e$.

It is easy to show that the identity element e is uniquely determined, and that every $a \in G$ has a unique inverse, denoted by a^{-1}. For most of the time we write $\cdot$ instead of $\circ$; in this case, when there is no confusion, we write 1 instead of e (multiplicative notation). If $g, h \in G$, we will often use the notation $g^h = h^{-1}gh$ for *conjugation* of g by h. If the set G is finite, then we say that G is a *finite group*, and $|G|$ is called the *order* of G.

A group G is *abelian* if $a \circ b = b \circ a$ for all $a, b \in G$. In this case we often write $+$ instead of $\circ$, and the identity element is denoted by 0 (additive notation).

A subset H of G is called a *subgroup* of G if it is a group under the same operation. We write $H \leq G$. One can verify directly that H is a subgroup of G if and only if $ab^{-1} \in H$ for all $a, b \in H$.

If H is a subgroup of G and $a \in G$, then we define *left (right) cosets* of H by

$$aH = \{ah \mid h \in H\},$$
$$Ha = \{ha \mid h \in H\}.$$

The set of all left cosets of H in G is denoted by G/H, and the set of all right cosets by $H\backslash G$. Different left (right) cosets form a partition of G. The number of left (= the number of

right) cosets of H in G is the *index* of H in G and is denoted by $|G : H|$. If G is a finite group then *Lagrange's theorem* says that $|G| = |H| \cdot |G : H|$. In particular, if $H \leq G$, then $|H|$ divides the order of G.

The intersection of a family of subgroups of a given group G is again a subgroup of G. Thus, if X is a non-empty subset of G, then there exists the smallest subgroup of G containing X. It is denoted by $\langle X \rangle$ and called the *subgroup generated by* X. We say that a group G is *finitely generated* if there exists a finite set X of its elements such that $G = \langle X \rangle$.

Let G_1 and G_2 be groups. A map $\phi : G_1 \to G_2$ is said to be a homomorphism of groups if it preserves group operation, that is,

$$(ab)^\phi = a^\phi b^\phi \text{ for all } a, b \in G_1,$$

where the products are calculated in the corresponding groups. The set

$$\ker \phi = \{x \in G_1 \mid x^\phi = 1\}$$

is said to be *the kernel* of ϕ and is a subgroup of G_1. The set

$$\operatorname{im} \phi = \{x^\phi \mid x \in G_1\}$$

is a subgroup of G_2 and is called the *image* of ϕ. A group homomorphism $\phi : G_1 \to G_2$ is said to be an *epimorphism* if $\operatorname{im} \phi = G_2$; *monomorphism* if $\ker \phi = \{1\}$; isomorphism if it is epimorhism and monomorphism; *endomorphism* if $G_1 = G_2$. A bijective endomorphism is also called an *automorphism*.

A subgroup H of G is said to be a *normal subgroup* of G if $xH = Hx$ for every $x \in G$. Equivalently, $x^{-1}Hx \subseteq H$ for all $x \in G$, i.e., H is *closed under conjugation* by the elements of G. If H is a normal subgroup of G then the sets of left and right cosets of H in G coincide, and we use the commonly accepted notation G/H for these. The operation on G/H given by $Ha \cdot Hb = H(ab)$ is well defined and turns G/H into a group called the *factor group* of G over H. The map $\rho : G \to G/H$ given by $g^\rho = Hg$ is a surjective homomorphism of groups with $\ker \rho = H$.

The intersection of a family of normal subgroups of G is again a normal subgroup of G. Thus, given a set $X \subseteq G$, there exists the smallest normal subgroup of G containing X. It is denoted by $\langle\langle X \rangle\rangle$ and called the *normal closure* of X in G.

Theorem 3.2.1 (First Isomorphism Theorem) *Let $\phi : G_1 \to G_2$ be a homomorphism of groups. Then $G_1/\ker \phi \cong \operatorname{im} \phi$.*

Theorem 3.2.2 (Second Isomorphism Theorem) *Let H be a subgroup and N a normal subgroup of G. Then $H \cap N \triangleleft H$, and $HN/N \cong H/(H \cap N)$.*

Theorem 3.2.3 (Third Isomorphism Theorem) *Let M and N be normal subgroups of G and let $N \leq M$. Then $M/N \triangleleft G/N$ and $(G/N)/(M/N) \cong G/M$.*

One can generalize the notion of normal subgroups as follows. A subgroup H of G is said to be *subnormal* in G if there exists a finite series $H = H_0 \triangleleft H_1 \triangleleft H_2 \triangleleft \cdots \triangleleft H_d = G$. The shortest length of such a series is called the *defect* of H in G. Subnormal subgroups of defect one are precisely normal subgroups.

Two other notions related to normal subgroups are the following. A subgroup H of G is said to be *fully invariant* if $H^\alpha \leq H$ for every endomorphism α of G. Similarly, H is *characteristic* in G if $H^\alpha \leq H$ for every automorphism α of G. The following is straightforward:

Lemma 3.2.4 *The properties of being a 'characteristic subgroup' and 'fully invariant subgroup' are transitive relations. If H is characteristic in K and K normal in G then $H \triangleleft G$.*

Let G be a group and $x, y \in G$. The *commutator* of x and y is defined by $[x,y] = x^{-1}y^{-1}xy = x^{-1}x^y$. The subgroup G' generated by all the commutators $[x,y]$, where $x, y \in G$, is called the *derived subgroup* or the *commutator subgroup* of G. Since $[x,y]^\alpha = [x^\alpha, y^\alpha]$ for all endomorphisms α of G, it follows that G' is a fully invariant subgroup of G. It is easy to verify that G/G' is abelian. Furthermore, if N is normal subgroup of G with G/N abelian, then $G' \leq N$. Thus G/G' can be seen as the largest abelian quotient of G. It is called the *abelianization* of G. If $G = G'$, then G is said to be a *perfect group*.

For a group G we define its *center* to be $Z(G) = \{g \in G \mid [g,x] = 1 \text{ for all } x \in G\}$. It is easy to verify that $Z(G)$ is a characteristic subgroup of G.

Let G_1 and G_2 be groups. The *direct product* $G_1 \times G_2$ is the group whose elements are all pairs $(g_1, g_2) \in G_1 \times G_2$, and the operation is given by

$$(a_1, a_2)(b_1, b_2) = (a_1 b_1, a_2 b_2).$$

Proposition 3.2.5 *Let G, G_1 and G_2 be groups. Then $G \cong G_1 \times G_2$ if and only if there exist normal subgroups H_1 and H_2 of G such that $H_i \cong G_i$ for $i = 1,2$, $H_1 \cap H_2 = 1$ and $H_1 H_2 = G$.*

More generally, $G \cong G_1 \times G_2 \times \cdots \times G_n$ if and only if there exist normal subgroups $H_1, \ldots, H_n$ of G such that $H_i \cong G_i$, $G = H_1 H_2 \cdots H_n$, and

$$H_i \cap H_1 \cdots H_{i-1} H_{i+1} \cdots H_n = \{1\}$$

for all i. This follows from Proposition 3.2.5 by induction.

Let X be a non-empty set, F a group, and $\iota : X \to F$ a function. Then F, together with ι, is said to be a *free group* on X if for each function α from X to a group G there exists a homomorphism $\beta : F \to G$ such that $\alpha = \iota\beta$. It is easy to show that ι has to be injective. Up to isomorphism, there is precisely one free group on a given set X. It can be constructed as a group whose elements are reduced words in $X \cup X^{-1}$, and the operation

is concatenation, followed by reduction of terms of the form $x^{\pm 1}x^{\mp 1}$ if necessary. For further details we refer to [8].

Let X be a set and let F be a free group on X. Choose a subset Y of F, and let $R = \langle\langle Y \rangle\rangle$ be its normal closure in F. Then we say that the group $G = F/R$ is given by *generators* X and *relations* Y. We write $G = \langle X \mid Y \rangle$.

The following result is simple but useful in recognizing groups from their presentations:

Lemma 3.2.6 (von Dyck's Lemma) *Let G be a group generated by $x_1, \ldots, x_m$ satisfying relators $r_1 = 1, \ldots, r_n = 1$. Let H be a group generated by $y_1, \ldots, y_m$, and suppose that $r_i(y_1, \ldots, y_m) = 1$ for all $i = 1, \ldots, n$. Then there exists a uniquely determined epimorphism $\phi : G \to H$ with $x_j^{\phi} = y_j$ for all $j = 1, \ldots, m$.*

A sample application von Dyck's lemma will be given in the next section.

3.2.2 Examples of groups and GAP

In this section we present some important examples of groups. Along the way we show how to use GAP to construct groups and study their properties. More information on how to obtain GAP and apply its commands can be found at [5].

Cyclic groups

A group generated by one element is called a *cyclic group*. If G is a cyclic group, two possibilites can occur. Either G is infinite, in which case it is isomorphic to $(\mathbb{Z}, +)$, or it is finite of order n, in which case it is isomorphic to $(\mathbb{Z}_n, +)$. In multiplicative notation, cyclic groups will be denoted by C_∞ and C_n, respectively.

In general, if G is an arbitrary group and $g \in G$, then the order of the cyclic subgroup $\langle g \rangle$ of G is called the *order* of g, and denoted by $|g|$.

In GAP, one can construct cyclic groups in several different ways. The standard one is as follows:

```
gap> G := CyclicGroup( 6 );
<pc group of size 6 with 2 generators>
gap> Elements( G );
[ <identity> of ..., f1, f2, f1*f2, f2^2, f1*f2^2 ]
```

The list of the elements above may be a bit unexpected, as it does not indicate that the group in question is cyclic. Rather, it reflects the fact that C_6 is isomorphic to $C_2 \times C_3$, and `f1` and `f2` are the corresponding generators of these factors.

It is possible to examine basic properties of the group we constructed above:

```
gap> Order( G );
6
gap> IsCyclic( G );
true
gap> IsAbelian( G );
true
```

Another way is to represent a cyclic group of order n with a generator x and relation $x^n = 1$. We first construct a free group on $\{x\}$ and then factor out the relation $x^n = 1$. For $n = 6$, this goes as follows:

```
gap> F := FreeGroup( "x" );
<free group on the generators [ x ]>
gap> AssignGeneratorVariables( F );
#I Assigned the global variables [ x ]
gap> G := F / [ x^6 ];
<fp group on the generators [ x ]>
gap> Order( G );
6
gap> StructureDescription( G );
"C6"
gap> Elements( G );
[ <identity ...>, x^3, x^2, x^-1, x^-2, x ]
```

Note that the groups in the first and second example both represent C_6, yet, in GAP's eyes they are not identical objects, because GAP represents them in different ways. The first example represents C_6 as a `pc group`, and the second one as an `fp group`.

Abelian groups

Finitely generated *abelian groups* are classified by the following result:

Theorem 3.2.7 (Fundamental Theorem of Abelian Groups) *Every finitely generated abelian group is a direct product of cyclic groups*

$$C_{m_1} \times C_{m_2} \times \cdots \times C_{m_r} \times C_\infty^k,$$

where $m_i | m_{i+1}$ for all $i = 1, \ldots, r-1$. Two groups of this form are isomorphic if and only if the numbers $m_1, \ldots, m_r$ and k are the same for the two groups.

Alternatively, all finite abelian groups are direct products of cyclic groups of prime power order. This follows from the fact that if m and n are relatively prime then $C_m \times C_n \cong C_{mn}$. A group that is isomorphic to the direct product of a number of copies of C_p is called an *elementary abelian p-group*. Every elementary abelian p-group (written additively) is also a vector space over $\mathrm{GF}(p)$. The scalar multiplication is given by

$$\lambda x = \underbrace{x + \cdots + x}_{\lambda \text{ times}}.$$

For example, one can construct $C_2 \times C_4 \times C_{12}$ in GAP as follows:

```
gap> G := AbelianGroup( [2, 4, 12 ] );
<pc group of size 96 with 3 generators>
gap> AbelianInvariants( G );
[ 2, 3, 4, 4 ]
```

The last command tells us that our group is isomorphic to $C_2 \times C_3 \times C_4 \times C_4$. In general, `AbelianInvariants( G );` returns a cyclic decomposition of G^{ab}.

Symmetric groups

If X is a non-empty set, then the set of all bijections $X \to X$ becomes a group under the operation of composition. It is denoted by $\mathrm{Sym}\, X$. If X is a finite set, then we can write $X = \{1,2,\ldots,n\}$, and we use the abbreviation S_n for $\mathrm{Sym}\, X$ in this case. The group S_n is called the *symmetric group* on n letters. Its elements are *permutations* that can be written as products of *cycles* of the form $(x_1\, x_2 \ldots x_k)$ that represents the map $x_1 \mapsto x_2 \mapsto \cdots \mapsto x_k \mapsto x_1$, and all other elements are fixed. The order of S_n is $n!$. If $n > 2$, then S_n is clearly a non-abelian group.

Let us use GAP to play around with S_4 and its elements:

```
gap> S4 := SymmetricGroup( 4 );
Sym( [ 1 .. 4 ] )
gap> Order( S4 );
24
gap> el := Elements( S4 );
[ (), (3,4), (2,3), (2,3,4), (2,4,3), (2,4), (1,2), (1,2)(3,4), (1,2,3),
  (1,2,3,4), (1,2,4,3), (1,2,4), (1,3,2), (1,3,4,2), (1,3), (1,3,4),
  (1,3)(2,4), (1,3,2,4), (1,4,3,2), (1,4,2), (1,4,3), (1,4), (1,4,2,3),
  (1,4)(2,3) ]
gap> a := el[ 4 ];
(2,3,4)
gap> b := el[ 7 ];
(1,2)
gap> a * b;
(1,2,3,4)
gap> a^(-1);
(2,4,3)
gap> a^b;
(1,3,4)
gap> Order( a );
3
```

We can also present symmetric groups in terms of generators and relations. Here is an example:

Example 3.2.8 *Let $G = \langle x, y \mid x^2 = y^3 = (xy)^2 = 1\rangle$. We claim that $G \cong S_3$. Denote $a = (1\,2)$ and $b = (1\,2\,3)$. Then $a^2 = b^3 = (ab)^2 = 1$. By von Dyck's Lemma, there exists a surjective homomorphism $\phi : G \to \langle a, b\rangle = S_3$. Now consider G. We have that $yx = xy^2$, hence every element of G can be written as $x^m y^n$, where $0 \le m \le 1$, $0 \le n \le 2$. It follows that $|G| \le 6$. Comparing the orders, we conclude that ϕ must be an isomorphism between G and S_3. Another proof can be done with* GAP*:*

```
gap> F := FreeGroup("x", "y");;
gap> AssignGeneratorVariables(F);;
#I Assigned the global variables [ x, y ]
gap> G := F / [x^2, y^3, (x*y)^2];;
gap> StructureDescription(G);
"S3"
```

In general, the group S_n has a following presentation:

$$\langle x_1, \ldots, x_{n-1} \mid x_i^2 = 1, [x_i, x_j] = 1, x_i x_{i+1} x_i = x_{i+1} x_i x_{i+1} \text{ for all } i \text{ and } j \neq i \pm 1\rangle.$$

Here x_i corresponds to the transposition $(i \;\; i+1)$. This is left as an exercise.

Using GAP, one can also construct subgroups generated by certain sets of elements, and normal closures of subgroups. It is also possible to test memberships to subgroups.

```
gap> G := SymmetricGroup( 5 );
Sym( [ 1 .. 5 ] )
gap> H := Subgroup( G, [(1, 2), (1, 3)]);
Group([ (1,2), (1,3) ])
gap> Order( H );
6
gap> (1,2,3,4) in H;
false
gap> N := NormalClosure(G, H);
Group([ (2,3), (1,3,2), (2,4), (3,5) ])
gap> Order( N );
120
gap> StructureDescription( H );
"S3"
gap> StructureDescription( N );
"S5"
```

The *parity* of a permutation $g \in S_n$ is defined to be the parity of the number $n - c(g)$, where $c(g)$ is the number of cycles of g (including the cycles of length 1). We regard the parity as an element of $\mathbb{Z}_2$. One can show that the parity is a homomorphism from S_n onto the group $\mathbb{Z}_2$. Its kernel consists of all permutations of even parity. It is denoted by A_n and called the *alternating group* on n letters.

Alternating groups can be constructed with GAP:

```
gap> G := AlternatingGroup( 4 );
Alt( [ 1 .. 4 ] )
gap> Order( G );
12
```

One can also locate A_4 within the list of all normal subgroups of S_4:

```
gap> G := SymmetricGroup( 4 );
Sym( [ 1 .. 4 ] )
gap> norm := NormalSubgroups( G );
[ Sym( [ 1 .. 4 ] ), Group([ (2,4,3), (1,4)(2,3), (1,3)(2,4) ]),
       Group([ (1,4)(2,3), (1,3)(2,4) ]), Group(()) ]
gap> List( norm, StructureDescription );
[ "S4", "A4", "C2 x C2", "1" ]
gap> Q := G / norm[ 2 ];
Group([ f1 ])
gap> StructureDescription( Q );
"C2"
```

We can also construct the natural homomorphism $S_4 \to S_4/A_4$ as follows:

```
gap> G := SymmetricGroup( 4 );;
gap> norm:= NormalSubgroups( G );;
gap> N:=norm[ 2 ];
Group([ (2,4,3), (1,4)(2,3), (1,3)(2,4) ])
gap> hom := NaturalHomomorphismByNormalSubgroup( G, N );
[ (1,2,3,4), (1,2) ] -> [ f1, f1 ]
gap> Kernel( hom ) = N;
true
gap> StructureDescription( Image( hom ) );
"C2"
```

Linear groups

Let F be a field. The set of all invertible $n \times n$ matrices over F is a group under multiplication. It is called the *general linear group* of dimension n over F, and denoted by $\mathrm{GL}(n,F)$. By Galois' theorem, the order of a finite field is alwasy a prime power, and if q is a prime power, then there is, up to isomorphism, a unique field of order q. It is denoted by $\mathrm{GF}(q)$. The group $\mathrm{GL}(n,\mathrm{GF}(q))$ is also denoted as $\mathrm{GL}(n,q)$.

The determinant map $\det : \mathrm{GL}(n,F) \to F^\times$ is clearly a surjective homomorphism of groups. Its kernel is denoted by $\mathrm{SL}(n,F)$ and called the *special linear group* of dimension n over F. Its elements are precisely all the matrices $A \in \mathrm{GL}(n,F)$ with $\det A = 1$.

Let us consider some examples using GAP:

```
gap> G := GL( 2, 4);
GL(2,4)
gap> Order( G );
180
gap> el := Elements( G );;
gap> a := el[ 5 ];
[ [ 0*Z(2), Z(2)^0 ], [ Z(2^2), 0*Z(2) ] ]
gap> b := el[ 7 ];
[ [ 0*Z(2), Z(2)^0 ], [ Z(2^2), Z(2^2) ] ]
gap> Determinant( a );
Z(2^2)
gap> a * b^2;
[ [ Z(2^2)^2, Z(2)^0 ], [ Z(2^2)^2, Z(2^2)^2 ] ]
gap> H := SL( 2, 4 );
SL(2,4)
gap> Order( H );
60
gap> StructureDescription( H );
"A5"
```

Proposition 3.2.9 $|\mathrm{GL}(n,q)| = (q^n-1)(q^n-q)\cdots(q^n-q^{n-1})$.

PROOF. A matrix is invertible if and only if its rows are linearly independent. This holds if and only if the first row is non-zero and, for $k = 2,\ldots,n$, the k-th row is not in the subspace spanned by the first $k-1$ rows. The number of possible rows is q^n, and the number lying in any k-dimensional subspace is q^k. So the number of choices for the first row is q^n-1, and for $k = 2,\ldots,n$, the number of choices for the k-th row is q^n-q^{k-1}. Multiplying these, we get the formula. ∎

Corollary 3.2.10 $|\mathrm{SL}(n,q)| = |\mathrm{GL}(n,q)|/(q-1)$.

PROOF. Let $F = \mathrm{GF}(q)$. We already saw above that $\mathrm{GL}(n,q)/\mathrm{SL}(n,q) \cong F^\times$, and this gives the result. ∎

Dihedral groups

A *symmetry* of a figure in Euclidian space is a rigid motion (or a combination of a rigid motion with reflection) of the space that carries the figure to itself. If we think of a rigid motion as a linear map of the real vector space, then it can be represented by a matrix. Alternatively, if we label the vertices of the figure, then a symmetry can be represented as a permutation of these labels.

The group of symmetries of a regular n-gon is called a *dihedral group* D_{2n}. If a denotes the rotation around the center by the angle $2\pi/n$, and b the reflection over a chosen diagonal, then the elements of D_{2n} can be written uniquely in the form $a^k b^\ell$ where $0 \leq k < n$ and $\ell \in \{0,1\}$. Thus $|D_{2n}| = 2n$. The group D_{2n} has a presentation

$$D_{2n} = \langle a, b \mid a^n = 1, b^2 = 1, a^b = a^{-1} \rangle.$$

In GAP, one can construct dihedral groups directly by

```
gap> G := DihedralGroup( 6 );
<pc group of size 6 with 2 generators>
gap> Order( G );
6
```

Another way is to present it by generators and relations. This is done by first constructing a free group on two generators and then factor out the relations.

```
gap> F := FreeGroup( "a", "b" );
<free group on the generators [ a, b ]>
gap> AssignGeneratorVariables(F);
#I Assigned the global variables [ a, b ]
gap> H := F / [ a^3, b^2, a^b / a^(-1) ];
<fp group on the generators [ a, b ]>
gap> StructureDescription( H );
"S3"
```

The last command tells us that $D_6 \cong S_3$. We can compare both constructions of D_6 above and see that they are not identical objects in GAP, yet they are isomorphic:

```
gap> H = G;
false
gap> IsomorphismGroups(G, H);
[ f1, f2 ] -> [ b, a ]
```

The reason is that GAP represents D_6 in two different ways, first as a `pc group` and then as an `fp group`. The reader should consult GAP's manual for further details.

3.2.3 Automorphisms

An *automorphism* of a group G is an isomorphism G to itself. There are special types of automorphisms called *conjugations* or *inner automorphisms*; they are of the form $c_g : x \mapsto g^{-1}xg$.

Proposition 3.2.11 *Let G be a group.*

(a) The set $\mathrm{Aut}(G)$ *of all automorphisms of G is a group under composition (from the right). This is the automorphism group of G.*

(b) The set $\mathrm{Inn}(G)$ *of all inner automorphisms of G is a normal subgroup of* $\mathrm{Aut}\,G$*. This is called the inner automorphism group of G.*

(c) $\mathrm{Inn}(G) \cong G/Z(G)$.

The proof is straightforward and we leave it as an exercise. The group $\mathrm{Out}(G) = \mathrm{Aut}(G)/\mathrm{Inn}(G)$ is the *outer automorphism group* of G. Note that its elements are not automorphisms, but rather right cosets $\mathrm{Inn}(G)\alpha$, where $\alpha \in \mathrm{Aut}(G)$.

GAP can deal with automorphisms very naturally:

```
gap> G := DihedralGroup( 12 );
<pc group of size 12 with 3 generators>
gap> A := AutomorphismGroup( G );
<group of size 12 with 3 generators>
gap> Elements( A );
[ [ f1*f3, f1*f2*f3^2, f1*f2*f3 ] -> [ f1*f2, f1*f3^2, f1 ],
  [ f1*f3, f1*f2*f3^2, f1*f2*f3 ] -> [ f1*f2*f3^2, f1*f3, f1 ],
  [ f1*f3, f1*f2*f3^2, f1*f2*f3 ] -> [ f1, f1*f2*f3, f1*f2 ],
  [ f1*f3, f1*f2*f3^2, f1*f2*f3 ] -> [ f1*f3, f1*f2*f3^2, f1*f2 ],
  [ f1*f3, f1*f2*f3^2, f1*f2*f3 ] -> [ f1*f2, f1*f3^2, f1*f3 ],
  [ f1*f3, f1*f2*f3^2, f1*f2*f3 ] -> [ f1*f2*f3, f1, f1*f3 ],
  [ f1*f3, f1*f2*f3^2, f1*f2*f3 ] -> [ f1*f3, f1*f2*f3^2, f1*f2*f3 ],
  [ f1*f3, f1*f2*f3^2, f1*f2*f3 ] -> [ f1*f3^2, f1*f2, f1*f2*f3 ],
  [ f1*f3, f1*f2*f3^2, f1*f2*f3 ] -> [ f1*f2*f3, f1, f1*f3^2 ],
  [ f1*f3, f1*f2*f3^2, f1*f2*f3 ] -> [ f1*f2*f3^2, f1*f3, f1*f3^2 ],
  [ f1*f3, f1*f2*f3^2, f1*f2*f3 ] -> [ f1, f1*f2*f3, f1*f2*f3^2 ],
  [ f1*f3, f1*f2*f3^2, f1*f2*f3 ] -> [ f1*f3^2, f1*f2, f1*f2*f3^2 ] ]
gap> StructureDescription( A );
"D12"
gap> inn := InnerAutomorphismsAutomorphismGroup( A );
<group with 3 generators>
gap> Order( inn );
6
gap> IsomorphismGroups( inn, G / Center( G ) );
CompositionMapping( [ (2,6)(3,5), (1,3,5)(2,4,6), (1,5,3)(2,6,4) ] ->
[ f1, f2^2, f2 ], <action isomorphism> )
```

Next we compute some automorphism groups:

Proposition 3.2.12 $\mathrm{Aut}\, C_n \cong C_{\phi(n)}$, *where* ϕ *is Euler's totient function.*

PROOF. Let $C_n = \langle g \rangle$ and take $\alpha \in \mathrm{Aut}\, G$. Then $g^\alpha = g^i$ for some $0 \le i \le n-1$, and since $\langle g^i \rangle = C_n$, this can only happen if $\gcd(i,n) = 1$. Conversely take an endomorphism α of C_n with $g^\alpha = g^i$, where $\gcd(i,n) = 1$. Then it is elementary to see that α is an automorphism. Thus the map $\mathrm{Aut}\, C_n \to \mathbb{Z}_n^\times$ given by $\alpha \mapsto i$ is an isomorphism of groups. This proves the result. ∎

Proposition 3.2.13 $\mathrm{Aut}(C_p^n) \cong \mathrm{GL}(n,p)$.

PROOF. This follows from the fact that C_p^n is an n-dimensional vector space over $\mathrm{GF}(p)$. ∎

3.2.4 Group actions and Sylow's theorems

Sylow theorems are central in the theory of finite groups, as they describe the structure of such groups in terms of their subgroups of prime power order. These theorems are closely related to another fundamental notion of group theory, actions.

Actions

An *action* of a group G on a non-empty set X is a map $\mu\colon X\times G\to X$ satisfying the following rules:

$$\mu(\mu(x,g),h)=\mu(x,gh),$$
$$\mu(x,1)=x$$

for all $x\in X$ and $g,h\in G$. We usually suppres μ and write $\mu(x,g)$ as xg. It is clear that the above definition is equivalent to the fact that the map $G\to \operatorname{Sym} X$ given by $g\mapsto(x\mapsto xg)$ is a homomorphism of groups. An action μ is *faithful* if the condition that $\mu(x,g)=\mu(x,h)$ for all $x\in X$ implies $g=h$.

Let G act on X. The relation $\equiv$ defined on X by $x\equiv y\iff \exists g\in G: xg=y$ is an equivalence relation on X. The equivalence class of $x\in X$ is called the *orbit* of x, and is denoted by $\operatorname{orb}_G(x)$. The set of orbits of G on X will be denoted by X/G. The action is said to be *transitive* if it has only one orbit, i.e., $|X/G|=1$. For $x\in X$, the *stabilizer* of x is

$$\operatorname{stab}_G(x)=\{g\in G\mid xg=x\}.$$

It is easy to see that $\operatorname{stab}_G(x)$ is a subgroup of G.

Example 3.2.14 *A group G acts on itself by right multiplication, i.e., we have an action $G\times G\to G$ given by $(g,h)\mapsto g\cdot h=gh$. It is not hard to see that this action is transitive and faithful.*

```
gap> G := Group((1,2,3),(2,3,4));;
gap> el := Elements( G);;
gap> OnRight(el[2], el[3]) = el[2] * el[3];
true
gap> orbit := Orbit(G, el[7], OnRight);
[ (1,3,2), (), (1,4,2), (1,2,3), (2,3,4), (1,4,3), (1,2)(3,4),
  (1,3)(2,4), (2,4,3), (1,4)(2,3), (1,3,4), (1,2,4) ]
gap> Size( orbit ) = Order( G );
true
```

Example 3.2.15 *A group G acts on itself by conjugation, i.e., $(g,h)\mapsto g^h$. The orbits of this actions are called the conjugacy classes of G. The stabilizer of $g\in G$ is denoted by $C_G(g)$ and called the centralizer of g in G.*

```
gap> G := DihedralGroup( 8 );;
gap> ConjugacyClasses( G );
[ <identity> of ...^G, f1^G, f2^G, f3^G, f1*f2^G ]
gap> el := Elements( G );;
gap> Centralizer( G, Subgroup( G, [ el[ 5 ] ] ) );
Group([ f1*f2, f3 ])
```

More generally, any subgroup $H \leq G$ acts on G by conjugation. At the other end of the scale, if N is a normal subgroup of G, then G, by definition, acts on N by conjugation.

Example 3.2.16 *A subgroup H of a group G acts on the set of all subgroups of G by conjugation; $(K,h) \mapsto K^h$. If $K \leq G$, then the stabilizer of K is under this action is the normalizer of K:*

$$N_H(K) = \{h \in H \mid K^h = K\}.$$

Example 3.2.17 *Let H be a subgroup of G and $H\backslash G$ the set of all right cosets of H in G. Then G acts on $H\backslash G$ by right multiplication: $(Hx)\cdot g = Hxg$.*

```
gap> G := Group((1, 2, 3, 4, 5), (1, 2) );;
gap> H := Subgroup( G, [ (1, 2) ] );;
gap> Index( G, H );
60
gap> act := FactorCosetAction( G, H );
<action epimorphism>
gap> Range( act );
<permutation group of size 120 with 2 generators>
gap> Kernel( act );
Group(())
```

Example 3.2.18 *Let X be a non-empty set and $G \leq \mathrm{Sym}\, X$. Then G acts on points of X by the rule $(x,g) \mapsto x^g$.*

```
gap> G := Group( (1, 2, 3), (2, 3, 4) );;
gap> Orbit(G, 1, OnPoints);
[ 1, 2, 3, 4 ]
```

Let G be a finite group acting on a set X. One can observe that there is a 1-1 correspondence between the elements of $\mathrm{orb}_G(x)$ and the right cosets of $\mathrm{stab}_G(x)$ in G. This implies the following fundamental result:

Theorem 3.2.19 (Orbit-stabilizer theorem) *Let G be a finite group acting on a set X. Choose $x \in X$. Then $|\mathrm{orb}_G(x)| \cdot |\mathrm{stab}_G(x)| = |G|$.*

In the special case when G acts on itself by conjugation, we obtain:

Corollary 3.2.20 (Class equation) *Let G be a finite group and let $x_1, \ldots, x_r$ be the representatives of conjugacy classes of non-central elements of G. Then*

$$|G| = |Z(G)| + \sum_{i=1}^{r} |G : C_G(x_i)|.$$

For $g \in G$ denote by $\mathrm{fix}(g)$ the number of fixed points of g (considered as an element of $\mathrm{Sym}\, X$). We have:

Theorem 3.2.21 (Orbit-counting Lemma) *Let a finite group G act on a set X. Then*

$$|X/G| = \frac{1}{|G|} \sum_{g \in G} \mathrm{fix}(g).$$

PROOF. We will count the pairs $(x, g) \in X \times G$ with the property that $xg = x$; let us call these pairs good pairs. On one hand, a given $g \in G$ is a member of $\mathrm{fix}(g)$ good pairs, hence the total number of good pairs is $\sum_{g \in G} \mathrm{fix}(g)$. On the other hand, $x \in X$ is a member of $|\mathrm{stab}_G(x)|$ good pairs. The orbit of x thus produces $|\mathrm{orb}_G(x)| \cdot |\mathrm{stab}_G(x)| = |G|$ good pairs, hence there are $|X/G| \cdot |G|$ good pairs in total. We get the result. ∎

Sylow theorems

Since the action of G on itself by right multiplication is faithful, we have that the corresponding homomorphism $G \to \mathrm{Sym}\, G$ is injective. In particular, we have:

Theorem 3.2.22 (Cayley's theorem) *Every finite group is isomorphic to a subgroup of S_n for some positive integer n.*

Another classical result that can be proved using actions is Cauchy's theorem which provides a basis for Sylow theorems. It goes as follows:

Theorem 3.2.23 (Cauchy's theorem) *Let G be a finite group. If a prime p divides $|G|$, then G contains an element of order p.*

Theorem 3.2.24 (Sylow's theorem) *Let G be a group of order $p^a \cdot m$, where m is not divisible by the prime p. Then the following holds:*

1. *G contains at least one subgroup of order p^a. Any two subgroups of this order are conjugate in G. They are called the Sylow p-subgroups of G.*
2. *For each $n \leq a$, G contains at least one subgroup of order p^n. Every such subgroup is contained in a Sylow p-subgroup.*
3. *Let s_p be the number of Sylow p-subgroups of G. Then $s_p \equiv 1 \mod p$ and s_p divides m.*

This result has numerous consequences for the structure of finite groups, see the problems at the end of this section. We mention here that GAP can compute a Sylow p-subgroup of a given group as follows:

```
gap> G := SymmetricGroup( 4 );;
gap> P := SylowSubgroup( G, 2 );
Group([ (1,2), (3,4), (1,3)(2,4) ])
```

How many Sylow 2-subgroups of S_4 are there? A consequence of Sylow's theorem is also that if P is a Sylow p-subgroup of G, then $s_p = |G : N_G(P)|$. Thus:

```
gap> Index( G, Normalizer( G, P ) );
3
```

Thus there are three Sylow 2-subgroups of S_4. All of them are conjugate to P:

```
gap> ConjugacyClassSubgroups( G, P );
Group( [ (1,2), (3,4), (1,3)(2,4) ] )^G
gap> Elements( last );
[Group([ (1,2), (3,4), (1,3)(2,4) ]), Group([ (2,3), (1,4), (1,3)(2,4)]),
  Group([ (1,3), (2,4), (1,4)(2,3) ]) ]
```

A finite group is said to be a *p-group* if every element has order a power of p. Equivalently, the order of the group is p^n for some n (exercise).

Proposition 3.2.25 *Let G be a p-group. Then $Z(G)$ is non-trivial, and G contains a normal subgroup of order p.*

PROOF. We may assume that G is non-abelian of order p^n. Let $x_1, \ldots, x_r$ be the representatives of non-central conjugacy classes of G. By the Class Equation,

$$p^n = |Z(G)| + \sum_{i=1}^{r} |G : C_G(x_i)|.$$

Since $C_G(x_i) \neq G$, the prime p divides $|G : C_G(x_i)|$ for all $i = 1, \ldots, r$. It follows that p divides $|Z(G)|$. The rest is now straightforward. ∎

Example 3.2.26 *There is only one group of order p, namely C_p. Let us show that all groups of order p^2 are abelian (hence there are only two possibilities, $C_p \times C_p$ and C_{p^2}). Suppose there exists a non-abelian group G of order p^2. Then $Z(G) \cong C_p$ and $G/Z(G) \cong C_p$. Let $Z(G)x$ be a generator of $G/Z(G)$. Then $G = Z(G)\langle x \rangle$, but the latter group is abelian, which is a contradiction.*

Example 3.2.27 *Let us classify all groups of order pq, where p and q are distinct primes (for $p = q$ see Example 3.2.26). Assume that $p > q$. Let P be a Sylow p-subgroup, and Q a Sylow q-subgroup of G. Then Sylow's theorem implies that $s_p = 1$, i.e., P is a normal subgroup of G. Similarly, $s_q \in \{1, p\}$, and $s_p = 1$ if and only if $p \equiv 1 \mod q$. We separate the two cases:*

Suppose $s_q = 1$. Denote $P = \langle a \rangle$ and $Q = \langle b \rangle$. Then $a^b = a^k$ and $b^a = b^\ell$ for some integers k and ℓ. Therefore $a^{k-1} = [a,b] = b^{-\ell+1}$. Since the orders and b are coprime, it follows that $[a,b] = 1$, hence $G \cong C_p \times C_q \cong C_{pq}$.

Now let $s_q = p$, that is, let q divide $p-1$. We still have $a^b = a^k$. By induction, $a^{b^s} = a^{k^s}$. Since $|b| = q$, we conclude that $k^q \equiv 1 \mod p$. There are exactly q solutions to this equation; if k is one of them, the others are powers of k. By replacing b by a power of itself we see that all these solutions give rise to the same group, namely, a group with presentation

$$\langle a, b \mid a^p = b^q = 1, a^b = a^k \rangle$$

for some k satisfying $k^q \equiv 1 \mod p$, $k \not\equiv 1 \mod p$.

More on finite p-groups will be discussed later on. We conclude with two useful lemmas which are of similar nature:

Lemma 3.2.28 (The Frattini argument) *Let G be a group and H a finite normal subgroup. If P is a Sylow p-subgroup of H, then $G = N_G(P)H$.*

PROOF. For $g \in G$ we have $P^g \leq H$ and $P^g = P^h$ for some $h \in H$. Thus $gh^{-1} \in N_G(P)$. ∎

Lemma 3.2.29 *If P is a Sylow p-subgroup of a finite group G and $N_G(P) \leqslant H \leqslant G$, then $H = N_G(H)$.*

PROOF. Clearly $P \leqslant H \lhd N_G(H)$. By Frattini's argument we have that $N_G(H) = N_{N_G(H)}(P)H$. But $N_{N_G(H)}(P) \leq N_G(P) \leq H$, hence the result. ∎

3.2.5 An estimate of the number of finite groups

In this short section we derive a rough bound for the number of of groups of order n.

Lemma 3.2.30 *A group G of order n can be generated by a set of at most $\log_2 n$ elements.*

PROOF. Choose a non-trivial element $g_1 \in G$, and let $G_1 = \langle g_1 \rangle$. If $G_1 = G$, then stop. Otherwise choose $g_2 \in G - G_1$ and let $G_2 = \langle g_1, g_2 \rangle$. Repeat the procedure until we find $g_1, \ldots, g_k \in G$ such that $G = \langle g_1, \ldots, g_k \rangle$.

We prove that $|G_i| \geq 2^i$ for all $i = 1, \ldots, k$; this suffices to prove our lemma. The proof is by induction on i, the case $i = 1$ being obvious, Suppose that $|G_i| \geq 2^i$. Since $|G_i|$ divides $|G_{i+1}|$ and $G_i \neq G_{i+1}$, we have $|G_{i+1}| \geq 2|G_i| \geq 2^{i+1}$, as required. ∎

Proposition 3.2.31 *The number of groups of order n is at most $n^{n \log_2 n}$.*

PROOF. By Cayley's theorem, every group of order n can be embedded as a subgroup of S_n, and can be generated by $k = \lfloor \log_2 n \rfloor$ elements. There are at most $n!$ choices for each g_i, so the number of subgroups of S_n is at most

$$(n!)^k \leq (n^n)^{\log_2 n} = n^{n \log_2 n},$$

as required. ∎

GAP offers a `Small Groups` library which gives access to all groups of certain "small" orders. The groups are sorted by their orders and they are listed up to isomorphism; that is, for each of the available orders a complete and irredundant list of isomorphism type representatives of groups is given. The library also has an identification function: it returns the library number of a given group. More on this can be found in GAP's manual. Here are some examples.

```
gap> AllSmallGroups( 16 );;
gap> NrSmallGroups( 512 );
10494213
gap> AllSmallGroups(Size, 16, IsAbelian, true);
[ <pc group of size 16 with 4 generators>,
  <pc group of size 16 with 4 generators>,
  <pc group of size 16 with 4 generators>,
  <pc group of size 16 with 4 generators>,
  <pc group of size 16 with 4 generators> ]
gap> List( last, StructureDescription );
[ "C16", "C4 x C4", "C8 x C2", "C4 x C2 x C2", "C2 x C2 x C2 x C2" ]
gap> G := DihedralGroup( 64 );
<pc group of size 64 with 6 generators>
gap> IdGroup( G );
[ 64, 52 ]
gap> H := SmallGroup( 64, 52 );
<pc group of size 64 with 6 generators>
gap> G = H;
false
gap> StructureDescription( H );
"D64"
```

3.2.6 Jordan-Hölder theorem

A group G is *simple* if $\{1\}$ and G are the only normal subgroups of G. The abelian simple groups are precisely C_p where p is a prime (exercise). More examples of finite simple groups will be exhibited in Section 3.3.

A *composition series* of a group G is a sequence of subgroups

$$\{1\} = G_0 \triangleleft G_1 \triangleleft G_2 \triangleleft \cdots \triangleleft G_r = G$$

such that all the factors G_{i+1}/G_i are simple groups. A related concept is that of *chief series*, where G_i are all normal in G and each G_{i+1}/G_i is a minimal normal subgroup of G/G_i.

The *Correspondence Theorem* says that if N is a normal subgroup of G then there is a bijection between subgroups of G/N and subgroups of G containing N. The bijection is canonical in the sense that all subgroups of G/N are of the form H/N, where H is a subgroup of G containing N. This result enables construction of a composition series of a finite group G as follows. Start with the series $\{1\} \triangleleft G$. If G is simple, we are done. Otherwise there is a proper non-trivial normal subgroup N of G. Now we repeat the procedure with $\{1\} \triangleleft N$ and $N \triangleleft G$. More precisely, if we have $G_i \triangleleft G_{i+1}$ and the corresponding quotient is not simple, then we choose (by the Correspondence Theorem) $N/G_i \triangleleft G_{i+1}/G_i$ with $N \neq G_i$ and $N \neq G_{i+1}$. In this way we refine the series, and since the group is finite, the procedure eventually results in a composition series of G. Given a composition series of G as above, we have r simple groups G_{i+1}/G_i.

Theorem 3.2.32 (Jordan-Hölder Theorem) *Any two composition series of a finite group G give rise, up to order and isomorphism type, to the same list of composition factors.*

PROOF. The proof is by induction on $|G|$. Let

$$G = G_0 \triangleright G_1 \triangleright G_2 \triangleright \cdots \triangleright G_r = \{1\}$$

and

$$G = H_0 \triangleright H_1 \triangleright G_2 \triangleright \cdots \triangleright H_s = \{1\}$$

be two composition series of G. If $G_1 = H_1$, then the parts of the series below this term are two composition series of G_1 and by induction they have the same length and composition factors. So assume from here on that $G_1 \neq H_1$. Let $K_2 = G_1 \cap H_1$. Let

$$K_2 \triangleright K_3 \triangleright \cdots \triangleright K_t = \{1\}$$

be a composition series of K_2. The group G_1H_1 is a normal subgroup of G and $G_1 < G$. It follows that $G = G_1H_1$. Therefore $G/G_1 = G_1H_1/G_1 \cong H_1/K_2$, and similarly also $G/H_1 \cong G_1/K_2$. Thus

$$G_1 \triangleright G_2 \triangleright \cdots \triangleright G_r = \{1\}$$

and

$$G_1 \triangleright K_2 \triangleright K_3 \triangleright \cdots \triangleright K_t = \{1\}$$

are two composition series of G_1 and hence they have the same length and same composition factors. A similar statement holds true for H_1, so each of the given series for G has the composition factors of K_2 together with G/G_1 and G/H_1. Therefore the result holds. ∎

Let us calculate a composition series of D_{32}:

```
gap> G := DihedralGroup( 32 );
<pc group of size 32 with 5 generators>
gap> cs := CompositionSeries( G );
[ Group([ f1, f2, f3, f4, f5 ]), Group([ f2, f3, f4, f5 ]),
  Group([ f3, f4, f5 ]), Group([ f4, f5 ]), Group([ f5 ]), Group([ ]) ]
gap> List( [1..5], i -> StructureDescription( cs[ i ] / cs[ i + 1 ] ) );
[ "C2", "C2", "C2", "C2", "C2" ]
```

The result is not surprising as D_{32} is a 2-group.

Solvable groups

A finite group is said to be *solvable* if all of its composition factors are cyclic of prime order. One can prove the following:

Theorem 3.2.33 *A finite group G is solvable if it has a series*

$$G = G_0 \triangleright G_1 \triangleright G_2 \triangleright \cdots \triangleright G_r = \{1\}$$

with all G_i/G_{i+1} abelian.

The statement of Theorem 3.2.33 is usually taken as the definition of solvable groups in the infinite case. Every abelian group is solvable. The smallest non-abelian solvable group is $1 \lhd A_3 \lhd S_3$. The smallest non-solvable group is A_5. The *derived length* of a solvable group G is the length of the shortest abelian series of G. A group is called *metabelian* if its derived length is no more than two.

Lemma 3.2.34 *The following hold:*

1. *A subgroup of a solvable group is solvable.*
2. *A homomorphic image of a solvable group is solvable.*
3. *If a normal subgroup and its factor are solvable, then the group is solvable.*

Lemma 3.2.35 *A product of two normal solvable subgroups of a group is again solvable.*

PROOF. Let $H \lhd G$ and $K \lhd G$ be solvable. Then $(KH)/K \simeq H/(H \cap K)$ is solvable by (2) above and consequently KH is solvable by (3). ∎

The following shows that A_5 is the only non-solvable group of order 60:

```
gap> l60 := AllSmallGroups( 60 );;
gap> List( l60, IsSolvable );
[ true, true, true, true, false, true, true, true, true, true, true,
  true, true ]
gap> notsolv := Filtered( l60, G -> not IsSolvable( G ) );
[ Group([ (1,2,3,4,5), (1,2,3) ]) ]
gap> StructureDescription( notsolv[ 1 ] );
"A5"
```

Examples of solvable groups include the following:

Theorem 3.2.36 *Let p, q, r be primes. Then all groups of orders $p^m q^n$ or pqr are solvable.*

We skip the proof. Solvability of groups of order $p^m q^n$ is also refered to as the *Burnside's $p^m q^n$-theorem*. It is proved using character theory.

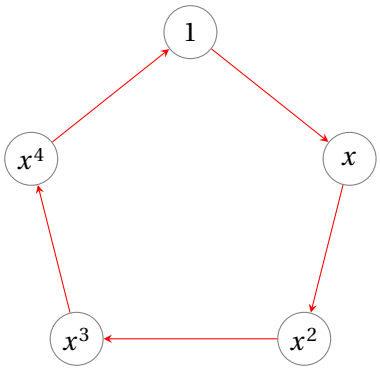
A celebrated theorem by Feit and Thompson says that *every group of odd order is solvable*. The proof is very long (about 255 pages) and represents a milestone in the classification of finite simple groups as it was a first significant indication that such a classification might be possible. We mention here that the Feit-Thompson theorem was recently reproved using interactive theorem prover `Coq`.

3.2.7 How to draw a group?

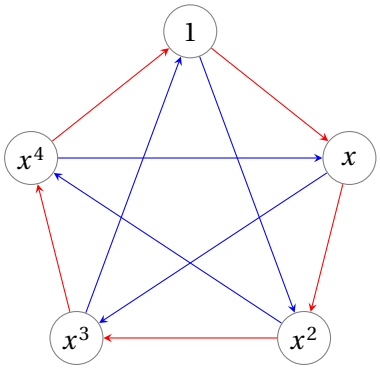
In this section we assume the reader is familiar with basic terminology of graph theory. Let G be a group generated by a set S. The *Cayley graph* $\Gamma = \mathrm{Cay}(G,S)$ is a colored directed graph given as follows: the vertex set of Γ is identified with G. To each $s \in S$ we assign a color c_s. The vertices g and sg are joined by a directed edge of color c_s for all $g \in G$ and $s \in S$. The set S is usually assumed to be finite, symmetric (i.e., $S = S^{-1}$) and not containing the identity element of the group. In this case, the uncolored Cayley graph is an ordinary graph: its edges are not oriented and it does not contain loops (single-element cycles).

One can modify the above definiton to the case when S is a set of elements of G that does not generate G. We still get a graph, but it may not be connected. From the definition of Cayley graphs it also follows that the Cayley graph of a given group clearly depends on the choice of a generating set S. Here are some examples that illustrate this.

Example 3.2.37 *If we take the cyclic group $C_n = \langle x \rangle$ of order n and $S = \{x, x^{-1}\}$, then $\mathrm{Cay}(C_n, S)$ is an undirected cycle $\mathscr{C}_n$ of length n. If we take $S = \{x\}$, then, unless $n = 2$, the corresponding Cayley graph is a directed cycle of length n. In the case $n = 5$ the diagram is as follows:*

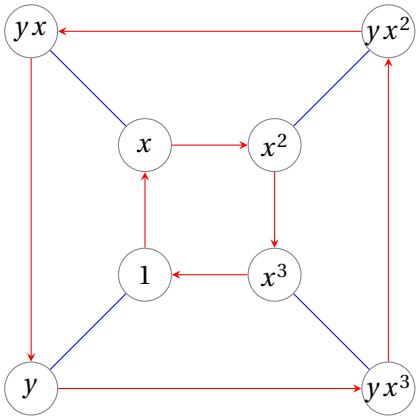


Every red directed edge between x^k and x^{k+1} resembles the fact that $x^{k+1} = x \cdot x^k$. If we take $S = \{x, x^2\}$, the corresponding graph is a directed circulant graph with jumps 1 and 2. Here is the diagram for $n = 5$:

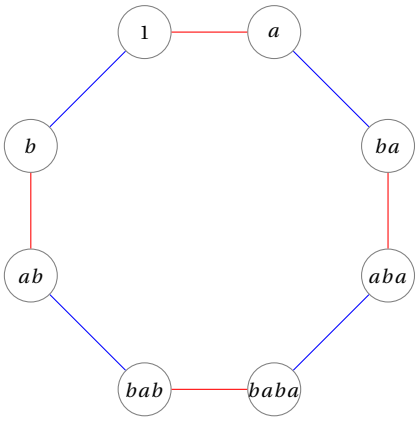


If we take $S = \{x^{\pm 1}, x^{\pm 2}\}$ then we get an undirected circulant graph with jumps 1 and 2. It turns out that undirected circulant graphs are precisely Cayley graphs of cyclic groups with respect to symmetric generating sets.

Example 3.2.38 *The dihedral group of order 8 has a presentation $D_8 = \langle x, y \mid x^4 = y^2 = 1, x^y = x^{-1} \rangle$. The Cayley graph* $\mathrm{Cay}(D_8, \{x, y\})$ *looks as follows:*



The red arrows represent multiplication by x from the left, and the blue edges represent multiplication by y; since $y = y^{-1}$, the blue edges are undirected. The dihedral group of order 8 can be also given by the following presentation: $D_8 = \langle a, b \mid a^2 = b^2 = 1, (ab)^2 = (ba)^2 \rangle$. In this case, $\mathrm{Cay}(D_8, \{a, b\})$ *is as follows:*



Cayley graphs can be constructed within GAP using a package called GRAPE. This package has to be loaded into GAP using LoadPackage. After that all the commands of the package are available. One can then construct Cayley graphs Cay(G,S); the result is a *record* that contains several attributes of the graph; we refer to GAP's manual for further details on records, and GRAPE's manual for further commands. Here we show how to construct a Cayley Graph of A_4 with respect to the generating set $\{(1\,2\,3),(1\,2\,4)\}$, and compute its adjacency matrix.

```
gap> LoadPackage("grape");;
--------------------------------------------------------------------------
Loading GRAPE 4.6.1 (GRaph Algorithms using PErmutation groups)
by Leonard H. Soicher (http://www.maths.qmul.ac.uk/~leonard/).
Homepage: http://www.maths.qmul.ac.uk/~leonard/grape/
--------------------------------------------------------------------------
gap> cay := CayleyGraph(AlternatingGroup(4), [(1,2,3),(1,2,4)]);
rec( adjacencies := [ [ 5, 6, 7, 10 ] ], group := Group([ (1,5,7)(2,4,8)
  (3,6,9)(10,11,12), (1,2,3)(4,7,10)(5,9,11)(6,8,12) ]), isGraph := true,
  isSimple := true,
  names := [ (), (2,3,4), (2,4,3), (1,2)(3,4), (1,2,3), (1,2,4), (1,3,2),
      (1,3,4), (1,3)(2,4), (1,4,2), (1,4,3), (1,4)(2,3) ], order := 12,
  representatives := [ 1 ],
  schreierVector := [ -1, 2, 2, 1, 1, 1, 1, 1, 2, 2, 2, 1 ] )
gap> CollapsedAdjacencyMat(cay);
[ [ 0, 0, 0, 0, 1, 1, 1, 0, 0, 1, 0, 0 ],
  [ 0, 0, 0, 1, 0, 0, 0, 1, 1, 1, 0, 0 ],
  [ 0, 0, 0, 1, 0, 0, 1, 0, 0, 0, 1, 1 ],
  [ 0, 1, 1, 0, 0, 0, 0, 1, 0, 0, 1, 0 ],
  [ 1, 0, 0, 0, 0, 0, 1, 0, 1, 0, 1, 0 ],
  [ 1, 0, 0, 0, 0, 0, 0, 1, 0, 1, 0, 1 ],
  [ 1, 0, 1, 0, 1, 0, 0, 0, 0, 0, 0, 1 ],
  [ 0, 1, 0, 1, 0, 1, 0, 0, 0, 0, 0, 1 ],
  [ 0, 1, 0, 0, 1, 0, 0, 0, 0, 1, 1, 0 ],
  [ 1, 1, 0, 0, 0, 1, 0, 0, 1, 0, 0, 0 ],
  [ 0, 0, 1, 1, 1, 0, 0, 0, 1, 0, 0, 0 ],
  [ 0, 0, 1, 0, 0, 1, 1, 1, 0, 0, 0, 0 ] ]
```

3.2.8 Problems

1. Supply the missing proofs in this section.

2. Let H be a subgroup of a group G with $|G:H|=2$. Prove that H is a normal subgroup of G.
3. Is it always true that if H is a subgroup of G with prime index, then $H \triangleleft G$?
4. Let p be the smallest prime that divides the order of a finite group G. If H is a subgroup of G of index p, then H is normal in G.
5. Find a group G and subgroups H and K with the property that $H \triangleleft K \triangleleft G$, but H is not normal in G.
6. Let H and K be subgroups of finite index in G. Prove that $|G:H\cap K| \leq |G:H|\cdot|G:K|$, with equality if and only if $G = HK$.
7. If H is a subgroup of G of finite index, then H contains a subgroup of finite index which is normal in G.
8. A group in which every non-trivial element has order 2 is abelian.
9. Let a and b be elements of order 2 of a finite group G. Prove that $\langle a,b\rangle$ is a dihedral group.
10. Find all subgroups of D_{12}. Which of these are normal subgroups?
11. Show that $\mathrm{GL}(2,2) \cong S_3$.
12. What is the largest order of an element of S_{12}?
13. Give an example of two non-isomorphic groups whose automorphism groups are isomorphic.
14. If G is a non-cyclic abelian group, then $\mathrm{Aut}\, G$ is non-abelian.
15. Let G act transitively on a set X, let H be a subgroup of G, and choose $x \in X$. Prove that the following are equivalent:

    (a) $G = H\,\mathrm{stab}_G(x)$,
    (b) $G = \mathrm{stab}_G(x)H$,
    (c) H acts transitively on X.

    Use this to find an alternative proof of Frattini's argument.
16. Let H be a subgroup of G. Show that $N_G(H)/C_G(H)$ is isomorphic to a subgroup of $\mathrm{Aut}\, H$.
17. Find the center and all conjugacy classes of D_{2n}.
18. Let P be a Sylow p-subgroup of a finite group G. Prove that if N is a normal subgroup of G, then $P\cap N$ is a Sylow p-subgroup of N, and PN/N is a Sylow p-subgroup of G/N.
19. Let P be a Sylow p-subgroup of a finite group G and $H \leq G$. Is it true that $P\cap H$ is always a Sylow p-subgroup of H?

20. Show that a group of order 40 cannot be simple. Do the same for groups of order 84.
21. Prove that S_n is given by a presentation listed in Example 3.2.8.
22. Show that A_4 has a presentation $\langle x, y \mid x^2 = y^3 = (xy)^3 = 1 \rangle$.
23. Identify the group $\langle x, y, z \mid z^y = z^2, x^z = x^2, y^x = y^2 \rangle$.
24. Find all the composition series of S_4.

3.3 Finite simple groups

Quote from `Wikipedia`:

> In mathematics, the classification of finite simple groups states that every finite simple group is cyclic, or alternating, or in one of 16 families of groups of Lie type, or one of 26 sporadic groups... These groups can be seen as the basic building blocks of all finite groups, in a way reminiscent of the way the prime numbers are the basic building blocks of the natural numbers. The Jordan-Hoölder theorem is a more precise way of stating this fact about finite groups. However, a significant difference with respect to the case of integer factorization is that such "building blocks" do not necessarily determine uniquely a group, since there might be many non-isomorphic groups with the same composition series or, put in another way, the extension problem does not have a unique solution.
>
> The proof of the theorem consists of tens of thousands of pages in several hundred journal articles written by about 100 authors, published mostly between 1955 and 2004. Gorenstein (d.1992), Lyons, and Solomon are gradually publishing a simplified and revised version of the proof.

3.3.1 Faithful primitive actions and Iwasawa's Lemma

In this section we prove Iwasawa's Lemma which provides a useful criterion for simplicity of a given finite group.

Transitive actions

Let H be a subgroup of G. Denote by $H\backslash G$ the set of right cosets of H in G (note that, unless H is a normal subgroup, $H\backslash G$ is only a set, not a group in general). The group G acts on $H\backslash G$ by right multiplication. This action is obviously transitive. Our first result shows that this example is, in a sense, generic. Before stating this in a precise form, we need a definition. Let G act on sets X_1 and X_2. An *equivalence* between these two actions is a bijection $f : X_1 \to X_2$ such that $(xg)^f = (x^f)g$ for all $x \in X_1$ and $g \in G$.

Proposition 3.3.1 *Any transitive action of a group G on a set X is equivalent to the action of G on $H\backslash G$, where $H = \mathrm{stab}_G(x)$ for some $x \in X$. Furthermore, the actions of G on $H\backslash G$ and $K\backslash G$ are equivalent if and only if H and K are conjugate.*

PROOF. Fix $x \in X$ and denote $H = \mathrm{stab}_G(x)$. Since the action is transitive, is straightforward to show there is an obvious bijection between X and the set of subsets $O(x,y) = \{g \in G \mid xg = y\}$ of G. Note that $O(x,y) = Hg$ for any $g \in O(x,y)$. It is now easy that the map $y \mapsto O(x,y)$ is an equivalence between the action of G on X, and the action of G on $H\backslash G$. The second part is left as an exercise. ∎

Suppose G acts transitively on a set X with $|X| > 1$. A *G-congruence* on X is an equivalence relation $\equiv$ on X that is compatible with the action, i.e., if $x \equiv y$, then $xg \equiv yg$ for all $g \in G$. An equivalence class of a G-congruence is called a *block*. There are two trivial G-congruences on X, namely, the *equality* $x \equiv y \Leftrightarrow x = y$, and the *universal relation* $x \equiv y$ for all $x, y \in X$. The action is called *imprimitive* if there is a non-trivial G-congruence on X, and *primitive* otherwise.

Examples of primitive actions can be obtained as follows. We say that an action of G on X is *doubly transitive* if for any two ordered pairs (x_1, x_2) and (y_1, y_2) of distinct elements of X there exists $g \in G$ such that $x_1 g = y_1$ and $x_2 g = y_2$.

Proposition 3.3.2 *A doubly transitive action is primitive.*

We leave the proof as an exercise. The following result provides a useful characterization of blocks:

Proposition 3.3.3 *Let G act transitively on X and let B be a non-empty subset of X. Then B is a block if and only if, for all $g \in G$, either $Bg = B$ or $Bg \cap B = \emptyset$.*

PROOF. If B is a block then Bg is also a block and the claim follows by the fact that different equivalence classes are disjoint.

Conversely, let B be a non-empty subset of X such that, for all $g \in G$, either $Bg = B$ or $Bg \cap B = \emptyset$. Since the action is transitive, all different Bg form a partition of X, which is the set of equivalence classes of a congruence. ∎

Proposition 3.3.4 *Let H be a proper subgroup of G. Then the action of G on $H\backslash G$ is primitive if and only if H is a maximal subgroup of G.*

PROOF. Suppose that G acts primitively on $H\backslash G$ and assume that $H < K < G$. Let B be the set of all cosets of H which are contained in K. By Proposition 3.3.3, B is a block which neither a singleton nor the whole $H\backslash G$, a contradiction.

Conversely, suppose that G acts imprimitively on $H\backslash G$. Let B be a block containing the coset H, and denote $K = \{g \in G \mid Bg = B\}$. Then $H < K < G$. ∎

Proposition 3.3.5 *Let G act primitively on X, and let N be a normal subgroup of G. Then either N acts trivially on X, or N acts transitively on X.*

PROOF. For $x, y \in X$ put $x \equiv y$ iff $xh = y$ for some $h \in N$. For any $g \in G$ we have $(xg)(g^{-1}hg) = yg$. By normality, $g^{-1}hg \in N$. Therefore $xg \equiv yg$, so $\equiv$ is a G-congruence. By primitivity, either all orbits have size 1 (i.e., N is in the kernel of the action), or there is a single orbit (i.e., N acts transitively on X). ∎

Minimal and maximal subgroups

The above discussion on actions provides some useful descriptions of minimal and maximal subgroups of finite groups.

Lemma 3.3.6 *A minimal normal subgroup of a finite group is isomorphic to the direct product of a number of copies of a simple group.*

PROOF. Let H be a minimal normal subgroup of G. By Lemma 3.2.4, H has no proper nontivial characteristic subgroups. Choose a minimal normal subgroup N of H of smallest possible order. Consider all subgroups of H of the form $N_1 \times \cdots \times N_n$, where $N_i \triangleleft H$, $N_i \cong N$. Let M be such group of largest possible order. If we show that $M = H$, then it follows from here that N is simple. For, if K is a normal subgroup of N, then it is a normal subgroup of $M = N_1 \times \cdots \times N_n = G$, and this contradicts the choice of N.

Thus it suffices to show that M is characteristic in H. Take $\phi \in \operatorname{Aut} H$. Then $N_i^\phi \cong N$. A straightforward argument shows that $N_i^\phi \triangleleft H$. If $N_i^\phi \not\leq M$, then $N_i^\phi \cap M \not\leq N_i^\phi$ and $|N_i^\phi \cap M| < |N|$. But $N_i^\phi \cap M \triangleleft H$, so the minimality of $|N|$ shows $N_i^\phi \cap M = \{1\}$. The subgroup $\langle M, N_i^\phi \rangle = M \times N_i^\phi$ is of the same type like M but of larger order, a contradiction. Thus M is characteristic in H. ∎

Corollary 3.3.7 *Let G be a finite solvable group. Then any maximal subgroup of G has prime power index.*

PROOF. Let H be a maximal subgroup of G and consider the action of G on $H\backslash G$. By Proposition 3.3.4, this action is primitive. The image of this action is a quotient of G, hence it is a solvable group. Therefore we may assume wlog that the action is faithful. Let N be a minimal normal subgroup of G. Then N is an elementary abelian p-group by Lemma 3.3.6. Snce G acts primitively, N acts transitively by Proposition 3.3.5. Using the

Orbit-Stabilizer Theorem, $|H\backslash G|$ is a power of p. ∎

Faithful actions and Iwasawa's Lemma

From here on we consider only faithful actions. We say that such an action of G on X is *regular* if it is transitive and the point stabilizer is trivial. From the above we see that a regular action of G is isomorphic to the action of G on itself by right multiplication.

Let G act faithfully on X and let N be a normal subgroup of G whose action on X is regular. Then we can identify X with N, so that N acts by right multiplication. To be more precise, choose $x \in X$ and observe there is a bijection between N and X under which $n \in N$ corresponds to $xn \in X$. Under the above bijection, the action of $\mathrm{stab}_G(x)$ on N by conjugation corresponds to the given action on X. To see this, take $g \in \mathrm{stab}_G(x)$ and suppose that $yg = z$. Let $h, k \in N$ correspond to $y, z \in X$ under the above bijection, that is, $xh = y$, $xk = z$. Then $x(g^{-1}hg) = xhg = yg = z$. Since the action is faithful, we conclude that $g^{-1}hg = k$, as required.

Theorem 3.3.8 (Iwasawa's Lemma) *Let G be a group with a faithful primitive action on X. Suppose there exists an abelian normal subgroup A of $\mathrm{stab}_G(x)$ with the property that the conjugates of A generate G. Then any non-trivial normal subgroup of G contains G'. In particular, if G is perfect, then it is simple.*

PROOF. Let N be a non-trivial normal subgroup of G. By Proposition 3.3.5, N acts transitively on X, therefore $N \not\leq \mathrm{stab}_G(x)$. By Proposition 3.3.4, $\mathrm{stab}_G(x)$ is a maximal subgroup of G. Hence $N\,\mathrm{stab}_G(x) = G$. Take $g \in G$ and write it as $g = nh$, where $n \in N$ and $h \in \mathrm{stab}_G(x)$. Then $gAg^{-1} = nhAh^{-1}n^{-1} = nAn^{-1}$. We conclude that $gAg^{-1} \leq NA$. By our assumption it follows that $G = NA$. Now, $G/N \cong A/(A \cap N)$ is abelian, hence $G' \leq N$. ∎

3.3.2 Symmetric groups and alternating groups

Here we examine the normal subgroups of S_n and prove that if $n \geq 5$, then the alternating group A_n is simple.

Proposition 3.3.9 *Two elements of S_n are conjugate if and only if they have the same cycle structure.*

PROOF. If $\pi \in S_n$ and $\gamma = (a_1\, a_2 \ldots a_k)$ is a cycle, then $\gamma^\pi = (a_1^\pi\, a_2^\pi \ldots a_k^\pi)$. ∎

Proposition 3.3.10 *The alternating group A_n is generated by the 3-cycles.*

PROOF. Note that 3-cycles are even permutations. If π is any even permutation, then it can be written as a product of an even number of transpositions. Thus we only need to consider products of two transpositions. If $a,b,c,d \in \{1,2,\ldots,n\}$ are pairwise different, then the following clearly hold:

$$\begin{aligned}(a\,b)(a\,b) &= 1,\\ (a\,b)(a\,c) &= (a\,b\,c),\\ (a\,b)(c\,d) &= (a\,b\,c)(a\,d\,c),\end{aligned}$$

and we are done. ■

Proposition 3.3.11 *The following are equivalent for $\pi \in A_n$:*

1. *The S_n conjugacy class of π splits into two A_n-conjugacy classes;*
2. *There is no odd permutation which commutes with π;*
3. *π has no cycles of even length, and all of its cycless have distinct lengths.*

PROOF. Let us proove that (1) is equivalent to (2). The group S_n acts transitively on A_n by conjugation. We have that $C_{A_n}(\pi) = C_{S_n}(\pi) \cap A_n$. If (2) holds, then $C_{A_n}(\pi) = C_{S_n}(\pi)$, therefore π has $|A_n : C_{A_n}(\pi)| = |S_n : C_{S_n}|/2$ conjugates in A_n. Thus (1) follows. If (2) does not hold then $|C_{A_n}(\pi)| = |C_{S_n}(\pi)|/2$, and π has $|A_n : C_{A_n}(\pi)| = |S_n : C_{S_n}|$ conjugates in A_n. Therefore (1) does not hold.

Now we prove that (2) and (3) are equivalent. If π has a cycle of even length, then this cycle is an odd permutation commuting with π. If π has only cycles of odd length, and two cycles of the same length ℓ, then a permutation interchanging them is a product of ℓ transpositions commuting with π. This proves that (2) implies (3). Assume now that (3) holds. Then any permutation commuting with π fixes each of its cycles and acts on it as a power of the corresponding cycle of π, hence it is an even permutation. ■

Proposition 3.3.12 *The group A_5 is simple.*

PROOF. A lazy proof is

```
gap> IsSimple( AlternatingGroup( 5 ) );
true
```

A formal proof goes as follows. The conjugacy classes of A_5 can be determined using Proposition 3.3.11:

- Representative $(*)(*)(*)(*)(*)$: this class has size 1 and does not split into two conjugacy classes of A_5;

- Representative $(*)(**)(**)$: this class has size 15 and does not split into two conjugacy classes of A_5;
- Representative $(*)(*)(***)$: this class has size 20 and does not split into two conjugacy classes of A_5;
- Representative $(*****)$: this class has size 24 and splits into two conjugacy classes of A_5, each of size 12.

A normal subgroup N of A_5 would have to be a union of conjugacy classes and contain the identity, plus its order would have to divide 60. Checking all the possibilities, we see that either N is trivial or $N = A_5$. ∎

It turns out that A_5 is the only simple group of order 60. A formal proof can be found in [4]. Here is a proof using GAP:

```
gap> Filtered(AllSmallGroups(60), IsSimple);
[ Alt( [ 1 .. 5 ] ) ]
```

Theorem 3.3.13 *If $n \geq 5$, then A_n is simple.*

PROOF. The proof goes by induction on n. The case $n = 5$ is covered by Proposition 3.3.12. Suppose N is a non-trivial normal subgroup of A_n. Since A_n clearly acts doubly transitively on $X = \{1, 2, \ldots, n\}$, this action is primitive by 3.3.2. Therefore N acts transitively on X by 3.3.5. It follows by Frattini's argument that $NA_{n-1} = A_n$. The intersection $N \cap A_{n-1}$ is a normal subgroup of A_{n-1}. By assumption, either $N \cap A_{n-1} = \{1\}$ or $A_{n-1} \leq N$. In the latter case, $A_n/N = NA_{n-1}/N \cong A_{n-1}/(A_{n-1} \cap N) = \{1\}$, hence $N = A_n$. So assume that $N \cap A_{n-1} = \{1\}$. In this case N acts regularly and so $|N| = n$ by a discussion above. By Lemma 3.2.30, N can be generated by at most $\lfloor \log_2 n \rfloor$ elements. An automorphism of N is determined by the images of generators, hence $|\mathrm{Aut}(N)| \leq n^{\log_2 n}$. On the other hand, A_{n-1} acts faithfully on N by conjugation, so $(n-1)! \leq n^{\log_2 n}$ which is impossible for $n \geq 6$. ∎

Corollary 3.3.14 *Let $n \geq 5$. Then the only normal subgroups of S_n are $\{1\}$, A_n and S_n.*

PROOF. Let N be a normal subgroup of S_n. Then $N \cap A_n$ is a normal subgroup of A_n, hence either $A_n \cap N = \{1\}$ or $A_n \leq N$. Suppose the first possibility holds. Then $N = N/(N \cap A_n) \cong NA_n/A_n$. If N is non-trivial then $NA_n = S_n$ and hence $N \cong C_2$. This is impossible as there would have to be a non-identity element of A_n in a conjugacy class of size 1. The remaining possibility is $A_n \leq N$, but in this case we either have $N = A_n$ or $N = S_n$, as A_n is a maximal subgroup of S_n. ∎

The remaining cases of S_n and A_n for $1 \leq n \leq 4$ are somewhat exceptional, but easy to deal with. We show here how to use GAP to examine these groups:

```
gap> for n in [ 1..4 ] do
> sn := SymmetricGroup( n );
> an := AlternatingGroup( n );
> Print("n = ", n, "\n");
> Print("A_n: ", StructureDescription( an ), " ", IsSimple( an ), "\n" );
> Print("S_n: ", StructureDescription( sn ), " ", NormalSubgroups( sn ), "\n" );
> od;
n = 1
A_n: 1 false
S_n: 1 [ Group( () ) ]
n = 2
A_n: 1 false
S_n: C2 [ SymmetricGroup( [ 1 .. 2 ] ), Group( () ) ]
n = 3
A_n: C3 true
S_n: S3 [ SymmetricGroup( [ 1 .. 3 ] ), Group( [ (1,2,3) ] ), Group( () ) ]
n = 4
A_n: A4 false
S_n: S4 [ SymmetricGroup( [ 1 .. 4 ] ),
  Group( [ (2,4,3), (1,4)(2,3), (1,3)(2,4) ] ),
  Group( [ (1,4)(2,3), (1,3)(2,4) ] ), Group( () ) ]
```

3.3.3 Simplicity of projective special linear groups

Unless stated otherwise, F will denote the Galois field GF(q), where q is a prime power. The *projective space* $\mathbb{P}^{n-1}(F)$ is the set of all one-dimensional subspaces of F^n. There are q^n-1 non-zero vectors in F^n, each of which spans a one-dimensional subspace. Each such space is spanned by any of its $q-1$ non-zero vectors, hence $|\mathbb{P}^{n-1}(F)| = (q^n-1)/(q-1)$. The group $\mathrm{GL}(n,F)$ acts on $\mathbb{P}^{n-1}(F)$ from the left as follows: $(A, \mathrm{span}(v)) \mapsto \mathrm{span}(Av)$.

Proposition 3.3.15 *The following conditions for $A \in \mathrm{GL}(n,F)$ are equivalent:*

1. *$A \in Z(\mathrm{GL}(n,F))$;*
2. *A is in the kernel of the action of $\mathrm{GL}(n,F)$ on $\mathbb{P}^{n-1}(F)$;*
3. *A is a scalar matrix, i.e., $A = \lambda I$ for some $\lambda \in F^\times$.*

PROOF. Clearly (3) implies (1). To see that the converse holds, take $A \in Z(\mathrm{GL}_n(F))$. Then, in particular, A has to commute with all matrices with 1 on the diagonal and the position (i,j), $i \neq j$, and zero elsewhere. Easy calculation then shows that A is a scalar matrix.

Let us prove that (2) and (3) are equivalent. Clearly every scalar matrix fixes all 1-dimensional subspaces of F^n. Conversely suppose that A fixes all 1-dimensional subspaces. Let $e_1, \ldots, e_n$ be a standard basis of F^n. Then $Ae_i = \lambda_i e_i$ for some non-zero $\lambda_i \in F$. Fix different i and j. There also exists $\lambda \in F^\times$ such that $A(e_i+e_j) = \lambda(e_i+e_j)$, and this implies $\lambda = \lambda_j = \lambda_i$. Consequently, A is a scalar matrix. ∎

We define the *projective general and projective special linear groups* by

$$\mathrm{PGL}(n,F) = \mathrm{GL}(n,F)/Z(\mathrm{GL}(n,F))$$

and

$$\mathrm{PSL}(n,F) = \mathrm{SL}(n,F)Z(\mathrm{GL}(n,F))/Z(\mathrm{GL}(n,F)).$$

Therefore the projective groups are the images of ther linear group counterparts in the action on the projective space, so we can think of them as subgroups of $\mathrm{Sym}\,\mathbb{P}^{n-1}(F)$. We see that $|\mathrm{PGL}(n,q)| = |\mathrm{GL}(n,q)|/(q-1) = |\mathrm{SL}(n,q)|$.

Proposition 3.3.16 $|\mathrm{PSL}(n,q)| = |\mathrm{SL}(n,q)|/\gcd(n,q-1)$.

PROOF. The kernel of the action of $\mathrm{SL}(n,q)$ on the corresponding projective space consists of scalar matrices with determinant one, i.e., matrices of the form λI with $\lambda^n = 1$. The multiplicative group of $\mathrm{GF}(q)$ is cyclic of order $q-1$, so the number of solution of $\lambda^n = 1$ is $\gcd(n,q-1)$. ∎

If we restrict to the case $n = 2$, we see that $\mathbb{P}^1(F)$ has $q+1$ points, so $\mathrm{PGL}(2,q)$ and $\mathrm{PSL}(2,q)$ are subgroups of S_{q+1}. Let us consider some small cases:

$q = 2$: $\mathrm{PSL}(2,2) = \mathrm{PGL}(2,2)$ is a subgroup of S_3 of order 6, hence $PSL(2,2) \cong S_3$.

$q = 3$: $\mathrm{PGL}(2,3)$ is a subgroup of S_4 of order 24, hence $\mathrm{PGL}(2,3) = S_4$. The group $\mathrm{PSL}(2,3)$ is a subgroup of index 2 in $\mathrm{PGL}(2,3)$, hence $\mathrm{PSL}(2,3) \cong A_4$.

$q = 4$: $\mathrm{PGL}(2,4) = \mathrm{PSL}(2,4)$ is a subgroup of S_5 of order 60, so it is isomorphic to A_5; one can double-check this with GAP:

```
gap> StructureDescription(PSL(2,4));
"A5"
```

$q = 5$: $\mathrm{PSL}(2,5) \cong A_5$:

```
gap> StructureDescription(PSL(2,5));
"A5"
```

We also remark here that there is another way of interpreting the actions of $\mathrm{PGL}(2,F)$ and $\mathrm{PSL}(2,F)$ on the projective line. The one-dimensional subspaces of F^2 can be spanned by either a unique vector of the form $(1,x)$, where $x \in F$, or the vector $(0,1)$. We identify points of the first type with F, and the point of the second type with ∞. Then the elements of $\mathrm{PGL}(2,F)$ can be identified with *linear fractional maps*

$$z \mapsto \frac{az+b}{cz+d},$$

where $a,b,c,d \in F$, $ad-bc \neq 0$. The group $\mathrm{PSL}(2,F)$ then consists of those linear fractional maps with $ad-bc = 1$.

We will prove the following result:

Theorem 3.3.17 *For $n \geq 2$ and any field F, the group* $\mathrm{PSL}(n,F)$ *is simple, except in the two cases, $n=2$, $F=GF(2)$ or $n=2$, $F=GF(3)$.*

We will only prove this theorem for $n=2$, the proof for $n>2$ is similar, but somewhat technical. Our proof will rely on Iwasawa's lemma applied to the action of $G=\mathrm{SL}(2,F)$ on $\mathbb{P}^1(F)$. We will show in a series of steps that all the conditions of the lemma are satisfied.

Proposition 3.3.18 *If $n \geq 2$, then* $\mathrm{SL}(2,F)$ *acts doubly transitively on* $\mathbb{P}^1(F)$.

Proof. Let $\mathrm{span}(v_1)$ and $\mathrm{span}(v_2)$ be two distinct 1-dimensional subspaces of F^2. For any other pair $\mathrm{span}(w_1)$ and $\mathrm{span}(w_2)$ there exists a linear map that maps $v_i \mapsto w_i$, $i=1,2$. One can modify this map to obtain one with determinant 1. We let the reader fill in the details. ∎

Let e_1, e_2 be a standard basis of F^2. Denote $x=\mathrm{span}(e_1)$. The stabilizer of x is

$$\begin{aligned}\mathrm{stab}_G(x) &= \{A \in \mathrm{SL}(2,F) \mid \mathrm{span}(e_1) = \mathrm{span}(Ae_1)\} \\ &= \left\{ \begin{pmatrix} a & b \\ 0 & 1/a \end{pmatrix} \mid a \in F^\times, b \in F \right\}.\end{aligned}$$

There is an abelian normal subgroup of $\mathrm{stab}_G(x)$ given as follows:

$$U = \left\{ \begin{pmatrix} 1 & b \\ 0 & 1 \end{pmatrix} \mid b \in F \right\}.$$

Its elements are called *transvections*.

Proposition 3.3.19 *The subgroup U and its conjugates generate* $\mathrm{SL}_2(F)$.

Proof. First we note that

$$\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}^{-1} U \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} = \left\{ \begin{pmatrix} 1 & 0 \\ b & 0 \end{pmatrix} \mid b \in F \right\} = U'.$$

Now pick $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(F)$. Suppose first that $b \neq 0$. Then

$$A = \begin{pmatrix} 1 & 0 \\ (d-1)/b & 1 \end{pmatrix} \begin{pmatrix} 1 & b \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ (a-1)/b & 1 \end{pmatrix} \in \langle U, U' \rangle.$$

If $c \neq 0$, then

$$A = \begin{pmatrix} 1 & (a-1)/c \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ c & 1 \end{pmatrix} \begin{pmatrix} 1 & (d-1)/c \\ 0 & 1 \end{pmatrix} \in \langle U, U' \rangle.$$

Finally assume that $b=c=0$. Then

$$A = \begin{pmatrix} 1 & 0 \\ (1-a)/a & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ (a-1) & 1 \end{pmatrix} \begin{pmatrix} 1 & -1/a \\ 0 & 1 \end{pmatrix} \in \langle U, U' \rangle.$$

This proves the result. ∎

Proposition 3.3.20 *If* $|F|>3$*, then* $\mathrm{SL}(2,F)$ *is a perfect group.*

PROOF. If $|F|>3$ there exists $a\in F$ such that $a^2\notin\{0,1\}$. Now we observe

$$\begin{pmatrix}1 & b(a^2-1)\\ 0 & 1\end{pmatrix}=\left[\begin{pmatrix}1/a & 0\\ 0 & a\end{pmatrix},\begin{pmatrix}1 & 0\\ -b & 1\end{pmatrix}\right].$$

Letting b run through F, we see that $U\leq \mathrm{SL}_2(F)'$. By Proposition 3.3.19 we conclude the result. ∎

PROOF.[Proof of Theorem 3.3.17 for $n=2$] This follows by previous propositions and Iwasawa's lemma. ∎

3.3.4 On the classification of finite simple groups (CFSG)

One of the greatest achievements of mathematics is a full classificiation of finite simple groups (CFSG) which was announced in the 1980's. Roughly speaking, the result says that all finite simple groups fall into one of the following four types:

1. *Cyclic groups of prime order*;
2. *Alternating groups* A_n for $n\geq 5$;
3. *Groups of Lie type*; these groups arise as automorphism groups of simple Lie algebras. An example is $\mathrm{PSL}(n,F)$.
4. *26 sporadic groups*; these do not fall into any infinite family of simple groups described above. They are usually defined as symmetry groups of various algebraic or combinatorial configurations. The largest of them has order

   $$808,017,424,794,512,875,886,459,904,961,710,757,005,754,368,000,000,000$$

   and is called the *Monster Group*.

Since a thorough account on these groups is beyond the purpose of these notes, we only exhibit some of their properties and how to use GAP to study them. The following are all non-abelian finite simple groups of order ≤ 1000000:

```
gap> AllSmallNonabelianSimpleGroups( [1..1000000] );
[ A5, PSL(2,7), A6, PSL(2,8), PSL(2,11), PSL(2,13), PSL(2,17), A7, PSL(2,19),
  PSL(2,16), PSL(3,3), PSU(3,3), PSL(2,23), PSL(2,25), M11, PSL(2,27),
  PSL(2,29), PSL(2,31), A8, PSL(3,4), PSL(2,37), PSp(4,3), Sz(8), PSL(2,32),
  PSL(2,41), PSL(2,43), PSL(2,47), PSL(2,49), PSU(3,4), PSL(2,53), M12,
  PSL(2,59), PSL(2,61), PSU(3,5), PSL(2,67), J_1, PSL(2,71), A9, PSL(2,73),
  PSL(2,79), PSL(2,64), PSL(2,81), PSL(2,83), PSL(2,89), PSL(3,5), M22,
  PSL(2,97), PSL(2,101), PSL(2,103), J_2, PSL(2,107), PSL(2,109), PSL(2,113),
  PSL(2,121), PSL(2,125), PSp(4,4) ]
```

Here is a construction of Mathieu groups M_{11} and M_{12} which are sporadic groups:

```
gap> p1 := (4,5,6)*(7,8,9)*(10,11,12);;
gap> p2 := (4,7,10)*(5,8,11)*(6,9,12);;
gap> p3 := (5,7,6,10)*(8,9,12,11);;
gap> p4 := (5,8,6,12)*(7,11,10,9);;
gap> p5 := (1,4)*(7,8)*(9,11)*(10,12);;
gap> p6 := (1,2)*(7,10)*(8,11)*(9,12);;
gap> p7 := (2,3)*(7,12)*(8,10)*(9,11);;
gap> m11 := Group(p1, p2, p3, p4, p5, p6);;
gap> IsSimple(m11);
true
gap> StructureDescription(m11);
"M11"
gap> m12 := Group(p1, p2, p3, p4, p5, p6,p7);;
gap> IsSimple(m12);
true
gap> StructureDescription(m12);
"M12"
```

There is a vast amount of properties of finite simple groups that follow from CFSG, too many to state here. Some of them are:

Theorem 3.3.21 *Let S be a finite non-abelian simple group.*

1. *S can be generated by two elements.*
2. $\mathrm{Out}(S)$ *is a solvable group (used to be Schreier's conjecture).*
3. *Every element of S is a commutator (used to be Ore's conjecture).*

CFSG also implies, that, given a positive integer n, there are at most two non-isomorphic finite simple groups of order n. It may happen that there are two non-isomorphic finite simple groups of the same order. For example, consider $\mathrm{PSL}(3,4)$ and $\mathrm{PSL}(4,2)$; they are both of order 20160, and

```
gap> G:=PSL(4,2);;
gap> H:=PSL(3,4);;
gap> IsomorphismGroups(G,H);
fail
```

Apart from using GAP, several useful information on finite simple groups can be obtained from *Atlas of Finite Group Representations* [1].

3.3.5 Problems

1. Complete the proof of Proposition 3.3.1.
2. Prove Proposition 3.3.2.
3. Let G act transitively on X. Suppose that the stabilizer of $x \in X$ acts transitively on $X - \{x\}$. Then G acts doubly transitively on X.

4. Let Ω be the set of 2-element subsets of $\{1,2,\ldots,n\}$. Then S_n acts on Ω by $\{i,j\}g = \{ig, jg\}$.
   (a) If $n=2$, then the action is not faithful.
   (b) If $n=3$, then the action is doubly transitive.
   (c) If $n=4$, then the action is imprimitive.
   (d) If $n \geq 5$, then the action is primitive, but not doubly transitive.
5. Let G be a group. The group $\mathrm{Aut}\,G$ acts naturally on the set G.
   (a) If $G-\{1\}$ is an orbit, prove that G is an elementary abelian p-group.
   (b) If $\mathrm{Aut}\,G$ acts doubly transitively on $G-\{1\}$, show that either G is a 2-group or $|G|=3$.
6. Let G be a group of order $2m$, where m is odd and $m > 1$. Prove that G is not simple.
7. Let $n \geq 2$. Show that the transpositions (1 2), (1 3), ..., (1 n) generate S_n.
8. Let $n \geq 3$. Show that the 3-cycles (1 2 3), (1 2 4), ..., (1 2 n) generate A_n.
9. Prove that there are no simple groups of order 312, 616, or 1960.
10. Show that the only simple group of order 60 is A_5.
11. Prove that $\mathrm{PSL}(4,2) \cong A_8$.
12. Prove by hand that $\mathrm{PSL}(3,4)$ has no elements of order 15, so it is not isomorphic to A_8.
13. Show that transvections in $\mathrm{SL}(2,F)$ need not be conjugate.

3.4 Some extension theory

Let N be a normal subgroup of G. Then we say that G is an extension of N by G/N. A precise definition of group extensions will be given in Section 3.4.1. The importance of extension theory can be outlined as follows. Let G be a finite group and $1 = G_0 \triangleleft G_1 \triangleleft G_2 \triangleleft \cdots \triangleleft G_r = G$ its composition series. By Jordan-Hölder theorem, the composition factors G_{i+1}/G_i are in a sense uniquely determined by G. On the other hand, these are simple groups, so they are known by CFSG. In order to build all finite groups with a given sequennce of composition factors, one can proceed as follows. Suppose we already know what G_i is, and we have a prescribed isomorphism type of the simple group G_{i+1}/G_i. If we knew how to build all the extensions (up to certain equivalence) of a given group by a (simple) group, then we would be able to construct all possible G_{i+1}. Proceeding this way, we would eventually be able to construct all finite groups. The trouble is that the problem of constructing all possible extensions is very difficult and still open.

We will briefly tackle the problem of classifying extensions of abelian groups. It will be shown that these are, up to equivalence, in 1-1 correspondence with the elements of a certain second cohomology group. Cohomological group theory is an area on its own, and we will not go deeply into it. We refer to [3] and [8] for further details.

3.4.1 Basic notions

A *group extension* of a group N by a group G is a short exact sequence

$$1 \longrightarrow N \xrightarrow{\mu} E \xrightarrow{\epsilon} G \longrightarrow 1.$$

From the above it clearly follows that μ is injective, ϵ is surjective, $M = \operatorname{im}\mu = \ker\epsilon$ is a normal subgroup of E, $M \cong N$, and $E/M \cong G$.

A *morphism* between extensions $N \overset{\mu}{\rightarrowtail} E \overset{\epsilon}{\twoheadrightarrow} G \longrightarrow 1$ and $\overline{N} \overset{\overline{\mu}}{\rightarrowtail} \overline{E} \overset{\overline{\epsilon}}{\twoheadrightarrow} \overline{G}$ is a triple of group homomorphisms (α, β, γ) such that the following diagram commutes:

$$\begin{array}{ccccc} N & \overset{\mu}{\rightarrowtail} & E & \overset{\epsilon}{\twoheadrightarrow} & G \\ \downarrow{\scriptstyle\alpha} & & \downarrow{\scriptstyle\beta} & & \downarrow{\scriptstyle\gamma} \\ \overline{N} & \overset{\overline{\mu}}{\rightarrowtail} & \overline{E} & \overset{\overline{\epsilon}}{\twoheadrightarrow} & \overline{G} \end{array}.$$

The collection of all group extensions and morphisms between them is a category. A morphism of the type

$$\begin{array}{ccccc} N & \overset{\mu}{\rightarrowtail} & E & \overset{\epsilon}{\twoheadrightarrow} & G \\ \downarrow{\scriptstyle 1} & & \downarrow{\scriptstyle\beta} & & \downarrow{\scriptstyle 1} \\ N & \overset{\overline{\mu}}{\rightarrowtail} & \overline{E} & \overset{\overline{\epsilon}}{\twoheadrightarrow} & G \end{array}$$

is said to be an *equivalence* of extensions.

3.4.2 Semidirect products

Suppose that H and N are groups and that we have a homomorphism $\alpha\colon H \to \mathrm{Aut}(N)$. The *(external) semidirect product* $H \ltimes_\alpha N$ of N and H is the set of all pairs (h,n), where $h \in H$, $n \in N$, with the operation

$$(h_1, n_1)(h_2, n_2) = (h_1 h_2, n_1^{h_2^\alpha} n_2).$$

This is a group with the identity element $(1_H, 1_N)$, and the inverse of (h,n) is $(h^{-1}, n^{-(h^\alpha)^{-1}})$.

We have embeddings $H \to H \ltimes_\alpha N$ and $N \to H \ltimes_\alpha N$ given by $h \mapsto (h, 1_N)$ and $n \mapsto (1_H, n)$, respectively. If H^* and N^* are images of these maps, then $N^* \triangleleft H \ltimes_\alpha N$, $H^* \cap N^* = 1$ and $H^* N^* = H \ltimes_\alpha N$. We say that $H \ltimes_\alpha N$ is the *internal semidirect product* of N^* and H^*.

The group H^* is said to be a *complement* of N^* in G. The group G is an extension of N^* by H^*; we say that this extension is a *split extension.*

GAP offers two ways of constructing semidirect products. The first one is directly via command `SemidirectProduct(H, alpha, N)`. In the special case when $N = \mathrm{GF}(q)^n$, `alpha` must be a homomorphism from H into a matrix group of $n \times n$ matrices over a subfield of GF(q), or into a permutation group. The second option is to use

`SemidirectProduct(H,N)`,

where $H \leq \mathrm{Aut}(N)$.

Let us build all possible semidirect products of $C_2 \times C_2$ by C_4:

```
gap> H := CyclicGroup(4);;
gap> N := AbelianGroup([2,2]);;
<pc group of size 4 with 2 generators>
gap> hom := AllHomomorphisms(H, AutomorphismGroup(N));;
gap> for map in hom do
> Print(IdGroup(SemidirectProduct(H, map, N)),"\n");
> od;
[ 16, 10 ]
[ 16, 3 ]
[ 16, 3 ]
[ 16, 3 ]
gap> StructureDescription(SmallGroup(16,10));
"C4 x C2 x C2"
gap> StructureDescription(SmallGroup(16,3));
"(C4 x C2) : C2"
```

Here are two more examples:

```
gap> SemidirectProduct(Group((1,2,3),(2,3,4)),GF(5)^4);
<matrix group of size 7500 with 3 generators>
gap> g:=Group((3,4,5),(1,2,3));;
gap> mats:=[[[Z(2^2),0*Z(2)],[0*Z(2),Z(2^2)^2]],
> [[Z(2)^0,Z(2)^0], [Z(2)^0,0*Z(2)]]];;
gap> hom:=GroupHomomorphismByImages(g,Group(mats),[g.1,g.2],mats);;
gap> SemidirectProduct(g,hom,GF(4)^2);
<matrix group of size 960 with 3 generators>
```

An important example of a semidirect product is the following. Let N be any group and $H = \mathrm{Aut}(N)$. Let $\alpha\colon H \to \mathrm{Aut}(N)$ be the identity mapping. Then the semidirect product $\mathrm{Aut}(N) \ltimes_\alpha N$ is called the *holomorph* of N.

Example 3.4.1 *Let $N = C_p^n$ be an elementary abelian p-group of order p^n. Its automorphism group is* $\mathrm{GL}(n,p)$. *The holomorph* $\mathrm{AGL}(n,p) = \mathrm{GL}(n,p) \ltimes C_p^n$ *is called the affine group of dimension n over $\mathbb{Z}_p$. Show that* $\mathrm{AGL}(2,2) \cong S_4$. *Here is a proof using* GAP*:*

```
gap> G := AbelianGroup([2,2]);;
gap> agl := SemidirectProduct(AutomorphismGroup(G), G);;
gap> StructureDescription(agl);
"S4"
```

Another construction related to semidirect products is that of a *wreath product*. Let G and H be groups and let H act on the set $X=\{x_1,x_2,\ldots,x_n\}$. We take

$$G^X=\prod_{i=1}^{n}G_{x_i}$$

to be the direct product of n copies of G indexed by the set X. Then H also acts on G^X by the rule

$$(g_{x_1},g_{x_2},\ldots,g_{x_n})h=(g_{x_1h},g_{x_2h},\ldots,g_{x_nh}).$$

Therefore we have a homomorphism $\alpha\colon H\to \mathrm{Aut}(G^X)$ and we can form the semidirect product $H\ltimes_\alpha G^X$ which is denoted by $G\wr_X H$ and called the *wreath product* of G by H.

A special case is when $X=H$, and H acts on X by right multiplication. Then the corresponding wreath product is denoted by $G\wr H$ and called the *regular (standard) wreath product*. Here is an example of how to build $C_2\wr C_4$ with GAP:

```
gap> G := StandardWreathProduct(CyclicGroup(2), CyclicGroup(4));
<group of size 64 with 3 generators>
gap> IdGroup(G);
[ 64, 32 ]
```

Alternatively, we can build $C_2\wr C_4$ as a semidirect product $C_4\ltimes C_2^4$, where we think of C_4 as the group $\langle(1\,2\,3\,4)\rangle$ acting on C_2^4 by permuting the indices:

```
gap> G := SemidirectProduct(Group((1,2,3,4)), GF(2)^4);
<matrix group of size 64 with 2 generators>
gap> IdGroup(G);
[ 64, 32 ]
```

Wreath products are important in the theory of extensions because of the following:

Theorem 3.4.2 *Every extension of G by H is isomorphic to a subgroup of $G\wr H$.*

We leave the proof as an exercise.

3.4.3 Extensions with abelian kernels

Consider

$$A\overset{\mu}{\rightarrowtail} E\overset{\epsilon}{\twoheadrightarrow} G\ ,$$

where A is an abelian group (written additively). When choosing a transversal $\mathscr{T}$ to $M=\mathrm{im}\,\mu=\ker\epsilon$ in E, we get a function $\tau\colon G\to E$ defined by $g^\tau=x$, where $x\in\mathscr{T}$ is such that $g=x^\epsilon$ (note that this is well defined). The function τ is called a *transversal function*. Note that τ is not necessarily a homomorphism. We also see that $\tau\epsilon=1_G$, and that any function $\tau\colon G\to E$ with the property $\tau\epsilon=1_G$ determines a transversal to M in E, namely $\{g^\tau\mid g\in G\}$.

Suppose that we have fixed τ. Then the elements $\{g^\tau : g \in G\}$ act on M by conjugation. Since $\mu: A \to M$ is an isomorphism, we can define $g^\chi \in \mathrm{Aut}(A)$ by the rule

$$(a^{g^\chi})^\mu = (g^\tau)^{-1} a^\mu (g^\tau)$$

for $a \in A$ and $g \in G$. We obtain a function $\chi: G \to \mathrm{Aut}(A)$. We prove that χ does not depend on the choice of τ. Here we will use the fact that A is abelian. Suppose that τ' is another transversal function. Then $(g^\tau (g^{\tau'})^{-1})^\epsilon = g^{\tau\epsilon}(g^{\tau'\epsilon})^{-1} = 1$, hence $g^{\tau'} = g^\tau m_g$ for some $m_g \in M$. If τ' induces $\chi': G \to \mathrm{Aut}(A)$ as above, then

$$(a^{g^{\chi'}})^\mu = (g^{\tau'})^{-1} a^\mu (g^{\tau'}) = m_g^{-1}((g^\tau)^{-1} a^\mu (g^\tau)) m_g,$$

hence $g^\chi = g^{\chi'}$. Thus χ is uniquely defined. We claim that χ is a homomorphism. Let $g_1, g_2 \in G$. Then $(g_1 g_2)^\tau \equiv g_1^\tau g_2^\tau \mod M$. Thus $(g_1 g_2)^\chi = g_1^\chi g_2^\chi$, hence χ is a homomorphism. We have proved:

Proposition 3.4.3 *Each extension* $A \overset{\mu}{\rightarrowtail} E \overset{\epsilon}{\twoheadrightarrow} G$ *, where A is abelian, determines a unique homomorphism $\chi : G \to \mathrm{Aut}(A)$ which arises by conjugation in* $\mathrm{im}\,\mu$ *by elements of E.*

Let $\chi: G \to \mathrm{Aut}(A)$ be a homomorphism. Then χ induces a G-action A given by $a \cdot g = a^{g^\chi}$. We say that A is a *G-module*. More precisely, let $g \in G$ and $x \in E$ such that $x^\epsilon = g$. Then

$$(a\,g)^\mu = x^{-1} a^\mu x$$

for $a \in A$ (well defined, since A is abelian). Note that this action is trivial precisely when $\mathrm{im}\,\mu$ is central in E, i.e., when the corresponding extension is a *central extension.*

Theorem 3.4.4 *Equivalent extensions of A by G, where A is abelian, induce the same G-module structure on A.*

Proof. Suppose we have equivalent extensions

$$\begin{array}{ccccc} A & \overset{\mu}{\rightarrowtail} & E & \overset{\epsilon}{\twoheadrightarrow} & G \\ \downarrow 1 & & \downarrow \beta & & \downarrow 1 \\ A & \overset{\overline{\mu}}{\rightarrowtail} & \overline{E} & \overset{\overline{\epsilon}}{\twoheadrightarrow} & G \end{array}$$

Let χ and $\overline{\chi}$ be the respective homomorphisms $G \to \mathrm{Aut}(A)$. Choose a transversal function $\tau: G \to E$. Let $\overline{\tau} = \tau\beta$. Then $\overline{\tau\epsilon} = \tau\beta\overline{\epsilon} = \tau\epsilon = 1_G$, hence $\overline{\tau}$ is a transversal function for the second extension. Then $(a^{g^\chi})^\mu = (g^\tau)^{-1} a^\mu (g^\tau)$ and $(a^{g^{\overline{\chi}}})^{\overline{\mu}} = (g^{\overline{\tau}})^{-1} a^{\overline{\mu}} (g^{\overline{\tau}})$ for $a \in A$ and $g \in G$. Applying β to the first equation and using the fact that $\mu\beta = \overline{\mu}$, we get $(a^{g^\chi})^{\overline{\mu}} = (g^{\tau\beta})^{-1} a^{\mu\beta} (g^{\tau\beta}) = (a^{g^{\overline{\chi}}})^{\overline{\mu}}$ and thus $g^\chi = g^{\overline{\chi}}$. ∎

Choose a transversal function $\tau: G \to E$, i.e., $\tau\epsilon = 1_G$. Then the above action can be rewritten as

$$(a\,g)^\mu = g^{-\tau} a^\mu g^\tau.$$

Let $x, y \in G$. As $x^\tau y^\tau$ and $(xy)^\tau$ belong to the same coset of $\ker\epsilon = \operatorname{im}\mu$ in E, we may write

$$x^\tau y^\tau = (xy)^\tau((x,y)\phi)^\mu$$

for some $(x,y)\phi \in A$. Thus we get a function $\phi: G \times G \to A$ defined by

$$((x,y)\phi)^\mu = (xy)^{-\tau} x^\tau y^\tau.$$

From the associative law $x^\tau(y^\tau z^\tau) = (x^\tau y^\tau)z^\tau$ we get that ϕ satisfies the identity

$$(x, yz)\phi + (y,z)\phi = (xy,z)\phi + (x,y)\phi \cdot z.$$

A function $\phi: G \times G \to A$ satisfiying this functional equation is called a *factor set* (or a *2-cocycle*). Note that we can assume without loss of generality that $1^\tau = 1$, therefore we can always assume that $(1,x)\phi = (x,1)\phi = 0$ for all $x \in G$. The set $Z^2(G,A)$ of all 2-cocycles in G with coefficients in the G-module A has the structure of an abelian group with the operation

$$(x,y)(\phi_1 + \phi_2) = (x,y)\phi_1 + (x,y)\phi_2.$$

Example 3.4.5 *In the situation above, what happens if $(x,y)\phi = 0$ for all $x, y \in G$? In this case, the transversal map $\tau: G \to E$ is a homomorphism. It is easy to see that the image of τ is then a complement of $\operatorname{im}\mu \cong A$ in E, therefore $E \cong G \ltimes_\chi A$.*

How does the choice of τ affect ϕ? Let τ' be another transversal function for given extension. Then we get another factor set ϕ', i.e., $x^{\tau'} y^{\tau'} = (xy)^{\tau'}((x,y)\phi')^\mu$. As x^τ and $x^{\tau'}$ belong to the same coset of $\ker\epsilon = \operatorname{im}\mu$, we can write

$$x^{\tau'} = x^\tau((x)\psi)^\mu$$

for some $(x)\psi \in A$. We get

$$(x,y)\phi = (x,y)\phi' + (xy)\psi - (x)\psi \cdot y - (y)\psi.$$

Define $\psi^*: G \times G \to A$ by

$$(x,y)\psi^* = (y)\psi - (xy)\psi + (x)\psi \cdot y,$$

so that $\phi' = \phi + \psi^*$. It follows that $\psi^* \in Z^2(G,A)$. The 2-cocycle ψ^* is called a *2-coboundary*. 2-coboundaries form a subgroup $B^2(G,A)$ of $Z^2(G,A)$. We have proved:

Proposition 3.4.6 *The extension $A \overset{\mu}{\rightarrowtail} E \overset{\epsilon}{\twoheadrightarrow} G$, where A is abelian, determines a unique element $\phi + B^2(G,A)$ of the group $Z^2(G,A)/B^2(G,A)$.*

Does every factor set induce an extension? Let A be a G-module and $\phi: G\times G\to A$ a factor set. Let $E(\phi)$ be (as a set) $G\times A$, with the operation

$$(x,a)(y,b)=(xy,ay+b+(x,y)\phi).$$

$E(\phi)$ becomes a group with identity element $(1,-(1,1)\phi)$ and inversion rule $(x,a)^{-1}=(x^{-1},-ax^{-1}-(1,1)\phi-(x,x^{-1})\phi)$. Define $\mu: A\to E(\phi)$ by the rule $a^{\mu}=(1,a-(1,1)\phi)$, and $\epsilon: E(\phi)\to G$ by the rule $(x,a)^{\epsilon}=x$. Then we have

$$A\rightarrowtail^{\mu} E(\phi)\overset{\epsilon}{\twoheadrightarrow} G\ .$$

Proposition 3.4.7 *Let A be a G-module and $\phi: G\times G\to A$ a factor set. Then the extension*

$$A\rightarrowtail^{\mu} E(\phi)\overset{\epsilon}{\twoheadrightarrow} G$$

induces the given G-module structure. There exists a transversal $\tau: G\to E(\phi)$ such that ϕ is the factor set for this extension with respect to τ.

PROOF. Let $g\in G$, $a\in A$. Note that $(g,0)^{\epsilon}=g$. By definition, the G-module structure induced by the extension is given by $(a\circ g)^{\mu}=(g,0)^{-1}a^{\mu}(g,0)=(1,ag-(1,1)\phi)=(ag)^{\mu}$, which gives the first part. For the second part, define $\tau: G\to E(\phi)$ by $g^{\tau}=(g,0)$. This is a transversal function and $x^{\tau}y^{\tau}=(xy)^{\tau}((x,y)\phi)^{\mu}$. ∎

By looking at factor sets, how can we determine which extensions are equivalent? Let A be a fixed G-module and let

$$A\rightarrowtail^{\mu_i} E_i\overset{\epsilon_i}{\twoheadrightarrow} G\ ,\qquad i=1,2$$

be two extensions realizing this module structure. Choose transversal functions τ_i and let ϕ_i be the resulting factor sets.

First suppose these extensions are equivalent:

$$\begin{array}{ccccc} A & \overset{\mu_1}{\rightarrowtail} & E_1 & \overset{\epsilon_1}{\twoheadrightarrow} & G \\ \downarrow 1 & & \downarrow \theta & & \downarrow 1 \\ A & \overset{\mu_2}{\rightarrowtail} & E_2 & \overset{\epsilon_2}{\twoheadrightarrow} & G \end{array}$$

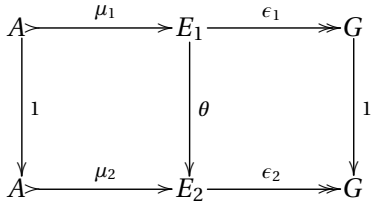
Then $\overline{\tau}_2=\tau_1\theta$ is a transversal for the second extension. Applying θ to

$$x^{\tau_1}y^{\tau_1}=(xy)^{\tau_1}((x,y)\phi_1)^{\mu_1},$$

we get $x^{\overline{\tau}_2}y^{\overline{\tau}_2}=(xy)^{\overline{\tau}_2}((x,y)\phi_1)^{\mu_2}$, hence $\overline{\tau}_2$ determines the factor set ϕ_1 for the second extension. As the factor sets of τ_2 and $\overline{\tau}_2$ belong to the same coset of $B^2(G,A)$, we get

$$\phi_1+B^2(G,A)=\phi_2+B^2(G,A).$$

Conversely, assume that $\phi_1 + B^2(G,A) = \phi_2 + B^2(G,A)$. Write $\phi_1 = \phi_2 + \psi^*$ for some $\psi: G \to A$ as above. Define $\theta: E_1 \to E_2$ by the rule $(x^{\tau_1} a^{\mu_1})^\theta = x^{\tau_2}(a + (x)\psi)^{\mu_2}$ for $x \in G$ and $a \in A$. θ is a well defined homomorphism, $\mu_1\theta = \mu_2$ and $\epsilon_1 = \theta\epsilon_2$. Hence we have a commutative diagram



and θ must be an isomorphism.

Theorem 3.4.8 *Let G be a group and A a G-module. Then there is a bijection between the set of equivalence classes of of extensions of A by G inducing the given module structure and the group $Z^2(G,A)/B^2(G,A)$. The split extension corresponds to $B^2(G,A)$.*

Let A be a G-module. We define $H^2(G,A) = Z^2(G,A)/B^2(G,A)$ to be the *second cohomology group* of G with coefficients in A. The elements of $H^2(G,A)$ thus correspond to equivalence classes of extensions of A by G. Unfortunately, different elements of $H^2(G,A)$ can still produce extensions of A by G that are isomorphic as groups.

Example 3.4.9 *Consider $\mathbb{Z}_p$ as a trivial C_p-module. From Example 3.2.26 it follows that there are only two non-isomorphic extensions of $A = \mathbb{Z}_p$ by $G = C_p$, namely $C_p \times C_p$ and C_{p^2}. On the other hand, one can show that $H^2(C_p, \mathbb{Z}_p) \cong C_p$.*

GAP can compute extensions of elementary abelian p-groups by solvable groups, which have to be presented as `pc groups`. One has to define an elementary abelian group A together with an action of G on A as a `MeatAxe` module for G over a finite field; we refer to GAP's manual for further information. The action of G on A can be represented by matrices over GF(p). It is a requirement that the matrices that define the module must correspond to the `pcgs` of the group G. In this case, $Z^2(G,A)$, $\mathrm{B}^2(G,A)$ and $H^2(G,A)$ are elementary abelian p-groups and can be considered as vector spaces over GF(p).

As another example we build all the extensions of $A = \mathbb{Z}_2 \oplus \mathbb{Z}_2$ by $G = D_8$, where we consider $\mathbb{Z}_2 \oplus \mathbb{Z}_2$ as a trivial D_8-module. Along the way we show commands for computing 2-cocycles, extensions corresponding to given 2-cocycles, and split extensions. The way we build the action is as follows. To each element of `Pcgs(G)` we assign 2×2 identity matrix over GF(2). Then we build the module using the command `GModuleByMats`. The other commands we use are self-evident:

```
gap> G := DihedralGroup(8);;
gap> mats := List( Pcgs( G ), x -> IdentityMat( 2, GF(2) ) );;
gap> A := GModuleByMats( mats, GF(2) );;
gap> co := TwoCocycles( G, A );;
gap> Extension( G, A, co[2] );;
gap> StructureDescription(last);
"C2 x (C4 : C4)"
```

```
gap> SplitExtension( G, A );;
gap> StructureDescription(last);
"C2 x C2 x D8"
gap> ext := Extensions( G, A );;
gap> Length(ext);
64
gap> DuplicateFreeList(List(ext, IdGroup));
[ [ 32, 46 ], [ 32, 40 ], [ 32, 22 ], [ 32, 39 ], [ 32, 9 ], [ 32, 23 ],
  [ 32, 13 ], [ 32, 41 ], [ 32, 10 ], [ 32, 2 ], [ 32, 14 ] ]
```

Here note that the notation `C4 : C4` means that the group in question is a semidirect product of C_4 by C_4. The command `TwoCocycles(G, A)` returns a list of vectors over the field underlying A, and the additive group *generated* by these vectors is the $Z^2(G,A)$. There is also a command `TwoCohomology(G, A)` that returns a record defining the second cohomology group as factor space of the vector space of cocycles by the subspace of coboundaries. We refer to GAP's manual for further details.

```
gap> z2 := AdditiveGroupByGenerators(co);;
gap> Length(Elements(z2));
256
gap> h2 := TwoCohomology(G, A);;
gap> h2.cohom;
<linear mapping by matrix, <vector space of dimension
8 over GF(2)> -> ( GF(2)^6 )>
gap> dimensionZ2 := Dimension(Source(h2.cohom));
8
gap> dimensionB2 := Dimension(Kernel(h2.cohom));
2
gap> dimensionH2 := Dimension(Image(h2.cohom));
6
```

The last line tells us that $H^2(G,A) \cong C_2^6$.

3.4.4 The Schur-Zassenhaus theorem

Let A and G be groups. We say that an extension of A by G *splits* if it is a semidirect product.

Theorem 3.4.10 *Suppose that A and G are finite groups satisfying* $\gcd(|A|,|G|) = 1$. *Then every extension of A by G splits.*

We will only prove this result in the case when A is abelian. In this form, the result was originally due to Schur. Zassenhaus improved it by showing that it suffices to assume that one of A or G is solvable. On the other hand, Feit-Thompson's Odd Order Theorem shows that this assumption is redundant.

PROOF.[Proof of Theorem 3.4.10 when A is abelian] Let $m = |A|$ and $n = |G|$. Let $\phi : G \times G \to A$ be a 2-cocycle representing an extension of A by G, and let $\chi : G \to \mathrm{Aut}(A)$ be the homomorphism that induces the corresponding G-module structure on A. We claim

that $n\phi \in B^2(G,A)$. Define a function $d: G \to A$ by

$$(g)d = \sum_{g_1 \in G} (g_1, g)\phi.$$

Consider the cocycle identity:

$$(g_1, g_2 g_3)\phi + (g_2, g_3)\phi = (g_1 g_2, g_3)\phi + (g_1, g_2)\phi \cdot g_3.$$

Sum this equation over $g_1 \in G$:

$$\begin{aligned}(g_2 g_3)d + n(g_2, g_3)\phi &= (g_2)d \cdot g_3 + \sum_{g_1 \in G} (g_1 g_2, g_3)\phi \\ &= (g_2)d \cdot g_3 + \sum_{g_1 g_2 \in G} (g_1 g_2, g_3)\phi \\ &= (g_2)d \cdot g_3 + (g_3)d.\end{aligned}$$

Therefore $n(g_2, g_3)\phi = (g_2)d \cdot g_3 + (g_3)d - (g_2 g_3)d$, which proves our claim. Now, there exist integers a and b with $am + bn = 1$. Since $|A| = m$, it follows that $m\phi = 0$. Therefore $\phi = (am + bn)\phi = bn\phi \in B^2(G,A)$. Thus every extension of A by G splits. ■

3.4.5 Problems

1. Let G_1, G_2 and G_3 be groups. Show that $(G_1 \wr G_2) \wr G_3$ may not be isomorphic to $G_1 \wr (G_2 \wr G_3)$.
2. Find a proof of Theorem 3.4.2.
3. Prove that a Sylow p-subgroup of S_{p^n} is isomorphic to $W(p,n) = (\cdots(C_p \wr Cp) \wr \cdots) \wr C_p$, the number of factors being n.
4. Prove that every group of order p^n is isomorphic to a subgroup of $W(p,n)$.
5. Let $1 \longrightarrow A \xrightarrow{\mu} E \xrightarrow{\epsilon} G \longrightarrow 1$ be a group extension, where A is abelian and $G = \langle g \rangle$ cyclic of order n. Choose $x \in E$ with $x^\epsilon = q$, and let $a = x^n$. Define a transversal function $\tau: G \to E$ by $(g^i)^\tau = x^i$ for $0 \leq i < n$. Prove that the corresponding factor set $\phi: G \times G \to A$ is given by
$$(g^i, g^j)\phi = \begin{cases} 0 & : \quad i + j < n \\ a & : \quad i + j \geq n \end{cases}.$$
6. Find all equivalence classes of extensions of C_4 by C_2 by hand. Which groups arise this way?
7. Find all equivalence classes of extensions of D_8 by C_2 by hand. Which groups arise this way?

8. Fill in the details in Example 3.4.9.

9. Let N be a normal subgroup of a finite group G, and assume that $|N| = n$ and $|G:N| = m$ are relatively prime. Let m_1 be a divisor of m. Then a subgroup of G of order m_1 is contained in a subgroup of order m.

3.5 Nilpotent groups and p-groups

Nilpotent groups are groups which can be constructed from abelian groups by repeatedly forming central extensions. We exhibit some of the classical theory of these groups, and show that they are closely related to finite p-groups. These form a very rich class of groups. We prove that there are lots of finite p-groups, hence there is little hope to classify them up to isomorphism.

3.5.1 Nilpotent groups

Definition and basic properties

We call $1 = G_0 \subset G_1 \subset \cdots \subset G_n = G$ a *normal series* of G if each of its members is a normal subgroup of G. A group G is *nilpotent* if it has a *central series*, i.e. a normal series $1 = G_0 \subset G_1 \subset \cdots \subset G_n = G$ in which each factor G_{i+1}/G_i is contained in the center of G/G_i. The length of the shortest central series of G is called the *nilpotency class* of G.

All nilpotent groups are solvable. Nilpotent groups of class no more than 1 are abelian. The smallest solvable non-nilpotent group is S_3.

Here is an example of how to manipulate nilpotent groups in GAP:

```
gap> l := AllSmallGroups(Size, 54, IsNilpotent, true);
[ <pc group of size 54 with 4 generators>,
  <pc group of size 54 with 4 generators>,
  <pc group of size 54 with 4 generators>,
  <pc group of size 54 with 4 generators>,
  <pc group of size 54 with 4 generators> ]
gap> NilpotencyClassOfGroup(l[2]);
1
gap> NilpotencyClassOfGroup(l[3]);
2
gap> ForAll(AllSmallGroups(54), IsNilpotent);
false
gap> G:= First(AllSmallGroups(54), x->not IsNilpotent(x));;
gap> StructureDescription(G);
"D54"
gap> List(l, StructureDescription);
[ "C54", "C18 x C3", "C2 x ((C3 x C3) : C3)", "C2 x (C9 : C3)",
  "C6 x C3 x C3" ]
```

From the above example we observe that all nilpotent groups of order 54 can be written as direct products of their Sylow p-subgroups. We will show later on that this

property characterizes finite nilpotent groups. We now exhibit a large class of nilpotent groups:

Lemma 3.5.1 *All finite p-groups are nilpotent.*

PROOF. We know that $Z(G)$ is nontrivial by Proposition 3.2.25. Now use induction on the order of G to show that $G/Z(G)$ is nilpotent. From here it easily follows that G is nilpotent as well. ∎

The following is straightforward to prove:

Lemma 3.5.2 *Subgroups, homomorphic images and finite direct products of nilpotent groups are nilpotent.*

We note that nilpotency is not closed under extensions, since S_3 is an extension of C_3 by C_2.

Commutators

The theory of nilpotent groups relies significantly on commutator calculus that we briefly develop here. A *simple commutator of length n* of elements $x_1,\ldots,x_n \in$G is defined inductively by $[x_1]=x_1$ and

$$[x_1,x_2,\ldots,x_n]=[[x_1,\ldots,x_{n-1}],x_n].$$

Lemma 3.5.3 *Let x,y,z be elements of a group. Then*

1. $[x,y]=[y,x]^{-1}$;
2. $[xy,z]=[x,z]^y[y,z]$ *and* $[x,yz]=[x,z][x,y]^z$;
3. $[x,y^{-1}]=([x,y]^{y^{-1}})^{-1}$ *and* $[x^{-1},y]=([x,y]^{x^{-1}})^{-1}$;
4. *(the Hall-Witt identity)* $[x,y^{-1},z]^y[y,z^{-1},x]^z[z,x^{-1},y]^x=1$.

PROOF. Let us only sketch the proof of the Hall-Witt identity. Observe that

$$[x,y^{-1},z]^y=x^{-1}y^{-1}xz^{-1}x^{-1}yxy^{-1}zy=u^{-1}v,$$

where $u=z^{x^{-1}}yx$ and we obtain v by cyclically permuting x,y,z in the definition of u. The rest is now immediate. ∎

These identities could also be proved using GAP. For example, in order to prove the identity $[xy,z]=[x,z]^y[y,z]$, it suffices that this holds in the free group generated by x,y,z:

```
gap> F:=FreeGroup( "x", "y", "z" );;
gap> AssignGeneratorVariables( F );;
gap> Comm( x * y, z ) = Comm( x, z )^y * Comm( y, z );
true
```

Let $X, Y \subset G$ be non-empty sets. Define the *commutator subgroup* of X and Y by $[X,Y] = \langle [x,y] \mid x \in X, y \in Y \rangle$ and note that $[X,Y] = [Y,X]$. For any $n \geqslant 2$ nonempty subsets $X_1, X_2, \ldots, X_n$ of G denote

$$[X_1, X_2, \ldots, X_n] = [[X_1, \ldots, X_{n-1}], X_n].$$

Note that $[G,G] = G'$ is just the derived subgroup of G. Define also $X^Y = \langle x^y \mid x \in X, y \in Y \rangle$. If X is a subset and $H \leqslant G$, then $X \subset X^H \lhd \langle X, H \rangle$. Thus, $X^H = X^{\langle X,H \rangle}$ is the normal closure of X in $\langle X, H \rangle$.

Here is an example:

```
gap> G := SmallGroup( 64, 52);;
gap> gen := GeneratorsOfGroup(G);;
gap> H := Subgroup(G, [gen[1]]);;
gap> K := Subgroup(G, [gen[2], gen[3]]);;
gap> C := CommutatorSubgroup(H,K);;
gap> Order(H);
2
gap> Order(K);
32
gap> Order(C);
16
```

Lemma 3.5.4 *Let $X \subset G$ and $H \leqslant G$. Then*

1. $X^K = \langle X, [X,K] \rangle$;
2. $[X,K]^K = [X,K]$;
3. *if $K = \langle Y \rangle$, then $[X,K] = [X,Y]^K$.*

PROOF. (1) Follows from $x^k = x[x,k]$.

(2) For $k, h \in K$ and $x \in X$ we have $[x,hk] = [x,k][x,h]^k$, so that $[x,h]^k \in [X,K]$.

(3) It suffices to show that $[x,k] \in [X,Y]^K$ what we prove for $k = y_1^{\pm 1} y_2^{\pm 1} \ldots y_r^{\pm 1}$ by induction on r. For $r = 1$ we get $[x, y_1^{-1}] = ([x,y_1]^{y_1^{-1}})^{-1} \in [X,Y]^K$. For the inductive step we write $k = k' y_r^{\pm 1}$. Then $[x,k] = [x, k' y_r^{\pm 1}] = [x, y_r^{\pm 1}][x,k']^{y_r^{\pm 1}} \in [X,Y]^K$ by induction. ∎

Corollary 3.5.5 *If $H = \langle X \rangle$ and $K = \langle Y \rangle$, then $[H,K] = [X,Y]^{HK}$.*

PROOF. This follows from Lemma 3.5.4, (3). ∎

Derived series, upper and lower central series

Define $G' = [G,G]$ and inductively $G^{(0)} = G$ and $G^{(n+1)} = (G^{(n)})'$. The *derived series* of G is the series

$$G^{(0)} \geqslant G^{(1)} \geqslant G^{(2)} \geqslant \cdots$$

of fully invariant (and therefore normal) subgroups of G. The derived series of a group is in close connection with its solvability:

Proposition 3.5.6 *If $1 = G_0 \lhd G_1 \lhd \cdots \lhd G_n = G$ is an abelian series of a solvable group G, then $G^{(i)} \leqslant G_{n-i}$ and, in particular, $G^{(n)} = 1$. The derived length of G is equal to the length of the derived series.*

PROOF. We prove this by induction, the case $i = 0$ being trivial. If the assertion is true for i, then $G^{(i+1)} = (G^{(i)})' \leqslant (G_{n-i})' \leqslant G_{n-i-1}$, as required. ∎

GAP can compute the derived series as follows:

```
gap> G := OneSmallGroup(Size, 120, IsAbelian, false, IsSolvable, true);;
gap> StructureDescription(G);
"C5 x (C3 : C8)"
gap> DerivedSeries(G);
[ C5 x (C3 : C8), Group([ f5 ]), Group([ ]) ]
gap> DerivedLength(G);
2
```

There are two canonical central series of a given group. Define $\gamma_1(G) = G$ and inductively $\gamma_{n+1}(G) = [\gamma_n(G), G]$. The result is the *lower central series*

$$G = \gamma_1(G) \geqslant \gamma_2(G) \geqslant \cdots$$

of fully invariant (and therefore normal) subgroups. The factor group $\gamma_n(G)/\gamma_{n+1}(G)$ lies in the center of $G/\gamma_{n+1}(G)$.

Define $Z_0(G) = 1$ and inductively $Z_{n+1}(G)/Z_n(G) = Z(G/Z_n(G))$. We obtain the *upper central series*

$$1 = Z_0(G) \leqslant Z_1(G) \leqslant Z_2(G) \leqslant \cdots$$

of characteristic (and therefore normal) subgroups of G. If G is finite, it terminates in a subgroup called the *hypercenter* of G.

Proposition 3.5.7 *If $1 = G_0 \leqslant G_1 \leqslant \cdots \leqslant G_n = G$ is a central series of a nilpotent group G, then*

1. *$\gamma_i(G) \leqslant G_{n-i+1}$, so that $\gamma_{n+1}(G) = 1$;*
2. *$G_i \leqslant Z_i(G)$so that $Z_n(G) = G$;*

3. *the nilpotency class of G equals the length of the upper central series which also equals the length of the lower central series.*

PROOF. (1). This is true for $i = 1$. Since $G_{n-i+1}/G_{n-i} \subset Z(G/G_{n-i})$, we have $[G_{n-i+1}, G] \subset G_{n-i}$. By induction,$\gamma_{i+1}(G) = [\gamma_i(G), G] \leqslant [G_{n-i+1}, G] \leqslant G_{n-i}$. The item (2) is another easy induction and (3) follows. ∎

Lemma 3.5.8 (The three subgroup lemma) *Let $H, K, L \leqslant G$. If two of the commutator subgroups $[H,K,L], [K,L,H], [L,H,K]$ are contained in a normal subgroup of G, then so is the third one.*

PROOF. By Corollary 3.5.5, $[H,K,L]$ is generated by conjugates of commutators of the form $[h,k^{-1},l]$. Apply the Hall-Witt identity. ∎

Proposition 3.5.9 *Let G be a group and $i, j \in \mathbb{N}$:*

1. $[\gamma_i(G), \gamma_j(G)] \leqslant \gamma_{i+j}(G)$.
2. $\gamma_i(\gamma_j(G)) \leqslant \gamma_{ij}(G)$.
3. $[\gamma_i(G), Z_j(G)] \leqslant Z_{j-i}(G)$ *if* $j \geqslant i$.
4. $Z_i(G/Z_j(G)) = Z_{i+j}(G)/Z_j(G)$

PROOF. (1) Both $[\gamma_i(G), \gamma_j(G), G]$ and $[G, \gamma_i(G), \gamma_j(G)]$ are inductively (on j) contained in $\gamma_{i+j+1}(G)$. By the three subgroup lemma the same holds true for $[\gamma_j(G), G, \gamma_i(G)] = [\gamma_i(G), \gamma_{j+1}(G)]$.

(2) This goes by induction on i: $\gamma_{i+1}(\gamma_j G) = [\gamma_i(\gamma_j(G)), \gamma_j(G)] \leqslant [\gamma_{ij}(G), \gamma_j(G)] \leqslant \gamma_{(i+1)j}(G)$.

(3) $[\gamma_{i+1}(G), Z_j(G)] = [\gamma_i(G), G, Z_j(G)] \leqslant [G, Z_j(G), \gamma_i(G)][Z_j(G), \gamma_i(G), G] \leqslant Z_{j-i-1}(G)$ by induction on i.

(4) Induction on i. ∎

Corollary 3.5.10 *For any group G we have that $G^{(i)} \leqslant \gamma_{2^i}(G)$. If G is nilpotent of class c, then its derived length is at most $\lfloor \log_2 c \rfloor + 1$.*

PROOF. Apply part (2) of the above proposition to

$$G^{(i)} = \underbrace{\gamma_2(\cdots(\gamma_2(G))\cdots)}_{i \text{ times}}$$

Now, let G be nilpotent of class c, let d be the derived length and let $2^i \geqslant c+1$. Then, $G^{(i)} \leqslant \gamma_{2^i}(G) \leqslant \gamma_{c+1}(G) = 1$. Since the smallest such i is $\lfloor \log_2 c \rfloor + 1$, it follows that $d \leqslant \lfloor \log_2 c \rfloor + 1$. ∎

Here is a sample computation of lower and upper central series of a group:

```
gap> G := SmallGroup(128, 50);;
gap> NilpotencyClassOfGroup(G);
4
gap> DerivedLength(G);
2
gap> LowerCentralSeriesOfGroup(G);
[ <pc group of size 128 with 7 generators>, Group([ f3, f5, f7 ]),
  Group([ f5, f7 ]), Group([ f7 ]), Group([ <identity> of ... ]) ]
gap> UpperCentralSeriesOfGroup(G);
[ Group([ f6, f7, f5, f3, f4, f1, f2 ]), Group([ f6, f7, f5, f3, f4 ]),
  Group([ f6, f7, f5 ]), Group([ f6, f7 ]), Group([ ]) ]
```

Unitriangular groups

Here is a ring-theoretic source of examples of nilpotent groups. Let S be a ring with identity and N a subring. Write $N^{(i)}$ for the set of all sums of products of i elements of N for $i > 0$, which is necessarily a subring. If $N^{(i)} = 0$ for some $i > 0$, then N is called *nilpotent*. Assume $N^{(n)} = 0$ and let U be the set of all elements of the form $1+x$ for $x \in N$. Then U is a group with respect to the ring multiplication, i.e.

$$(1+x)(1+y) = 1+(x+y+xy)$$

and

$$(1+x)^{-1} = 1+(-x+x^2-\cdots+(-x)^{n-1}).$$

Define $U_i = \{1+x \mid x \in N^{(i)}\}$ and observe that U_i is an increasing series of subgroups. We want to show that this is actually a central series of U. Let $x \in N^{(r)}$ and $y \in N^{(s)}$, then

$$[1+x,1+y] = (1+x+y+yx)^{-1}(1+x+y+xy).$$

We let $u = x+y+xy$ and $v = x+y+yx$:

$$[1+x,1+y] = (1-v+v^2-\cdots+(-v)^{n-1})(1+u) =$$

$$1+(1-v+v^2-\cdots+(-v)^{n-2})(u-v)+(-v)^{n-1}u.$$

Now, $u-v = xy-yx \in N^{(r+s)}$ and $(-v)^{n-1}u = 0$. We have thus shown that $[U_r, U_s] \leqslant U_{r+s}$ implying that U is nilpotent of class no more than $n-1$.

For an even more concrete example, let us take S to be the ring of all $n \times n$ matrices over a commutative ring with identity R. Further, let N be the subring of all strictly upper

triangular matrices. It is not hard to see that the class of U in this case is exactly $n-1$ showing that there are nilpotent groups of arbitrary class. We note here that in the case $n=3$ we call the group U a *Heisenberg group* over R.

Observe that U_i consists of all upper unitriangular matrices whose first $i-1$ super diagonals are zero. It easily follows that

$$U_i/U_{i+1} \simeq \underbrace{R \oplus R \oplus \cdots \oplus R}_{n-i \text{ times}}.$$

In the case that $R = \mathrm{GF}(p)$ we find U to be a finite p-group of order $p^{n(n-1)/2}$. On the other hand, if $R=\mathbb{Z}$, then U is a finitely generated torsion-free nilpotent group.

Now, let T denote the group of all upper triangular invertible matrices over R. Let $\theta : T \to (R^*)^n$ be the projection of a matrix to its diagonal. So, this is an epimorphism whose kernel is precisely equal to U and whose image is an abelian group. It follows that T is solvable, with the derived length being no more than $[\log_2(n-1)+2]$.

Properties of nilpotent groups

Lemma 3.5.11 *If G is a nilpotent group and $1 \neq N \lhd G$, then $N \cap Z(G) \neq 1$.*

Proof. Let i be the smallest natural number s.t. $N \cap Z_i(G) \neq 1$. Then, $[N \cap Z_i(G), G] \leqslant N \cap Z_{i-1}(G) = 1$, so that $N \cap Z_i(G) \leqslant N \cap Z_1(G) \neq 1$ implying equality. ∎

Corollary 3.5.12 *A minimal normal subgroup of a nilpotent group is contained in the center.*

Proposition 3.5.13 *If A is a maximal normal abelian subgroup of the nilpotent group G, then $A = C_G(A)$.*

Proof. Clearly $A \leqslant C = C_G(A)$. Suppose that $A \neq C$. Then C/A is a nontrivial normal subgroup of the nilpotent G/A. By Lemma 3.5.11 there is an $A \neq Ax \in (C/A) \cap Z(G/A)$. Now $\langle x, A \rangle$ is abelian and normal leading to a contradiction. ∎

Theorem 3.5.14 *The following conditions are equivalent for a finite group G:*

1. *G is nilpotent;*
2. *every subgroup of G is subnormal;*
3. *Every proper subgroup H of G is properly contained in its normalizer;*
4. *Every maximal subgroup of G is normal;*

5. G is the direct product of its Sylow subgroups.

PROOF. (1) $\Rightarrow$ (2). Let G be nilpotent with class c. If $H \leqslant G$, then $HZ_iG \lhd HZ_{i+1}G$ since $Z_{i+1}G/Z_iG = Z(G/Z_iG)$. So, HZ_iG is the series proving subnormality of H.

(2) $\Rightarrow$ (3). Let $H = H_0 \lhd H_1 \lhd \cdots \lhd H_n = G$ be the series proving subnormality of the proper subgroup H. Let i be the smallest integer s.t. $H \neq H_i$. Then, $H = H_{i-1} \lhd H_i \leqslant N_G(H)$.

(3) $\Rightarrow$ (4). If $M < G$ is maximal, then $M < N_G(M)$ implying $N_G(M) = G$.

(4) $\Rightarrow$ (5). Assume P is a non-normal Sylow subgroup. Then $N_G(P)$ is proper and therefore contained in a maximal subgroup M. Then $M \lhd G$ contradicting Lemma 3.2.29. Thus, Sylow p-subgroup is normal and consequently unique for each p. Their product is clearly direct and equal to G.

(5) $\Rightarrow$ (1). This follows since every p-group is nilpotent and direct sum of nilpotent groups is nilpotent. ∎

In the case of infinite groups, properties (2) to (5) are weaker than (1). Using the above result, one can refine Corollary 3.3.7 as follows:

Corollary 3.5.15 *A maximal subgroup M of a finite nilpotent group G has prime index.*

PROOF. We known that $M \triangleleft G$, and $|G : M| = p^k$ by Corollary 3.3.7. If $k > 1$, then there exists $H < G$ containing M such that $|H : M| = p$ which is a contradiction. ∎

The Fitting Subgroup

Theorem 3.5.16 (Fitting) *Let M and N be normal nilpotent subgroups of a group G. If c and d are nilpotency classes of M and N, then $L = MN$ is nilpotent of class $\leq c + d$.*

PROOF. By induction on i we show that

$$\gamma_i(L) = \prod_{X_j \in \{M,N\}} [X_1, \ldots, X_i].$$

Taking $i = c + d + 1$ and noting that $[A, G] \leq A$ for all $A \triangleleft G$, we conclude that each $[X_1, \ldots, X_i]$ is contained in either $\gamma_{c+1}(M)$ or $\gamma_{d+1}(N)$, both of which equal to 1. ∎

The subgroup $\mathrm{Fit}(G)$ generated by all the normal nilpotent subgroups of a group G is called the *Fitting subgroup* of G. If the group G is finite, then $\mathrm{Fit}(G)$ is nilpotent. In these cases, $\mathrm{Fit}(G)$ is the unique largest normal nilpotent subgroup of G. Note also that $\mathrm{Fit}(G) = 1$ if and only if G is semisimple.

Let $N \leq H \leq G$ and $N \triangleleft G$. Define $C_G(H/N) = \{g \in G : [H, g] \leq N\}$. Clearly $C_G(H/N) \leq G$.

Theorem 3.5.17 *Let G be a finite group. For a prime p let $O_p(G)$ be the largest normal p-subgroup of G. The following groups are then equal to* Fit(G)*:*

(a) The direct product of all $O_p(G)$, where p divides $|G|$.

(b) The intersection of the centralizers of the chief factors of G.

PROOF. (a) If $N \triangleleft G$ is nilpotent, then $N = \times O_p(N)$. As the group $O_p(N)$ is a characteristic subgroup of N, it follows that $O_p(N) \triangleleft G$. Therefore $O_p(N) \leq O_p(G)$, and thus $N \leq \times O_p(G)$.

(b) Let $1 = G_0 \leq G_1 \leq \cdots \leq G_n = G$ be a chief series of G and denote

$$I = \bigcap_i C_G(G_{i+1}/G_i).$$

Since $[G_{i+1}, I] \leq G_i$ for all i, we get $\gamma_{n+1}(I) = 1$, hence $I \leq \mathrm{Fit}(G)$. Conversely, let $F = \mathrm{Fit}(G)$. Since G_1 is a minimal normal subgroup of G, we get either $[G_1, F] = 1$ or $[G_1, F] = G_1$. In the latter case, $G_1 \leq \gamma_{c+1}(F) = 1$ for some c, a contradiction. Thus $[G_1, F] = 1$. Induction on n shows that $F \leq C_G(G_{i+1}/G_i)$ for all i. ∎

```
gap> G := SmallGroup(96, 10);;
gap> IsNilpotent(G);
false
gap> F := FittingSubgroup(G);;
gap> Order(F);
48
gap> StructureDescription(F);
"C12 x C4"
```

The Frattini subgroup

The *Frattini subgroup* Frat(G) of G is the intersection of all maximal subgroups of G (if G does not have maximal subgroups, then we define $\mathrm{Frat}(G) = G$). Clearly Frat(G) is a characteristic subgroup of G. We say that $g \in G$ is a *nongenerator* of G if $G = \langle g, X \rangle$ implies $G = \langle X \rangle$ for every $X \subseteq G$.

Theorem 3.5.18 Frat(G) *equals the set of nongenerators of G.*

PROOF. Let $g \in \mathrm{Frat}(G)$, $G = \langle g, X \rangle$, but $G \neq \langle X \rangle$. There exists $M \leq G$ which is maximal subject to $\langle X \rangle \leq M$ and $g \notin M$. M is a maximal subgroup of G, hence $g \in M$, a contradiction.

Let g be a nongenerator and $g \notin \mathrm{Frat}(G)$. Thus $g \notin M$ for some maximal subgroup M. It follows $\langle g, M \rangle = G$, hence $G = M$, a contradiction. ∎

Proposition 3.5.19 *Let G be a finite group.*

(a) *If* $N \triangleleft G$, $H \leq G$ *and* $N \leq \mathrm{Frat}(H)$, *then* $N \leq \mathrm{Frat}(G)$.

(b) *If* $K \triangleleft G$, *then* $\mathrm{Frat}(K) \leq \mathrm{Frat}(G)$.

(c) *If* $N \triangleleft G$, *then* $\mathrm{Frat}(G/N) \geq \mathrm{Frat}(G)N/N$, *with equality if* $N \leq \mathrm{Frat}(G)$.

(d) *If* A *is an abelian normal subgroup of* G *such that* $\mathrm{Frat}(G) \cap A = 1$, *there exists* $H \leq G$ *such that* $G = HA$ *and* $H \cap A = 1$.

PROOF.

(a) If not, then there exists a maximal subgroup M such that $N \not\leq M$. Then $G = MN$, $H = (H \cap M)N$, thus $H \leq M$, therefore $N \leq M$, a contradiction.

(b) Apply (a) with $N = \mathrm{Frat}(K)$ and $H = K$.

(c) By definition.

(d) Let H be minimal subject to $G = HA$. Then $H \cap A \triangleleft G$. If $H \cap A \leq \mathrm{Frat}(H)$, then we claim that $H \cap A = 1$ by (a). Namely, if this were false, there would exist a maximal subgroup M of H such that $H \cap A \not\leq M$. Then $H = M(H \cap A)$ and $G = MA$, contrary to the minimality of H.

■

Theorem 3.5.20 (Gaschütz) *Let* G *be a group.*

(a) *If* $\mathrm{Frat}(G) \leq H \leq G$, *where* H *is finite and* $H/\mathrm{Frat}(G)$ *is nilpotent, then* H *is nilpotent.*

(b) *If* G *is finite,* $\mathrm{Frat}(G)$ *is nilpotent.*

(c) *Define* $\mathrm{FFrat}(G)$ *by* $\mathrm{FFrat}(G)/\mathrm{Frat}(G) = \mathrm{Fit}(G/\mathrm{Frat}(G))$. *If* G *is finite, then* $\mathrm{FFrat}(G) = \mathrm{Fit}(G)$.

(d) *If* G *is finite,* $\mathrm{FFrat}(G)/\mathrm{Frat}(G)$ *is the product of all the abelian minimal normal subgroups of* $G/\mathrm{Frat}(G)$.

PROOF.

(a) Let P be a Sylow subgroup of H, $F = \mathrm{Frat}(G)$, and $K = PF \leq H$. K/F is a Sylow subgroup of H/F, hence K/F is characteristic in H/F. Hence K is normal in G. By the Frattini argument, $G = N_G(P)K = N_G(P)F = N_G(P)$.

(b) Follows from (a).

(c) Denote $H = \mathrm{FFrat}(G)$. H is nilpotent by (a), thus $H \leq \mathrm{Fit}(G)$.

(d) Taking quotients, we may assume that $\mathrm{Frat}(G) = 1$. Write $L = \mathrm{Fit}(G)$. $L/\mathrm{Frat}(L)$ is abelian, hence $L' \leq \mathrm{Frat}(L) \leq \mathrm{Frat}(G) = 1$. Thus L is abelian. Let N be the product of all the abelian minimal normal subgroups of G. Then $N \leq L$. There exists $H \leq G$ such that $G = HN$ and $N \cap H = 1$. $H \cap L$ is normal in $HL = G$. Since $H \cap L \cap N = 1$, it follows that $H \cap L = 1$ by the minimality. Then $L = L \cap (HN) = N$.

■

Proposition 3.5.21 *Let G be a finite group. Then G is nilpotent if and only if $G' \leq \mathrm{Frat}(G)$.*

PROOF. If G is nilpotent and M a maximal subgroup of G, then $G' \leq M$. Conversely, if $G' \leq \mathrm{Frat}(G)$ then every maximal subgroup of G is normal. ∎

```
gap> G := SmallGroup(96, 10);;
gap> F := FrattiniSubgroup(G);;
gap> StructureDescription(F);
"C4 x C2"
```

3.5.2 Finite p-groups

Basic properties

Proposition 3.5.22 *Let G be a group of order p^{m+1}.*

(a) If G is nilpotent of class $c > 1$, then $G/Z_{c-1}(G)$ is not cyclic.

(b) $c \leq m$.

(c) If $0 \leq i \leq j \leq m+1$, every subgroup of order p^i is contained in some subgroup of order p^j.

(d) G has subgroups of every order dividing p^{m+1}.

PROOF. (a) If $G/Z_{c-1}(G)$ were cyclic, $G/Z_{c-2}(G)$ would be abelian, hence $Z_{c-1}(G) = G$, a contradiction.

(b) $|G : Z_{c-1}(G)| \geq p^2$ by (a), all upper central factors have order $\geq p$.

(c) Let H be a subgrup of order p^i. As H is subnormal in G, it is a part of a composition series $1 = H_0 \leq \cdots \leq H_i = H \leq \cdots \leq H_{m+1} = G$ by Jordan-Hölder's theorem. All composition factors have order p, hence the assertion.

(d) Follows from (c). ∎

Lemma 3.5.23 *Let G be an elementary abelian p-group. Then* $\mathrm{Frat}(G) = 1$.

PROOF. Let $G = C_p^n$ and let $M_i = \{(x_1, \ldots x_{i-1}, 1, x_{i+1}, \ldots, x_n) : x_j \in C_p\}$ for $i = 1, \ldots, n$. Then M_i are maximal subgroups of G and $\bigcap_{i=1}^n M_i = 1$, hence $\mathrm{Frat}(G) = 1$. ∎

Theorem 3.5.24 (The Burnside Basis Theorem) *Let G be a finite p-group. Then* $\mathrm{Frat}(G) = \gamma_2(G)G^p$, *where $G^p = \langle g^p \mid g \in G \rangle$. Also if $|G : \mathrm{Frat}(G)| = p^r$, then every set of generators of G has a subset of r elements which also generates G.*

PROOF. Let M be a maximal subgroup of G. Then $M \triangleleft G$ and $|G:M| = p$. It follows that $\gamma_2(G)G^p \leq M$, hence $\gamma_2(G)G^p \leq \mathrm{Frat}(G)$. On the other hand, $G/\gamma_2(G)G^p$ is an elementary abelian p-group, hence $\mathrm{Frat}(G/\gamma_2(G)G^p) = 1$. It follows that $\mathrm{Frat}(G) \leq \gamma_2(G)G^p$.

Let $G = \langle x_1, \ldots, x_s \rangle$ and $F = \mathrm{Frat}(G)$. Then $\overline{G} = G/F = \langle Fx_1, \ldots, Fx_s \rangle$. The group $\overline{G}$ is a vector space over $\mathrm{GF}(p)$, hence it has a basis $\{Fx_{i_1}, \ldots, Fx_{i_r}\}$. Write $Y = \langle x_{i_1}, \ldots, x_{i_r} \rangle$. Then $G = \langle Y, F \rangle$, hence $G = \langle Y \rangle$. ∎

Let G be a finite p-group. By the Burnside Basis Theorem, we can think of $G/\mathrm{Frat}(G)$ as a vector space over $\mathrm{GF}(p)$.

Corollary 3.5.25 *Let G be a finite p-group and d the minimal number of generators of G. Then $d = \dim_{GF(p)} G/\mathrm{Frat}(G)$.*

Extraspecial p-groups

A finite p-group is said to be *extraspecial* if $G' = Z(G) \cong C_p$.

Proposition 3.5.26 *Let G be a nonabelian group of order p^3. If p is odd, then G is isomorphic with*

$$\langle x, y \mid x^p = y^p = 1, [x,y]^x = [x,y]^y = [x,y] \rangle$$

or

$$\langle x, y \mid x^{p^2} = 1 = y^p, x^y = x^{1+p} \rangle.$$

These groups have exponent p and p^2 respectively. If $p = 2$, then G is isomorphic with D_8 or quaternion group Q_8. In particular, all non-abelian groups of order p^3 are extraspecial.

PROOF. All the groups given above have order p^3. For $p = 2$, the assertion follows from the description of all groups of order 8 (exercise).

Assume that p is odd. We consider two cases:

Case 1. All elements of G have order p. Let $z \in Z(G)\backslash\{1\}$ and let $x \notin Z(G)$. Then $\langle z, x \rangle = \langle z \rangle \times \langle x \rangle$ is a subgroup of order p^2, hence it is a maximal subgroup and thus normal in G. Choose $w \notin \langle z, x \rangle$. Then $G = \langle z, x, w \rangle$. We have that $x^w = x^a z^b$ for some $0 \leq a, b < p$. If $a = 0$, then $x^y \in Z(G)$, hence $x \in Z(G)$, a contradiction. Thus there exists c such that $ac \equiv 1 \mod p$. Let $t = w^c$. We have that $G = \langle z, x, t \rangle$, and $x^t = xz^{b'}$ for some $0 \leq b' < p$. As G is nonabelian, $b' \neq 0$, hence there exists d such that $b'd \equiv 1 \mod p$. Put $y = t^d$. Then we get $[x,y] = z$ and $G = \langle x, y \rangle$. We have

$$x^p = y^p = 1, [x,y]^x = [x,y]^y = [x,y],$$

as required.

Case 2. G contains an element x of order p^2. Let $N = \langle x \rangle$. As N is a maximal subgroup of G, N is normal in G. Choose $z \in G\backslash N$ of order p. There exists $a \in \mathbb{Z}$ such that $x^z = x^a$.

Since $x = x^{z^p}$, it follows that $a^p \equiv 1 \mod p^2$, hence $a \equiv 1 \mod p$. Write $a = 1 + kp$. Let l be such that $kl \equiv 1 \mod p$. Let $y = z^l$. Then $x^y = x^{1+p}$. Since $N \cap \langle y \rangle = 1$, we have $N\langle y \rangle = G$.

All the groups above are clearly extraspecial. ∎

A group G is said to be the *central product* of its normal subgroups $G_1, \ldots, G_n$ if $G = G_1 \cdots G_n$, $[G_i, G_j] = 1$ for $i \neq j$, and $G_i \cap \prod_{j \neq i} G_j = Z(G)$.

Theorem 3.5.27 *An extraspecial p-group is a central product of n nonabelian subgroups of order p^3, and has order p^{2n+1}. Conversely, a finite central product of nonabelian groups of order p^3 is an extraspecial p-group.*

PROOF. Let $C = Z(G) = G'$, and let c be a generator of C. The group $V = G/C$ is elementary abelian, hence a vector space over GF(p). We have a well defined skew-symmetric bilinear form $f : V \times V \to \mathrm{GF}(p)$ induced by

$$[x, y] = c^{(Cx, Cy)f}.$$

If $(Cx, Cy)f = 0$ for all $y \in G$, then $x \in C$, thus f is nondegenerate. Thus there exists a decomposition $V = V_1 \oplus \cdots \oplus V_n$ where V_i is a 2-dimensional space with basis $\{u_i, v_i\}$, such that

$$\begin{aligned} (u_i, v_i)f &= 1, \\ (u_i, v_j)f &= 0 \text{ for } i \neq j, \\ (u_i, u_j)f &= 0, \\ (v_i, v_j)f &= 0. \end{aligned}$$

Write $u_i = Cx_i$, $v_i = Cy_i$. Then $G_i = \langle x_i, y_i \rangle$ is a nonabelian group of order p^3. We have that G is the central product of $G_1, \ldots G_n$. Clearly $G/C = G_1/C \times \cdots \times G_n/C$, hence $|G| = p^{2n+1}$.

Conversely, let G be the central product of $G_1, \ldots, G_n$, where each G_i is a nonabelian group of order p^3. Since $Z(G_i) \leq Z(G)$, it follows that $Z(G) = Z(G_i) \cong C_p$. Beside that, $[G_i, G_j] = 1$ for $i \neq j$, and $[G_i, G_i] = Z(G_i) = Z(G)$ for all i. Hence

$$[G, G] = [G_1 \cdots G_n, G_1 \cdots G_n] = Z(G),$$

therefore G is extraspecial. ∎

3.5.3 Enumeration of finite p-groups

It turns out that most of the finite groups are p-groups. The proof is beyond the scope of these notes. To illustrate this result, there are 49,910,529,484 different isomorphism

classes of groups of order at most 2000, and 49,487,365,422, or just over 99%, are groups of order 1024. We mention here that Phillip Hall proved that the number of isomorphism classes of groups of order p^n is

$$p^{\frac{2}{27}n^3+O(n^{8/3})}.$$

We will not prove this result. Instead we will derive some good upper and lower bounds on the number of finite p-groups of given order. We refer to [2] for a wealth of further estimates.

Preliminary results

Let r be a positive integer and F_r a free group on $\{x_1,\dots,x_r\}$. Denote

$$G_r = F_r/F_r^{p^2}\gamma_2(F_r)^p\gamma_3(F).$$

We identify x_i with their images in G_r, so $x_1,\dots,x_r$ generate G_r.

A finite p-group G is said to have Φ-*class 2* if there exists a central elementary abelian subgroup H of G such that G/H is elementary abelian. In other words, G is a central extension of an elementary abelian group by an elementary abelian group. Our first result shows that every group of Φ-class 2 is a homomorphic image of some G_r:

Lemma 3.5.28 *Let H be a group of Φ-class 2, and let $y_1,\dots,y_r \in H$. There is a homomorphism $\phi : G_r \to H$ such that $x_i^\phi = y_i$ for all $i = 1,\dots,r$.*

PROOF. As F_r is free there exists a unique homomorphism $F_r \to H$ with $x_i \mapsto y_i$. As $F_r^{p^2}\gamma_2(F_r)^p\gamma_3(F)$ is contained in the kernel of this map, we get the result. ∎

Lemma 3.5.29 *The group G_r is a finite p-group. The Frattini subgroup* $\mathrm{Frat}(G_r)$ *is central of order $p^{r(r+1)/2}$ and index p^r. Moreover, any automorphism $\alpha \in \mathrm{Aut}(G_r)$ that induces an identity mapping on $G_r/\mathrm{Frat}(G_r)$ fixes* $\mathrm{Frat}(G_r)$ *pointwise.*

PROOF.[Sketch of proof] The group $G_r^p\gamma_2(G_r)$ is a central elementary abelian p-subgroup of G_r, and the quotient by it is also elementary abelian. Thus G_r is a p-group. Observe that $\mathrm{Frat}(G_r)$ is generated by x_i^p and $[x_j,x_i]$, where $1 \le i < j \le r$. It is straightforward but technical to prove that this generating set is a minimal one, we skip the details. It follows that $\mathrm{Frat}(G_r)$ is central of order $p^{r(r+1)/2}$ and index p^r.

Now take $\alpha \in \mathrm{Aut}(G_r)$ that induces an identity mapping on $G_r/\mathrm{Frat}(G_r)$. So there exist $h_1,\dots,h_r \in \mathrm{Frat}(G_r)$ such that $x_i^\alpha = h_i x_i$. Since $\mathrm{Frat}(G_r)$ is central and $\mathrm{Frat}(G_r)^p = \{1\}$, we have

$$(x_i^p)^\alpha = (x_i^\alpha)^p = (h_i x_i)^p = h_i^p x_i^p = x_i^p$$

and

$$[x_j,x_i]^\alpha = [x_j^\alpha, x_i^\alpha] = [h_j x_j, h_i x_i] = [x_j,x_i].$$

Thus α fixes every generator of $\operatorname{Frat} G_r$ and we are done. ∎

Lemma 3.5.30 *Let N_1 and N_2 be subgroups of $\operatorname{Frat} G_r$. Then $G_r/N_1 \cong G_r/N_2$ if and only if there exists $\alpha \in \operatorname{Aut} G_r$ such that $N_1^\alpha = N_2$.*

PROOF. It is obvious that if there exists $\alpha \in \operatorname{Aut} G_r$ such that $N_1^\alpha = N_2$, then it induces an isomorphism $G_r/N_1 \to G_r/N_2$. Conversely, suppose there is an isomorphism $\alpha' \colon G_r/N_1 \to G_r/N_2$. Let $y_1, \ldots, y_r \in G$ be such that $(N_1 x_i)^{\alpha'} = N_2 y_i$. By Lemma 3.5.28 there exists a homomorphism $\alpha \colon G_r \to G_r$ with $x_i^\alpha = y_i$. Since α' is an isomorphism, $G_r = \langle y_1, \ldots, y_r \rangle N_2$. But $N_2 \leq \operatorname{Frat} G_r$, therefore $G_r = \langle y_1, \ldots, y_r \rangle$. Thus α is surjective. Since G_r is finite, this implies that α is an isomorphism. It remains to show that $N_1^\alpha = N_2$. By definition, $N_2 x^\alpha = (N_1 x)^{\alpha'}$ for all $x \in G_r$, and the result follows easily from here. ∎

A lower bound

A similar argument as in the proof of 3.2.9 shows the following:

Lemma 3.5.31 *Let V be a vector space over $GF(q)$ of dimension d. For $0 \leq k \leq d$, let $n_{k,d}$ be the number of subspaces of V of dimension k. Then*

$$n_{k,d} = \frac{(q^d-1)(q^d-q)\cdots(q^d-q^{k-1})}{(q^k-1)(q^k-q)\cdots(q^k-q^{k-1})}.$$

In particular, $q^{k(d-k)} \leq n_{k,d} \leq q^{k(d-k+1)}$.

Proposition 3.5.32 *Let r be a positive integer, and s an integer such that $1 \leq s \leq r(r+1)/2$. Then there are at least $p^{rs(r+1)/2-r^2-s^2}$ isomorphism classes of groups of order p^{r+s}.*

PROOF. Let G_r be as above. Let X be the set of subgroups $N \leq \operatorname{Frat} G_r$ of index p^s in $\operatorname{Frat} G_r$. Each $N \in X$ gives rise to a group G_r/N of order p^{r+s}. Furthermore, Lemma 3.5.30 implies that the set of isomorphism classes of these groups is in 1-1 correspondence with the set of orbits of $\operatorname{Aut} G_r$ acting on X.

Let $\theta \colon \operatorname{Aut} G_r \to \operatorname{Aut}(G_r/\operatorname{Frat} G_r)$ be the natural homomorphism. By Lemma 3.5.29 every $\alpha \in \ker\theta$ fixes $\operatorname{Frat} G_r$ pointwise and so acts trivially on X. Therefore $\ker\theta$ is contained in the stabilizer of every element of X, and so the length of any orbit of $\operatorname{Aut} G_r$ acting on X is at most

$$|\operatorname{Aut} G_r|/|\ker\theta| \leq |\operatorname{Aut}(G_r/\operatorname{Frat} G_r)| = |\operatorname{Aut} C_p^r| = |\operatorname{GL}(r,p)| \leq p^{r^2}.$$

From Lemma 3.5.31 we conclude that $|X| \geq p^{s(r(r+1)/2-s}$, therefore there are at least

$$p^{s(r(r+1)/2-s}/p^{r^2}$$

orbits of $\mathrm{Aut}\, G_r$ on X. This gives the desired bound. ∎

Proposition 3.5.32 yields roughly $p^{x^2yn^3/2}$ groups with Frattini subgroup of index p^{xn} and order p^{yn}. Maximizing the function $z = x^2y/2$ under the constraint $x+y=1$ yields the maximum value $z = 2/27$.

Theorem 3.5.33 *The number $f(p^n)$ of groups of order p^n is at least*

$$p^{\frac{2}{27}n^2(n-6)}.$$

PROOF. We may assume $n > 6$. Define $s = (n+2(n \mod 3))/3$ and $r = n-s$. Then Proposition 3.5.32 gives $f(p^n) \geq p^{rs(r+1)/2-r^2-s^2} \geq p^{2n^2(n-6)/27}$. ∎

An elementary upper bound

Let G be a group of order p^n and let

$$G = G_0 \geq G_1 \geq \cdots \geq G_{n-1} \geq G_n = \{1\}$$

be its chief series. For each i choose $g_i \in G_{i-1} - G_i$. Then every $g \in G$ may be written uniquely in *normal form* $g = g_1^{\alpha_1} \cdots g_n^{\alpha_n}$, where $\alpha_i \in \{0,1,\ldots,p-1\}$. Furthermore, $g \in G_i$ iff $\alpha_1 = \cdots = \alpha_i = 0$.

Observe that, given $1 \leq i < j \leq n$, we have that $g_i^p \in G_i$ and $[g_j, g_i] \in G_j$. Hence we may write these elements in normal form, that is,

$$g_i^p = g_{i+1}^{\beta_{i,i+1}} \cdots g_n^{\beta_{i,n}} \tag{3.1}$$

and

$$[g_j, g_i] = g_{j+1}^{\gamma_{i,j,j+1}} \cdots g_n^{\gamma_{i,j,n}} \tag{3.2}$$

for some $\beta_{i,j}, \gamma_{i,j,k} \in \{0,1,\ldots,p-1\}$. It is easy to see that the generators $g_1,\ldots,g_n$ and all the relations of the form (3.1) and (3.2) form a presentation for G (called a *power commutator presentation* or *polycyclic presentation*). One has to prove that a product of two elements in normal form can again be written in normal form. This can be done using *collection process* described in [9].

We remark that GAP calls the groups given by power-commutator presentations `pc groups`. Here is an example of how GAP prints out presentations of `pc groups`:

```
gap> PrintPcpPresentation(PcGroupToPcpGroup(DihedralGroup(16)));
g1^2 = id
g2^2 = g3
g3^2 = g4
g4^2 = id
g2 ^ g1 = g2 * g3 * g4
g3 ^ g1 = g3 * g4
```

Note that the conjugation relations can be rewritten into commutator ones using the identity $x^y = x[x,y]$, and that the trivial commutator relations are left out.

The above discussion leads to the following:

Theorem 3.5.34 *We have that*

$$f(p^n) \leq p^{\frac{1}{6}(n^3-n)}.$$

PROOF. Let G be as above. The isomorphism class of G is determined by the values of $\beta_{i,j}$ and $\gamma_{i,j,k}$. There are at most p choices for each of these $(n^3-n)/6$ elements, so there are at most $p^{\frac{1}{6}(n^3-n)}$ isomorphism classes of groups of order p^n. ∎

3.5.4 Coclass

As we have seen so far, there are many p-groups of given order, too many to classify them all up to isomorphism. In recent years there has been an idea to clasify p-groups according to coclass. This has lead to coclass theory which has produced some fascinating results. In this section we will briefly describe some of the main features of the theory, omitting almost all details. We refer to [7] for proofs and further results.

Let G be a group of order p^n. Then its nilpotency class c is strictly smaller of n by Proposition 3.5.22. The difference $n-c$ is called the *coclass* of G. Finite p-groups of coclass 1 are also known as *p-groups of maximal class*. An example of a p-group of maximal class is $C_p \wr C_p$; its order is p^{p+1} and the nilpotency class is precisely p (exercise).

Example 3.5.35 *Define*

$$Q_{2^n} = \langle x, y \mid y^{2^{n-1}} = 1, x^2 = y^{2^{n-2}}, y^x = y^{-1} \rangle$$

to be the generalized quaternion group of order 2^n (check that this is indeed its order). The group Q_8 is known as the quaternion group. Similarly, the group

$$SD_{2^n} = \langle x, y \mid y^{2^{n-1}} = 1, x^2 = 1, y^x = y^{2^{n-2}-1} \rangle$$

is said to be the semi-dihedral group of order 2^n. A classical result of the coclass theory is that 2-groups of maximal class are precisely dihedral, semi-dihedral, and generalized quaternion 2-groups.

The goal of coclass theory is to study common properties of finite p-groups of fixed coclass. To this purpose we study the so-called *coclass graph* $\mathscr{G}(p,r)$ whose vertices correspond to the isomorphism types of p-groups of coclass r. Two vertices G and H are joined by a directed edge from G to H if and only if $G \cong H/\gamma_c(H)$, where c is the nilpotency class of H. In order to understand this graph, we need a notion of pro-p-groups:

Definition 3.5.36 *A topological group G is a pro-p group if it is compact and has a basis of open neighborhoods of the identity consisting of normal subgroups of G of p-power index.*

Definition 3.5.37 *An inductively ordered set is a partially ordered set I with the property that for all $i, j \in I$ there exists $k \in I$ with $k > i$ and $k > j$. An inverse system of groups is a family $\{G_i \mid i \in I\}$ of groups, where I is an inductively ordered set, with surjections $\theta_{ij} : G_i \to G_j$ whenever $i > j$, satisfying $\theta_{ij}\theta_{jk} = \theta_{ik}$ for all $i > j > k$.*

Definition 3.5.38 *Let $\{G_i \mid i \in I\}$ be an inverse system of groups. The inverse limit of this system is*

$$\operatorname{proj\,lim} G_i = \left\{ (g_i) \in \prod_{i \in I} G_i \mid g_i\theta_{ij} = g_j \text{ for all } i > j \right\},$$

equipped with the product topology.

If G is a pro-p group and $\mathcal{N}$ the set of all normal subgroups of G of p-power index, then $\mathcal{Q} = \{G/N \mid N \in \mathcal{N}\}$ forms an inverse system, where the homomorphisms are the natural ones. We have that G is the inverse limit of $\mathcal{Q}$. This property in fact characterizes pro-p groups.

Definition 3.5.39 *If a group is an inverse limit of p-groups of coclass r, then it said to be a pro-p group of coclass r.*

It turns out [7] that every infinite pro-p group S of coclass r determines a *maximal coclass tree* $\mathcal{T}(S)$ in $\mathcal{G}(p,r)$, namely, the subtree of $\mathcal{G}(p,r)$ consisting of all descendants of $S/\gamma_i(S)$, where i is minimal such that $S/\gamma_i(S)$ has coclass r and $S/\gamma_i(S)$ is not a quotient of another infinite pro-p group R of coclass r not isomorphic to S.

In 1980, Leedham-Green and Newman posed five conjectures (A–E) about the stucture of the coclass graph. These are now all theorems [7]. We state them as follows:

Theorem 3.5.40

E *Given p and r, there are only finitely many isomorphism types of infinite solvable pro-p groups of coclass r.*

D *Given p and r, there are only finitely many isomorphism types of infinite pro-p groups of coclass r.*

C *Pro-p groups of finite coclass are solvable.*

B *For some function g, every finite p-group of coclass r has derived length bounded by $g(p,r)$.*

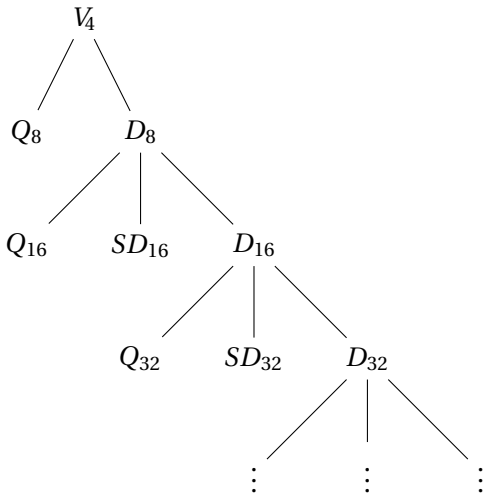
A *For some function f, every finite p-group of coclass r has a normal subgroup N of class 2 (1 if $p = 2$) whose index is bounded by $f(p,r)$.*

The coclass theorems in particular imply that $\mathscr{G}(p,r)$ consists of finitely many maximal coclass trees and finitely many groups lying outside these trees.

The next results shows that there is a certain kind of periodicity within coclass graphs. Let S be an infinite pro-p group of coclass r. The subtree $\mathscr{T}(S,k)$ of $\mathscr{T}(S)$ containing all groups of distance at most k from the main line is called a **shaved tree**. We denote its branches by $\mathscr{B}_j(S,k)$.

Theorem 3.5.41 (Theorem P (du Sautoy, 2001)) *Let S be an infinite pro-p group of coclass r. Then there exist integers $d = d(\mathscr{T}(S,k))$ and $f = f(\mathscr{T}(S,k))$ such that $\mathscr{B}_j(S,k)$ and $B_{j+d}(S,k)$ are isomorphic as rooted trees for all $j \geq f$.*

The simplest case are 2-groups of coclass 1. The graph $\mathscr{G}(2,1)$ has an isolated vertex C_4 and one infinite tree:



The periodicity in this tree is self-evident, even without shaving the tree.

3.5.5 Problems

1. Prove that the Pauli spin matrices
$$i = \begin{pmatrix} \sqrt{-1} & 0 \\ 0 & -\sqrt{-1} \end{pmatrix},\ j = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix},\ k = \begin{pmatrix} 0 & \sqrt{-1} \\ \sqrt{-1} & 0 \end{pmatrix}$$
generate a subgroup of $\mathrm{GL}(2,\mathbb{C})$ that is isomorphic to Q_8.
2. Let a group G be generated by $a_1,\ldots,a_d$. Show that $\gamma_i(G)$ is the normal closure in G of the set $\{[x_{j_1},\ldots,x_{j_i}] \mid 1 \leq j_k \leq i\}$.
3. Let $G = \langle a_1,\ldots,a_d\rangle$ be a nilpotent group. Then every element of G' can be written as $[x_1,a_1]\cdots[x_d,a_d]$ for some $x_1,\ldots,x_d \in G$.
4. Suppose that $G = HN'$, where $H \leq G$ and $N \triangleleft G$. Prove that $G = H\gamma_i(N)$ for all i.

5. Prove that the group $C_p \wr C_{p^n}$ is nilpotent of class precisely p^n.
6. Let G be a group of order p^n. If G has a unique subgroup of order p^m for all $1 < m < n$, prove that G is cyclic.
7. Let G be a group of order p^n, wher $n \geq 3$, and of maximal class. Prove the following:
   (a) G^{ab} is an elementary abelian p-group of order p^2 and $|\gamma_i(G) : \gamma_{i+1}(G)| = p$ for $2 \leq i \leq n-1$. The group G can be generated by two elements.
   (b) For every $i \geq 2$ we have that $\gamma_i(G)$ is the only normal subgroup of G of index p^i.
   (c) $Z_i(G) = \gamma_{n-i}(G)$ for all $i = 0, \ldots, n-1$.
8. Let G be a group in which $x^2 \in Z(G)$ for every $x \in G$. Prove the following:
   (a) G is nilpotent of class ≤ 2.
   (b) Every element of G' has order at most 2.
   (c) For all $x, y \in G$, the element $(xy)^2 y^{-2} x^{-2}$ belongs to G'.
   (d) For every $x, y \in G$ we have that $(xy)^4 = x^4 y^4$.
9. Let G be a metabelian group and $x, y, z, z_1, \ldots, z_n \in G$. Prove:
   (a) $[x, y, z_1, \ldots, z_n] = [x, y, z_{\pi(1)}, \ldots, z_{\pi(n)}]$ for every $\pi \in S_n$.
   (b) $[x, y, z][y, z, x][z, x, y] = 1$.
10. Let G be a group in which $x^3 = 1$ for all $x \in G$. Prove that $[x, y, y] = 1$ for all $x, y \in G$.
11. Let G be a finite group and F its Fitting subgroup.
    (a) Let N/F be an abelian normal subgroup of G/F such that $N \leq C_G(F)F$. Prove that $N = F(N \cap C_G(F))$.
    (b) Let N be as in (a). Prove that $N/(N \cap C_G(F))$ is nilpotent.
    (c) Let c be the nilpotency class of $N/(N \cap C_G(F))$, where N is as above. Show that N is nilpotent of class $\leq c+1$.
    (d) Conclude that $C_G(F)F/F$ contains no non-trivial abelian normal subgroup.
    (e) If G is solvable, show that $C_G(F) \leq F$.
12. Let G be a finite nilpotent group and N a non-trivial normal subgroup of G. Show the following:
    (a) $[N, G]$ is a proper subgroup of N.
    (b) Some maximal proper subgroup of N is normal in G.
    (c) Suppose that G is a p-group and M and N normal subgroups of G with $N < M$. Prove that there exists $K \triangleleft G$ such that $N \leq K < M$ and $|M : K| = p$.

13. Supply a proof of Lemma 3.5.31.
14. Use GAP to explore the number $f(m)$ of groups of order m for small m, and in the case when $m = p^n$ for small primes p and integers n.

3.6 References

[1] *Atlas of Finite Group Representations*, http://brauer.maths.qmul.ac.uk/Atlas/v3/.

[2] S. R. Blackburn, P. M. Neumann, and G. Venkataraman, *Enumeration of finite groups*, Cambridge University Press, Cambridge, 2007.

[3] K. S. Brown, *Cohomology of groups*, Springer-Verlag, New York, 1982.

[4] P. J. Cameron, *Notes on finite group theory*, October 2013.

[5] The GAP Group, GAP – Groups, Algorithms, and Programming, Version 4.7.4; 2014, (http://www.gap-system.org).

[6] I. M. Isaacs, *Finite group theory*. Graduate Studies in Mathematics, 92. American Mathematical Society, Providence, RI, 2008.

[7] C. R. Leedham-Green, and S. McKay, *The structure of groups of prime power order*, Oxford University Press, New York, 2002.

[8] D. J. S. Robinson, *A course in the theory of groups*, 2nd. ed., Springer-Verlag, New York, 1996.

[9] C. C. Sims, *Computation with finitely presented groups*, Cambridge University Press, Cambridge, 1994.

Chapter 4

Symmetric Key Cryptography and its Relation to Graph Theory

Enes Pasalic
University of Primorska, Slovenia

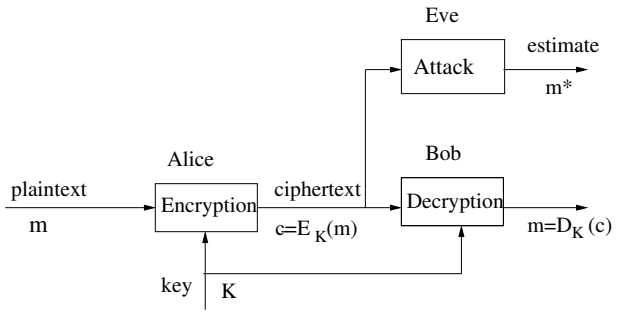
SUMMARY

Modern cryptology relies on many scientific disciplines such as information theory, probability theory, discrete mathematics among others. In addition, many public cryptosystems are based on some hard graph theoretic problems such as graph coloring for instance. While not directly derived from the concepts related to graphs, the most important cryptographic properties of certain discrete structures may be defined and analyzed in the graph theoretic framework which might give at least different insight at these structures. We will give a short survey of cryptography with the emphasis on these discrete structures being basic primitives in the so-called symmetric key cryptography. Booolean functions, vectorial mappings over finite structures and permutations over finite fields, as the most important representatives of these structures, will be considered in real-life encryption schemes. Their cryptographic properties will be stated both in a classical way using some suitable tools in cryptology and these will be then translated in the graph theoretic language. The students will also get a brief insight in the state-of-the-art research in this direction.

4.1 Introduction

A modern cryptology relies on many disciplines such as information theory, computer science, probability theory, number theory and abstract algebra. An information theoretical foundation of modern cryptology was established in the late forties. In his celebrated paper [9] from 1948 Claude E. Shannon laid the theoretical foundations of information theory. One of the greatest contribution of his work was a new concept of measuring the information. In his second work [10], among other important notion, Shannon introduced the concept of *u*nconditional security of symmetric ciphers. Unconditional security means that even if an adversary is assumed to have unlimited computational resources he still cannot defeat the cryptosystem. A necessary condition for a symmetric-key encryption scheme to be unconditionally secure is that the encryption key is at least as long as the message, which obviously restricts the practical use of such a system. Also, Shannon introduced two extremely important concepts which have been extensively used in design of modern ciphers, namely *c*onfusion and *d*iffusion.

A standard cryptosystem model used for achieving confidentiality (secrecy), also called symmetric-key cryptosystem transforms the plaintext message m into the ciphertext message c so that $c = E_K(m)$, where E_K denotes the encryption function, see Figure 4.1.



Model of a classic cryptosystem

Figure 4.1: Symmetric-key cryptosystem

The ciphertext message received by Bob is now supposed to be decrypted before reading. Equipped with the same key as Alice, Bob performs the following. He applies the decryption algorithm D_K on the encrypted message, i.e., $m = D_K(E_K(m))$ and retrieves the original message. The cryptanalyst Eve, not knowing the actual key K, may perform various attacks on the cryptosystem. The most trivial one, is called *e*xhaustive search which checks for all possible keys in the key space to decrypt the message.

As an example of an insecure symmetric-key cryptosystem we consider the Vigenère cipher. It is assumed that both the message and key symbols are letters from the English

alphabet, i.e., $\mathscr{M},\mathscr{K} \in \{A,B,\ldots,Z\}$. A sequence of message symbols $\mathbf{m} = m_0, m_1, \ldots$ is encrypted by this scheme into an encrypted sequence $\mathbf{c} = c_0, c_1, \ldots$ as follows. In order to express the encryption mathematically a simple transformation is performed, namely the letters are replaced by integers such that, $A \longleftrightarrow 0$, $B \longleftrightarrow 1, \ldots, Z \longleftrightarrow 25$. The same transformation is applied to the key $\mathbf{K} = K_0, K_1, \ldots, K_{l-1}$ and the corresponding message and key sequence are denoted $\mathbf{m}'$ and $\mathbf{K}'$, respectively. Then, the encrypted integer sequence $\mathbf{c}' = c'_0, c'_1, \ldots$ is obtained using,

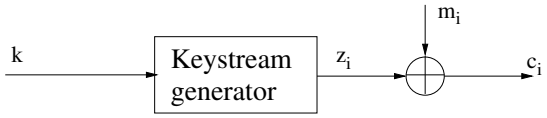
$$c'_i = m'_i + K'_{i \bmod l} \bmod 26, \;\; i = 0,1,2,\ldots. \tag{4.1}$$

Now the ciphertext $\mathbf{c}$ is derived from $\mathbf{c}'$ using the reverse transformation, $0 \longleftrightarrow A, 1 \longleftrightarrow B, \ldots, 25 \longleftrightarrow Z$. To recover the sequence of the original message, a similar transformation is applied to the encrypted sequence by the recipient,

$$m'_i = c'_i + (26 - K'_{i \bmod l}) \bmod 26, \;\; i = 0,1,2,\ldots.$$

Then the same transformation as above is applied to $\mathbf{m}'$ to retrieve the sequence of alphabetic letters $\mathbf{m}$.

Nevertheless, practical encryption schemes use more sophisticated approaches of implementing Shannon's concepts of confusion and diffusion. The encryption is rather performed on a bit level (or on a block of bits) by either "expanding" the secret key of finite length into a pseudo random sequence (running key sequence) z_i using keystream generator (*s*tream ciphers), see Figure 4.2.



General model of a binary additive stream cipher

Figure 4.2: Additive (binary) stream cipher

Alternatively, an encryption scheme can be designed by implementing a pseudo random permutations that substitutes a block of data (typically 128 bits) by a block of ciphertext bits of the same length (*b*lock ciphers) by repeating substitution (S) and permutation (P) through sevral rounds, see Figure 4.3.

In both cases an essential cryptographic primitive for embedding the concept of confusion is so-called Boolean function. Denoting by $\mathbb{F}_2$ the binary Galois field (thus $\mathbb{F}_2 = \{0,1\}$) and the n-dimensional vector space over $\mathbb{F}_2$ by $\mathbb{F}_2^n$, a Boolean function is defined as $f : \mathbb{F}_2^n \to \mathbb{F}_2$. A vectorial Boolean function $F : \mathbb{F}_2^n \mapsto \mathbb{F}_2^m$, also known as substitution box (S-box), is widely used primitive in the design of block ciphers. For instance, the

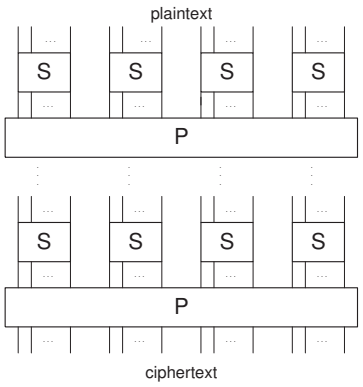


Figure 4.3: Substitution permutation network using S-boxes - a block cipher

S-boxes of DES (Data Encryption Standard) use $F : \mathbb{F}_2^6 \mapsto \mathbb{F}_2^4$, whereas the new standard AES (Advanced Encryption Standard) use $F : \mathbb{F}_2^8 \mapsto \mathbb{F}_2^8$. Since S-boxes are commonly the only nonlinear components of the block cipher, their design is crucial from the security point of view.

4.2 LFSR based stream ciphers and basic definitions

Stream ciphers which make use of a Boolean function are classically divided into two major groups: *n*onlinear combination generator and *n*onlinear filter generators, see Figure 4.4.

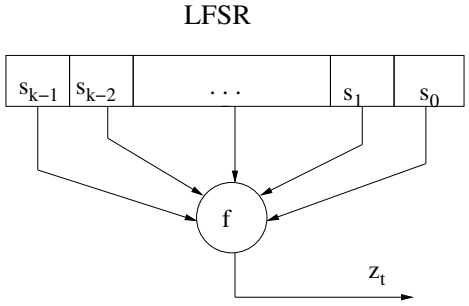


Figure 4.4: Nonlinear filtering generator

Both schemes have in common the use of a *l*inear feedback shift register (LFSR) as a main constituent block for producing sequences of large period. LFSRs are very well suited for hardware implementation and they can produce sequences with very good

statistical properties. In relation to Figure 4.5, the update procedure performed in any LFSR (at the time instance controlled by the system clock) may be summarized as follows:

1. The content of stage 0 is output and forms a part of the output sequence s_i, and at the same time the new content of stage $k-1$ is computed using a linear recursion $s_k = \sum_{i=0}^{k-1} s_i c^{k-i}$.
2. The content of stage i is moved to stage $i-1$, for each $1 \leq i \leq k-1$. The next state of the LFSR is therefore $S = (s_k, \ldots, s_1)$ seen from left to right in Figure 4.5.

For a given length of the LFSR, the period and statistical properties of the sequence depend entirely on the *c*onnection polynomial used. The use of a primitive connection polynomial $c(x) \in \mathbb{F}_2[x]$ results in the *s*equence of maximum length (the length is $2^L - 1$ for an LFSR of length L) with good statistical properties. Informally, a primitive polynomial $p(x) = a_0 + a_1 x + \ldots + a_k x^k$ of degree k can be defined as an irreducible polynomial over $\mathbb{F}_2$ with the property that $\{x^i \pmod{p(x)} : i = 0, \ldots, 2^k - 2\} = \mathbb{F}_2^k \setminus \{0\}$, using the representation $x^i \pmod{p(x)} = r(x) = r_0 + r_1 x + \ldots + r_{k-1} x^{k-1}$ and identifying $(r_0, \ldots, r_{k-1})$ with the elements of $\mathbb{F}_2^k$.

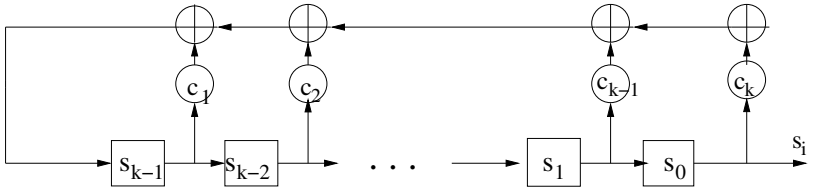


Figure 4.5: LFSR of length k with connection polynomial

Let s denote an infinite binary sequence whose terms are $s_0, s_1, \ldots$, whereas its truncated version of finite length n is denoted by s^n, that is, $s^n = s_0, s_1, \ldots, s_{n-1}$. The following definitions, taken from [6], will be useful in the sequel.

Definition 4.2.1 *An LFSR is said to generate a sequence s if there is some initial state of LFSR for which the output sequence of the LFSR is s. Similarly, an LFSR generates s^n if for some initial state the first n terms of the output sequence of the LFSR coincide with s^n.*

Definition 4.2.2 *The linear complexity of an infinite binary sequence s, denoted $L(s)$, is the length of the shortest LFSR that generates s.*

Example 4.2.3 *For $k = 4$ (or $L = 4$) and the primitive connection polynomial $C(x) = x^4 + x + 1$ if we start the LFSR with $S = (s_0, s_1, s_2, s_3) = (1,1,1,0)$ it produces the sequence*

$$1,1,1,0,1,0,1,1,0,0,1,0,0,0,1|1,1,1,0\ldots$$

The sequence is of maximum length $15 = 2^4 - 1$ *and contains exactly* $2^{k-1} = 8$ *ones and* $2^{k-1} - 1$ *zeros, why ? Check what happens if we use irreducible polynomial* $C(x) = x^4 + x^3 + x^2 + x + 1$ *!*

However, any sequence generated by a finite-state machine has a finite linear complexity. Moreover, due to Elwyn R. Berlekamp and James L. Massey [5], there exists an efficient polynomial-time synthesis algorithm, which computes the linear complexity of a given binary sequence. When the length L of LFSR is known then a sequence of length $2L$ is required to compute the connection polynomial, either using the Berlekamp-Massey algorithm or a direct matrix equation. If L is not known, then the Berlekamp-Massey algorithm can be used to determine L and the connection polynomial. In either case the adversary must obtain a subsequence of length $2L$.

In reference to Figure 4.2, we assume that an adversary mounts a known or chosen-plaintext attack on additive binary stream cipher where the running-key generator is implemented by using an LFSR. Then the adversary can obtain the subsequence of $\mathbf{z}$ of length L, by computing $z_i = m_i \oplus c_i$, $i = 0, \ldots, L-1$ (since m_i are known). Then, an LFSR of length L, with the connection polynomial computed with the Berlekamp-Massey algorithm, can be initialized with this subsequence to generate the remainder of the sequence $\mathbf{z}$.

Thus, a necessary but not sufficient condition for any keystream generator is the requirement for a large linear complexity. This cannot be achieved using a single LFSR, and general methods for destroying the linear properties of LFSRs are:

- using a *n*onlinear combining function at the outputs of several LFSRs;
- using a *n*onlinear filtering function on the contents of a single LFSR;and
- using the output of one/several LFSRs to control clocking of one/several LFSRs.

As mentioned earlier the first two methods take advantage of a Boolean function to introduce the nonlinearity to the keystream. A general construction of a nonlinear combination generator is illustrated in Figure 4.6, where for the sake of generality we consider $F : \mathbb{F}_2^n \to \mathbb{F}_2^m$, for $m \geq 1$.

In this set up the outputs of n LFSRs, $x^{(1)}, \ldots, x^{(n)}$ are used as the inputs to a nonlinear vectorial Boolean function, denoted F, and the keystream sequence is then generated by this function. More formally, $z_i \stackrel{\triangle}{=} f_i(x_i^{(1)}, \ldots, x_i^{(n)})$, and the function $F : \mathbb{F}_2^n \mapsto \mathbb{F}_2^m$ (actually an S-box) is a collection of m Boolean functions $F = (f_1, \ldots, f_m)$. A Boolean function $f(x_1, \ldots, x_n)$ can be represented as the output column of its *t*ruth table f, i.e., a binary string of length 2^n, $f = [f(0,0,\cdots,0), f(1,0,\cdots,0), f(0,1,\cdots,0), \ldots, f(1,1,\cdots,1)]$.

The truth table representation may be suitable for Boolean function in small number of variables. Thus, for moderate to large values of n, $f \in \mathscr{B}_n$ is usually represented by its

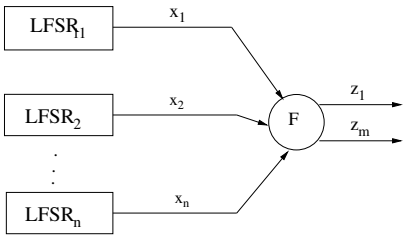


Figure 4.6: Nonlinear combination generator

*a*lgebraic normal form (ANF):[1]

$$f(x_1,\ldots,x_n)=\sum_{u\in\mathbb{F}_2^n}\lambda_u\left(\prod_{i=1}^{n}x_i^{u_i}\right),\ \ \lambda_u\in\mathbb{F}_2\,,u=(u_1,\ldots,u_n). \tag{4.2}$$

There are 2^n different terms $x_1^{u_1}x_2^{u_2}\cdots x_n^{u_n}$ for different u's. As λ_u is binary it gives $\#\mathscr{B}_n=2^{2^n}$ different functions in n variables $x_1,\ldots,x_n$ (denoting by $\mathscr{B}_n$ the set of all Boolean functions in n variables), implying that a search for "good" functions becomes infeasible already for $n=6$!

Example 4.2.4 *For $n=3$ there are $2^8=256$ distinct functions specified by λ_u,*

$$\mathscr{B}_3=\{\lambda_0 1\oplus\lambda_1x_1\oplus\lambda_2x_2\oplus\lambda_3x_3\oplus\lambda_4x_1x_2\oplus\lambda_5x_1x_3\oplus\lambda_6x_2x_3\oplus\lambda_7x_1x_2x_3\}.$$

The *a*lgebraic degree of f, denoted by $deg(f)$ or sometimes simply d, is the maximal value of the Hamming weight of u such that $\lambda_u\neq 0$. There is a one-to-one correspondence between the truth table and the ANF via so called inversion formulae.

x_3	x_2	x_1	$f(x)$
0	0	0	0
0	0	1	0
0	1	0	0
0	1	1	1
1	0	0	1
1	0	1	1
1	1	0	0
1	1	1	1

The truth table of the Boolean function $f(x_1,x_2,x_3)=x_1x_2+x_2x_3+x_3$.

The easiest way to obtain the ANF from the truth table (without involving Möbius transform) is to expand the ANF of f when $f(x)=1$ and add these together. For the above example we have:

$$f(x)=x_1x_2(1+x_3)+(1+x_1)(1+x_2)x_3+x_1(1+x_2)x_3+x_1x_2x_3=x_1x_2+x_2x_3+x_3,$$

[1]Addition operator over $\mathbb{F}_2$ denoted by "$\oplus$" is often replaced with usual addition operator "+".

after cancelling identical terms. A *b*alanced Boolean function has equally many zeros and ones in its truth table, i.e., $\{f(x)=0 : x \in \mathbb{F}_2^n\} = \{f(x)=1 : x \in \mathbb{F}_2^n\} = 2^{n-1}$. What can be said about the upper bound on degree of balanced Boolean functions in $\mathscr{B}_n$ then ?

The reason why we require a high algebraic degree is related to the following attack scenario. Recall that the basic goal of the attacker is to recover the secret state bits located in LFSR. Since both LFSR, its connection polynomial $c(x)$, the filtering function $f(x)$ and a portion of the output keystream sequence (known-plaintext attack) are known we have the following. At each time instance the known keystream bit $z_i^t = f(x_1^t, \ldots, x_n^t)$, where the time dependency of the inputs to f is due to the structure of LFSR. Anyway, any x_i^t is a linear function of the initial secret state bits $s_0, \ldots, s_{L-1}$, say $x_i^t = \sum_{j=0}^{L-1} a_j^{(i,t)} s_j$, due to the linear update function of LFSR. Thus given f of degree d, whose ANF contains at most $T = \binom{n}{0} + \binom{n}{1} + \ldots + \binom{n}{d}$ terms, we get one equation of degree d is secret state bits. Using so-called linearization we can introduce (at most) T new variables in $s_0, \ldots, s_{L-1}$ and solve a linear system with respect to unknown and secret s_i. Since there are L secret state variables after the above substitution our linear system has at most $T' = \binom{L}{0} + \binom{L}{1} + \ldots + \binom{L}{d}$ terms. The complexity of solving a linear system of size $\approx \binom{L}{d}$ is of order $(\binom{L}{d})^3$ using Gauss elimination. Therefore, a large d is desirable but the implementation cost increases !

Assume now, that n maximum-length LFSRs as in Figure 4.6, whose lengths

$$L_1, L_2, \ldots, L_n$$

are relatively prime, are combined by a nonlinear Boolean function $f(x_1, \ldots, x_n)$. Then the linear complexity of the keystream sequence z is $f(L_1, \ldots, L_n)$, where the expression is computed over the integers [6,12]. Since this expression is directly dependent on the degree of f, then obviously a large linear complexity of the keystream is obtained by functions of high degree.

Example 4.2.5 *(Geffe generator) Assume that the lengths of LFSRs are relatively prime for the scheme in Figure 4.6, with $n=3$. Let the nonlinear combining function be $f(x_1,x_2,x_3) = x_1x_2 \oplus x_2x_3 \oplus x_3$. The function f is obviously of degree 2. The Geffe generator is cryptographically weak because the information about the states of $LFSR_1$ and $LFSR_3$ leaks to the output. For fixed $x_3=0$ the output is x_1x_2 and therefore 75% zeros and 25% of ones are outputted in this case.*

The observation in the above example leads to another important criteria for Boolean functions used as a nonlinear combining function, which is the concept of *c*orrelation immunity.

Definition 4.2.6 [11] *Let $x_1, x_2, \ldots, x_n$ be a set of independent uniformly distributed binary random variables. A Boolean function $f(x_1, x_2, \ldots, x_n)$ is called m th order correlation*

immune if for each subset of at most m input variables $x_{i_1},\ldots,x_{i_k}$, $1 \le i_1 \cdots \le i_k \le n$, $k \le m$, the mutual information between the keystream $z = f(x_1,\ldots,x_n)$ and the subset $x_{i_1},\ldots,x_{i_k}$ is equal to zero, i.e. $I(z;x_{i_1},\ldots,x_{i_k}) = 0$. Expressed in terms of probability we have that

$$Prob(x_{i_1} \oplus x_{i_2} \cdots \oplus x_{i_k} = z) = \frac{1}{2}, \quad z \in \mathbb{F}_2, \quad \textit{for any } k = 1,\ldots,m.$$

Another important measure of cryptographical strength of Boolean functions is *nonlinearity*. The nonlinearity of f, denoted by $\mathcal{N}_f$, is defined to be the minimum Hamming distance [2] to the set of affine functions. For an n-input variable function the set of affine functions is given as $\mathcal{A}_n = \{a_1x_1 \oplus \cdots \oplus a_nx_n \oplus b,\ a \in \mathbb{F}_2^n;\ b \in \mathbb{F}_2\}$. The set of all n variable linear functions, when $b = 0$, is denoted by $\mathcal{L}_n$. Thus, the nonlinearity of f is given by,

$$\mathcal{N}_f = \min_{g \in \mathcal{A}_n} d_H(f,g). \tag{4.3}$$

Prof. James Massey formulated it nicely once upon a time " The linearity is the curse of the cryptographer". Any cryptographic primitive somehow implements Shannon's concept of confusion which for our scheme (almost) directly corresponds to nonlinearity.

The linear functions will be represented by means of the scalar (inner) product,

$$\varphi_\alpha : x \in \mathbb{F}_2^n \longmapsto \alpha \cdot x = \sum_{i=1}^{n} \alpha_i x_i.$$

Definition 4.2.7 *A t-th order correlation immune function Boolean function f which is balanced is called a t-resilient function.*

The properties of Boolean functions are most comprehensibly viewed through the *W*alsh transform.

Definition 4.2.8 *The Walsh transform of $f \in \mathcal{B}_n$ in point $\alpha \in \mathbb{F}_2^n$ is denoted by $\mathcal{F}(f+\varphi_\alpha)$ and calculated as,*

$$\alpha \in \mathbb{F}_2^n \longmapsto \mathcal{F}(f+\varphi_\alpha) = W_f(\alpha) = \sum_{x \in \mathbb{F}_2^n} (-1)^{f(x)+\varphi_\alpha(x)}. \tag{4.4}$$

The values of these coefficients form the *W*alsh-spectrum of f, and clearly f is balanced if and only if $W_f(0) = 0$. Notice that $\varphi_\alpha(x) = \alpha \cdot x$ uniquely identifies one linear function, see also relation (4.5).

Exercise 4.2.9 *Show that the Hamming distance between a Boolean function $f(x)$ and an affine function $g(x) = \alpha \cdot x + b$ ($\alpha \in \mathbb{F}_2^n$ and $b \in \mathbb{F}_2$), can be calculated via the Walsh transform as $d_H(f,g) = 2^{n-1} - \frac{(-1)^b \mathcal{F}(f+\varphi_\alpha)}{2}$.*

[2] The Hamming distance between two binary strings of the same length, say f and g, is the number of positions where these strings differ, i.e., $d_H(f,g) = \#\{x | f(x) \neq g(x)\}$.

A closely related concept, known as the Hadamard transform and denoted by W_f^H, simply uses the values $f(x)$ instead of $(-1)^{f(x)}$, that is $W_f^H(\alpha)=\sum_{x\in\mathbb{F}_2^n} f(x)(-1)^{\varphi_\alpha(x)}$. A simple relationship between the two transforms is given as an exercise.

Exercise 4.2.10 *Show that $W_f(\alpha)=-2W_f^H(\alpha)+2^n\Delta(\alpha)$ for any $\alpha\in\mathbb{F}_2^n$, where $\Delta(\alpha)=1$ if $\alpha=0$, and zero otherwise.*

The values of Walsh and Hadamard spectra of $f\in\mathscr{B}_n$ are easily obtained through $W_f=H_nf^T$, respectively, $W_f^H=H_n(-1^f)^T$, where f^T denotes the transpose of the truth table of f and H_n is the Hadamard matrix of size $2^n\times 2^n$ defined recursively,

$$H_1=\begin{pmatrix}1 & 1\\ 1 & -1\end{pmatrix},\quad H_n=\begin{pmatrix}H_{n-1} & H_{n-1}\\ H_{n-1} & -H_{n-1}\end{pmatrix}.$$

It is easy to show that $HH^T=2^nI$ and also $H^TH=2^nI$, where I is the identity matrix whose diagonal elements are ones.

The nonlinearity of $f(x)$ can be obtained via the Walsh transform as,

$$\mathscr{N}_f=2^{n-1}-\frac{1}{2}\max_{\alpha\in\mathbb{F}_2^n}|\mathscr{F}(f+\varphi_\alpha)|. \tag{4.5}$$

Lemma 4.2.11 [13] *Let $f\in\mathscr{B}_n$ and let t be some positive integer. The function f is said to be correlation immune (CI) of order t if and only if $\mathscr{F}(f+\varphi_\alpha)=0$ for any $a\in\mathbb{F}_2^n$ such that $1\le wt(\alpha)\le t$.*

An important property of the Walsh spectra, referred to as Parseval's equality [4], states that for any Boolean function $f\in\mathscr{B}_n$, $\sum_{\alpha\in\mathbb{F}_2^n}\mathscr{F}^2(f+\varphi_\alpha)=2^{2n}$.

Exercise 4.2.12 *Use a similar technique as in the proof of Proposition 4.2.14 to show Parseval's equality. Consider the sum $\sum_{u\in\mathbb{F}_2^n} W_f(u)W_f(u\oplus v)$ and show it is 2^{2n} if $v=0$ and zero otherwise.*

We illustrate the cryptographic criteria discussed above with a detailed examination of the nonlinear combining function used in the Geffe generator, see also Example 4.2.5.

Example 4.2.13 *Consider the function $f(x_1,x_2,x_3)=x_1x_2+x_2x_3+x_3$ used in the Geffe generator. The truth table and the Walsh spectra are given in Table 4.1. Note that the linear functions φ_α are determined by x values. For instance the entry $(x_1,x_2,x_3)=(1,0,0)$ will yield $\varphi_\alpha=(x_1,x_2,x_3)\cdot(1,0,0)=x_1$. Then, the nonlinearity $\mathscr{N}_f=2^{n-1}-\frac{1}{2}\max_{\alpha\in\mathbb{F}_2^n}|\mathscr{F}(f+\varphi_\alpha)|=2$. The function is balanced but not correlation immune since $\mathscr{F}(f+x_1)=\mathscr{F}(f+x_3)\neq 0$.*

Notice that the Walsh spectra, constrained by Parseval's equality, is integer valued and obviously we cannot design cryptographically strong Boolean functions by specifying

x_1	x_2	x_3	$f(x)$	$\mathscr{F}(f+\varphi_\alpha)$
0	0	0	0	0
0	0	1	0	-4
0	1	0	0	0
0	1	1	1	-4
1	0	0	1	-4
1	0	1	1	0
1	1	0	0	4
1	1	1	1	0

Table 4.1: The truth table and the Walsh spectra of the Boolean function $f(x_1,x_2,x_3) = x_1x_2 + x_2x_3 + x_3$.

the values (placing zeros and controlling maximum values) in the Walsh spectra (even though Parseval's equality is satisfied). This means that the Boolean space is only a small subspace of a more general mapping from $\mathbb{Z}^n$ to $\mathbb{Z}$.

Proposition 4.2.14 *Given the Walsh spectra $\{W_f(\alpha)\}$ of $f \in \mathscr{B}_n$ the inverse Walsh transformation can be computed as,*

$$(-1)^{f(x)} = 2^{-n} \sum_{\alpha \in \mathbb{F}_2^n} W_f(\alpha)(-1)^{\alpha \cdot x} \quad \text{for all } x \in \mathbb{F}_2^n. \tag{4.6}$$

PROOF. Let us substitute $W_f(\alpha) = \sum_{y \in \mathbb{F}_2^n} (-1)^{f(y)+\alpha \cdot y}$ in $f(x)$ so that,

$$\begin{aligned} \sum_{\alpha \in \mathbb{F}_2^n} W_f(\alpha)(-1)^{\alpha \cdot x} &= \sum_{\alpha \in \mathbb{F}_2^n} \sum_{y \in \mathbb{F}_2^n} (-1)^{f(y)+\alpha \cdot y}(-1)^{\alpha \cdot x} \\ &= \sum_{y \in \mathbb{F}_2^n} (-1)^{f(y)} \sum_{\alpha \in \mathbb{F}_2^n} (-1)^{\alpha \cdot (x+y)} \\ &= 2^n (-1)^{f(x)}, \end{aligned}$$

since since the sum $\sum_{\alpha \in \mathbb{F}_2^n} (-1)^{\alpha \cdot (x+y)}$ is equal to zero unless $x = y$ in which case it is equal to 2^n. The statement follows. ∎

A special class of functions achieving the upper bound on nonlinearity is known as *b*ent functions. They exist only for even n and have a uniform spectra, that is, f is bent if and only if $W_f(\alpha) = \pm 2^{n/2}$, for all $\alpha \in \mathbb{F}_2^n$. It is easily understood that since $\sum_{\alpha \in \mathbb{F}_2^n} W_f(\alpha)^2 = 2^{2n}$, then $\{W_f(\alpha)$ is minimized with respect to its maximum absolute value if the spectra is flat. These functions are not balanced however, since $W_f(0) = \pm 2^{n/2}$, but they posses many other desirable properties and have several connections to difference sets, Kerdock codes, symmetric design etc. (their modified balanced versions are also used in symmetric key primitives). Bent functions correspond to strongly distance regular Cayley graphs, this connection is discussed later.

For any bent function f one may define its dual $\tilde{f}$ as $(-1)^{\tilde{f}(x)} = 2^{-n/2} W_f(x)$ for all $x \in \mathbb{F}_2^n$.

Proposition 4.2.15 *The dual bent function $\tilde{f}$ of a bent function f is again bent.*

PROOF. If f is bent the inverse Walsh transform gives, $(-1)^{f(x)} = 2^{-n} \sum_{\alpha \in \mathbb{F}_2^n} W_f(\alpha)(-1)^{\alpha \cdot x}$, for all $x \in \mathbb{F}_2^n$. Replacing $W_f(\alpha) = 2^{n/2}(-1)^{\tilde{f}(\alpha)}$ from the definition of $\tilde{f}$, we get

$$2^{n/2}(-1)^{f(x)} = \sum_{\alpha \in \mathbb{F}_2^n} (-1)^{\tilde{f}(\alpha)} (-1)^{\alpha \cdot x} = \sum_{\alpha \in \mathbb{F}_2^n} (-1)^{\tilde{f} + \alpha \cdot x} = W_{\tilde{f}}(\alpha),$$

thus $W_{\tilde{f}}(\alpha) \in \{-2^{n/2}, 2^{n/2}\}$ and $\tilde{f}$ is bent. ∎

One class of bent functions of particular importance, known as the Maiorana-McFarland class, is specified as follows. Let us, for $n = 2k$, identify $\mathbb{F}_2^n$ with $\mathbb{F}_2^k \times \mathbb{F}_2^k$. Suppose $\pi : \mathbb{F}_2^k \to \mathbb{F}_2^k$ is a permutation and $g \in \mathscr{B}_k$. A function $f : \mathbb{F}_2^k \times \mathbb{F}_2^k \to \mathbb{F}_2$ defined by

$$f(x, y) = x \cdot \pi(y) + g(y), \text{ for all } x, y \in \mathbb{F}_2^k, \tag{4.7}$$

is a bent function and this class is denoted as $\mathscr{M}$.

Proposition 4.2.16 *The function f defined by (4.7) is a bent function.*

PROOF. The Walsh transform at $(a, b) \in \mathbb{F}_2^k \times \mathbb{F}_2^k$ equals to:

$$W_f(a, b) = \sum_{x \in \mathbb{F}_2^k} \sum_{y \in \mathbb{F}_2^k} (-1)^{f(x,y) + (a,b) \cdot (x,y)} = \sum_{y \in \mathbb{F}_2^k} (-1)^{g(y) + b \cdot y} \sum_{x \in \mathbb{F}_2^k} (-1)^{x \cdot \pi(y) + a \cdot x}.$$

For any fixed y the sum $\sum_{x \in \mathbb{F}_2^k} (-1)^{x \cdot \pi(y) + a \cdot x} = \sum_{x \in \mathbb{F}_2^k} (-1)^{x \cdot (\pi(y) + a)} = 0$, unless $\pi(y) = a$ which happens exactly for one $y = \pi^{-1}(a)$. In the case $\pi(y) = a$ the sum

$$\sum_{x \in \mathbb{F}_2^k} (-1)^{x \cdot (\pi(y) + a)} = 2^k,$$

and therefore $W_f(a, b) = 2^k (-1)^{g(\pi^{-1}(a)) + b \cdot \pi^{-1}(a)}$, thus f is bent. ∎

Notice that the class $\mathscr{M}$ contains as a subclass a class of bent functions, but it can also generate resilient functions with high nonlinearity. To see this we modify the above definition as follows,

Definition 4.2.17 *For any positive integers p, q such that $n = p + q$, a function $f \in \mathscr{B}_n$ in the Maiorana McFarland class is defined by*

$$f(x, y) = \phi(y) \cdot x \oplus g(y), \quad x \in \mathbb{F}^p, \; y \in \mathbb{F}^q, \tag{4.8}$$

where ϕ is any mapping from $\mathbb{F}^q$ to $\mathbb{F}^p$, $g \in \mathscr{B}_q$ is arbitrary.

Proposition 4.2.18 *Let f be defined as above and for $p > q$ assume that π is injective. Then, $N_f = 2^{n-1} - 2^{p-1}$. In addition, if $wt(\phi(y)) \geq t+1$ for all $y \in \mathbb{F}_2^q$ then f is t-resilient.*

PROOF. Let $\mathbb{F}_2^n = \mathbb{F}_2^p \times \mathbb{F}_2^q$. All we have to do is to show that $\max_{(a,b)\in\mathbb{F}_2^p\times\mathbb{F}_2^q} \mid W_f(a,b) \mid = 2^p$. We have,

$$W_f(a,b) = \sum_{y\in\mathbb{F}_2^q}\sum_{x\in\mathbb{F}_2^p}(-1)f(x,y)+(a,b)\cdot(x,y) = \sum_{y\in\mathbb{F}_2^q}(-1)^{g(y)+b\cdot y}\sum_{x\in\mathbb{F}_2^p}(-1)\phi(y)\cdot x + a\cdot x.$$

Then again, for any fixed $y \in \mathbb{F}_2^q$ the sum $\sum_{x\in\mathbb{F}_2^p}(-1)\phi(y)\cdot x + a\cdot x = 0$, unless $\pi(y) = a$. Since π is injective then $\#\{y \in \mathbb{F}_2^q : \pi(y) = a\}$ is either 0 or 1. In the case $\pi(y) = a$ we have $\sum_{x\in\mathbb{F}_2^p}(-1)\phi(y)\cdot x + a\cdot x = 2^p$, and the first part follows. The second part is left as an exercise. ∎

Example 4.2.19 *Let $n = 6$, $p = 4$, $q = 2$ and $(x,y) \in \mathbb{F}_2^4 \times \mathbb{F}_2^2$. Define injective $\pi : \mathbb{F}_2^2 \to \mathbb{F}_2^4$ as $\pi(00) = (1100)$, $\pi(10) = (0110)$, $\pi(01) = (1010)$, $\pi(1) = (10011)$. Then, for any fixed y the function $f(x,y)$ is a linear function in $x_1,\ldots,x_4$. More precisely, $f(x,00) = x_1 + x_2$, $f(x,10) = x_2+x_3$, $f(x,01) = x_1+x_3$, $f(x,11) = x_1+x_3+x_4$. Then f is 1-resilient, $deg(f) = 3$ (check this !), and $\mathcal{N}_f = 24$.*

More advanced construction methods are not treated here due to their tedious representation. The currently best known methods are given recently by Pasalic and Zhang based on the use of disjoint linear codes (resilient S-boxes) and a subtle modification of the Maiorana-McFarland construction for resilient Boolean functions.

4.3 Equivalence classes of Boolean functions

The group of all invertible $\mathbb{F}_2$-linear transformations on $\mathbb{V}_n$ is denoted by $GL(\mathbb{V}_n)$.

Definition 4.3.1 *Two Boolean functions $f, g \in \mathcal{B}_n$ are said to be affine equivalent if and only if there exist $A \in GL(\mathbb{V}_n)$ and $b \in \mathbb{V}_n$ such that*

$$g(x) = f(Ax+b) \text{ for all } x \in \mathbb{V}_n. \tag{4.9}$$

The affine general linear group $AGL(\mathbb{V}_n)$ consists of all the element of the form (A,b). It can be verified that the transformation $f(x) \mapsto f(Ax+b)$ is a group action of $AGL(\mathbb{V}_n)$ on $\mathcal{B}_n$.

Definition 4.3.2 *Two Boolean functions $f, g \in \mathcal{B}_n$ are said to be extended affine equivalent (EA-equivalent, or, equivalent) if and only if apart from A and b as above there exist $\mu \in \mathbb{V}_n$ and $\epsilon \in \mathbb{F}_2$ such that*

$$g(x) = f(Ax+b) + \mu\cdot x + \epsilon \text{ for all } x \in \mathbb{F}_2^n. \tag{4.10}$$

Given any two Boolean functions $f, g \in \mathscr{B}_n$ deciding whether they are EA-equivalent or not is an important open problem. A direct verification requires a search over all the elements of $AGL(\mathbb{V}_n)$ and therefore its computational complexity is $O(2^{n^2})$. Since an exhaustive search over all the elements of $AGL(\mathbb{V}_n)$ is not feasible for $n \geq 7$, the decision problem involving equivalence of Boolean functions is attempted by using carefully chosen invariants. Algebraic degree of a non-affine Boolean function is an invariant with respect to affine transformations and addition of affine functions. Therefore, two Boolean functions with algebraic degree greater than or equal to 2 are EA-inequivalent if their algebraic degrees are different. It is well known [3] that the absolute Walsh spectra of any Boolean function f are invariants with respect to the action of $AGL(\mathbb{V}_n)$ and the addition by an affine function. Unfortunately these invariants are not useful to determine affine inequivalence of Boolean functions having the same algebraic degree and absolute Walsh spectra. The problem of classifying Boolean functions and bent functions in particular seems to be elusive.

Open Problem 4.3.3 *Find new classes of bent functions by proving their affine non-equivalence to already known classes. The problem may also be viewed in terms of suitable subgroups of permutations of the Walsh spectra. Indeed, since the dual bent function is also bent it implies that either* $\{\alpha : W_f(\alpha) = 2^{n/2}\} = 2^{n-1} - 2^{n/2-1}$ *and* $\{\alpha : W_f(\alpha) = -2^{n/2}\} = 2^{n-1} + 2^{n/2-1}$ *or vice versa. This is also related to a group action on the (multi)set of the Walsh spectra.*

Open Problem 4.3.4 *A related concept to the above is so-called algebraic thickness which refers to the most compact representation of a function by its ANF. For instance, the function* $f(x_1, \ldots, x_n) = x_1 x_2 \cdots x_n$ *(which is cryptographically disastrous, why ?) is obviously affine equivalent to the function* $f(x_1, \ldots, x_n) = (x_1+1)(x_2+2)\cdots(x_n+1)$. *While the former contains a single term in its ANF, the latter contains all possible* 2^n *terms in its ANF. Of course, if we would implement such a function we would prefer the former one. Given any function* $f \in \mathscr{B}_n$ *find efficiently its affine equivalent containing the least number of ANF terms !*

4.4 Vectorial Boolean functions - substitution boxes

The *n*onlinearity of $F = (f_1, f_2, \ldots, f_m)$, denoted by N_F, is defined as the minimum among the nonlinearities of all nonzero linear combinations of the component functions of F, i.e.,

$$nl(F) = \min_{\tau \in \mathbb{F}_2^{m*}} nl(\sum_{j=1}^{m} \tau_j f_j(x)), \text{ where } \tau = (\tau_1, \ldots, \tau_m) \in \mathbb{F}_2^{m*}. \tag{4.11}$$

The algebraic degree of F is defined as the minimum of degrees of all nonzero linear combinations of the component functions of F, namely,

$$deg(F) = \min_{\tau \in \mathbb{F}_2^{m*}} deg(\sum_{j=1}^{m} \tau_j f_j(x)). \tag{4.12}$$

The two measures defined above in terms of linear combinations of the component functions obviously make the design of cryptographically strong vectorial Boolean functions much harder than in the Boolean case. In certain situations one may use additional algebraic structures in those cases such structures are available, but usually one prefer to involve the structure of finite fields and to consider mappings F over $\mathbb{F}_{2^n}$ so that isomorphically $F : \mathbb{F}_2^n \to \mathbb{F}_2^n$ is equivalent to $F : \mathbb{F}_{2^n} \to \mathbb{F}_{2^n}$ (once the basis of the finite field is fixed).

Example 4.4.1 *Consider the mapping F over $\mathbb{F}_{2^n}$, for n odd, given as a polynomial $F(x) = x^3$, thus $\mathbb{F}_{2^n} \ni x \mapsto x^3 \in \mathbb{F}_{2^n}$. Since* $\gcd(3, 2^n) = 1$ *for odd k, F is a permutation. Furthermore, $N_F = 2^{n-1} - 2^{\frac{n-1}{2}}$ which is exceptionally high nonlinearity and such functions are called almost bent (AB) for this reason. The mapping x^{2^k+1} is also known as Gold mapping, when* $\gcd(k, n) = 1$.

Another important property of substitution boxes is their differential table. Actually, this property of having low uniformity of differentials is of the same importance as nonlinearity in the design of S-boxes since it leads to differential cryptanalysis which is one of the most powerful cryptanalytic tools.

Definition 4.4.2 *Let F be an (n, m) S-box, that is $F : \mathbb{F}_2^n \to \mathbb{F}_2^m$. For any $a \in \mathbb{F}^n$ and $b \in \mathbb{F}^m$, we denote*

$$\delta_F(a, b) = \#\{x \in \mathbb{F}^n,\ F(X_n + a) + F(X_n) = b\} \tag{4.13}$$

where $\#S$ is the cardinality of any set S. We define

$$\delta(F) = \max_{a \neq 0, b \in \mathbb{F}^n} \delta_F(a, b). \tag{4.14}$$

The smaller the $\delta(F)$, the better the differential properties of F.

The above definition is more generally stated in terms of vector space mappings, since when $m \nmid n$ where is no corresponding finite field representation. In the Boolean case, when $m = 1$, the above differentials are commonly denoted as $D_{a,f}(x) = f(x + a) + f(x)$, which is a derivative of f in direction $a \neq 0$, and obviously $D_{a,f}(x) \in \mathscr{B}_n$.

Exercise 4.4.3 *Show that if* $\deg(f) = d$ *then* $\deg(D_{a,f}) \leq d - 1$.

Referring back to our finite filed representation we now assume that $n = m$ and consider the derivative of $F(x) \in \mathbb{F}_{2^n}[x]$ (the ring of polynomials with coefficients in $\mathbb{F}_2^n$). That is, for $F(x) \in \mathbb{F}_{2^n}[x]$ we consider the number of solutions to $F(x+a)+F(x) = b$, where $a \in \mathbb{F}_{2^n}^*$ and $b \in \mathbb{F}_{2^n}$. Notice that if x_0 is a solution to this equation for some fixed a and b then $x_0 + a$ is a solution as well. Also, if a is fixed then clearly $\sum_{b \in \mathbb{F}_{2^n}} \delta_F(a,b) = 2^n$. Therefore, the functions for which $\delta(F) = 2$ attains the lowest possible differential spectra and are called almost perfect nonlinear (APN) functions.

Remark 4.4.4 *The term perfect nonlinear functions is reserved for polynomials over $\mathbb{F}_q$ where the prime characteristic of the filed $p \neq 2$. In this case, there exists mappings $F(x) \in \mathbb{F}_q[x]$ such that $F(x+a) - F(x)$ is a permutation over F_q for any $a \in \mathbb{F}_q^*$, thus $F(x+a) - F(x) = b$ has exactly one solution for any $a \in \mathbb{F}_q^*$ and $b \in \mathbb{F}_q$. Such mappings are called planar mappings and the known classes mainly come from linearized polynomials. For instance, the mapping $F(x) = x^2$ is planar over $\mathbb{F}_{p^n}$, for $p \neq 2$, since $F(x+a) - F(x) = x^2 + 2ax + a^2 - x^2 = 2ax + a^2$ due to the fact that $\alpha x + \beta$ is a permutation over $\mathbb{F}_{p^n}$ for any nonzero α and any β.*

Example 4.4.5 *Let $F(x) = x^3$ over $\mathbb{F}_{2^n}$, where n is odd. Then, F is an APN permutation. The permutation property being clear, we need to show that $F(x+a)+F(x) = b$ admits at most two solutions for any $a \in \mathbb{F}_{2^n}^*$ and $b \in \mathbb{F}_{2^n}$. Indeed, $F(x+a)+F(x) = (x+a)^3 + x^3 = ax^2 + a^2x + a^3$ so that $ax^2 + a^2x + a^3 = b$ is of degree 2 and can have at most two solutions in the field.*

Since for any $\alpha \in \mathbb{F}_{2^n}$ we have $\alpha^{2^n-1} = 1$ it is sufficient to consider polynomials of degree up to $2^n - 1$, that is the polynomials of the form $F(x) = \sum_{i=0}^{2^n-1} a_i x^i$, where $a_i \in \mathbb{F}_{2^n}$. Notice that this global degree of a polynomial in $\mathbb{F}_{2^n}[x]$ does not correspond to the algebraic degree of F defined previously. More precisely, the algebraic degree of F corresponds to the largest Hamming weight of i for which $a_i \neq 0$, see Carlet [1] which is an excellent reference for all topics treated here. To realize this consider $F(x) = x^4$ over $\mathbb{F}_{2^n}$ whose algebraic degree is only 1 since it belongs to the class of linearized polynomials over the finite filed of the form $L(x) = \sum_{i=0}^{n-1} a_i x^{2^i}$. If $\alpha_1, \ldots, \alpha_n$ is a basis of $\mathbb{F}_{2^n}$ (through the isomorphism of $\mathbb{F}_2^n$ and $\mathbb{F}_{2^n}$) so that any element $x \in \mathbb{F}_{2^n}$ can be uniquely represented as $x = x_1\alpha_1 + \ldots + x_n\alpha_n$, where $x_i \in \mathbb{F}_2$, then,

$$x^4 = (x_1\alpha_1 + \ldots + x_n\alpha_n)^4 = x_1^4\alpha_1^4 + \ldots + x_n^4\alpha_n^4 = x_1\alpha_1^4 + \ldots + x_n\alpha_n^4,$$

since in the Boolean ring $x_i^2 = x_i$. In this representation we actually consider $F : \mathbb{F}_2^n \to \mathbb{F}_2^n$, where $x = (x_1, \ldots, x_n) \mapsto (f_1(x_1, \ldots, x_n), \ldots, f_n(x_1, \ldots, x_n))$, and each f_i is a linear Boolean function. Notice that $\alpha_1^4, \ldots, \alpha_n^4$ is just a linear transformation of the basis (Forbenius automormhism).

Example 4.4.6 *Let $F(x)=x^3$ over $\mathbb{F}_{2^3}$ defined by a primitive polynomial $p(x)=x^3+x+1$ over $\mathbb{F}_2$. Let α be primitive element of $\mathbb{F}_{2^3}$, i.e., $\alpha^3=\alpha+1$ and let $\{1,\alpha,\alpha^2\}$ be a polynomial basis of $\mathbb{F}_{2^3}$. Then the component functions of $F(x)=1\cdot f_1(x_1,x_2,x_3)+\alpha f_2(x_1,x_2,x_3)+\alpha^2 f_3(x_1,x_2,x_3)$ are derived as,*

$$\begin{aligned} F(x) &= x^3=(x_0+\alpha x_1+\alpha^2 x_2)^3= \\ &= (x_0+\alpha x_1+\alpha^2 x_2)(x_0+\alpha x_1+\alpha^2 x_2)^2= \\ &= (x_0+\alpha x_1+\alpha^2 x_2)(x_0+\alpha^2 x_1+\alpha^4 x_2) \overset{\alpha^3=\alpha+1}{=} \ldots \\ &= (x_0+x_1+x_2+x_1x_2)+\alpha(x_1+x_0x_1+x_0x_2)+\alpha^2(x_2+x_0x_1) \end{aligned}$$

Notice that the algebraic degree of F above is 2 since the binary (Hamming) weight of 3 is $wt(3)=2$. Concludingly, even though x^3 is an APN permutation and an AB function as well (thus achieving the maximum nonlinearity) its algebraic degree is low and therefore its use in block ciphers is not recommended. We conclude this section with one of the most elegant problem in the theory of finite fields (related to cryptography) which is the existence of APN permutations for even n.

Open Problem 4.4.7 *For even $n>6$, find a class (or single function) which is an APN permutation or disprove their existence !! Only recently, Dillon [2] exceptionally confirmed the existence of such mappings for $n=6$ using very sophisticated connections with coding theory.*

4.5 Vectorial bent functions

While the construction of Boolean bent functions (at least those in $\mathscr{M}$ class was easy and generic, the construction of $F:\mathbb{F}_2^n\to\mathbb{F}_2^k$ is not that obvious. Now we have to ensure that for $F(x)=(f_1(x),\ldots,f_k(x))$ all nonzero linear combinations of the form $a_1f_1(x)+\ldots+a_kf_k(x)$ are bent, where f_i are Boolean functions. The bound on k for which it is possible to find such a collection was given by Nyberg [8], that is, $k\leq n/2$. The design of such functions achieving the upper bound on k, that is $k=n/2$, was only given in terms of sequences and the representation of these functions in [14] is not univariate (meaning that their representation as polynomials over finite fields is unclear). In a recent work [7], the structure of the cyclic group of the 2^k+1 roots of the unity was used to derive one complete class of vectorial bent functions $F:\mathbb{F}_2^n\to\mathbb{F}_2^{n/2}$ in a univariate representation.

Let us define the trace function $Tr_m^n:\mathbb{F}_{2^n}\to\mathbb{F}_{2^m}$, a mapping to a subfield $\mathbb{F}_{2^m}$ when $m\mid n$, is defined as

$$Tr_m^n(x)=x+x^{2^m}+x^{2^{2m}}+\ldots+x^{2^{(n/m-1)m}},\text{ for all } x\in\mathbb{F}_{2^n}. \tag{4.15}$$

The absolute trace $Tr_1^n:\mathbb{F}_{2^n}\to\mathbb{F}_2$, also denoted by Tr, then maps to the prime field. Let also $n=2k$, and denote by L the field $\mathbb{F}_{2^n}$ and its subfield $\mathbb{F}_{2^k}$ by K. Let $\mathscr{U}=$

$\{u \in L : u^{2^k+1} = 1\}$ be the cyclic subgroup of L of order 2^k+1, which is essentially the group of (2^k+1)th primitive roots of unity. Then, $\alpha^{2^k-1} = \beta$ is a generator of $\mathscr{U}$, and $\mathscr{U} = \{\alpha^{s(2^k-1)}, s = 0,\ldots,2^k\}$, where $\alpha \in L$ is a primitive element. Now, any element $x \in L^*$ can be uniquely represented as $x = \gamma u$, where $\gamma \in K^*$ and $u \in \mathscr{U}$, and furthermore $\cup_{u\in\mathscr{U}} uK^* = L^*$. For convenience, we denote $P(x) = \sum_{i=1}^{t} a_i x^{i(2^k-1)}$ so that $F(x) = Tr_k^n(P(x))$. The following result specify three equivalent necessary and sufficient conditions (we only state two here) for F to be vectorial bent [7].

Theorem 4.5.1 *Let $n = 2k$, and define $F(x) = Tr_k^n(P(x))$, where $P(x) = \sum_{i=1}^{t} a_i x^{i(2^k-1)}$ and $t \leq 2^k$. Then the following conditions are equivalent:*

1. *F is a vectorial bent function of dimension k.*
2. *$\sum_{u\in\mathscr{U}} (-1)^{Tr_1^k(\lambda F(u))} = 1$ for all $\lambda \in K^*$.*
3. *There are two values $u \in \mathscr{U}$ such that $F(u) = 0$, and furthermore if $F(u_0) = 0$, then F is one-to-one and onto from $\mathscr{U}_0 = \mathscr{U} \setminus u_0$ to K.*

The proof is rather tedious but relies on the nice property of the exponents (known as Dillon exponent) of the terms $x^{i(2^k-1)}$. Indeed, since $x \in GF(2^n)$ can be written as $x = uy$ for $u \in U$, $y \in GF(2^k)$, then $F(uy) = \sum_{i=1}^{t} a_i (uy)^{i(2^k-1)} = \sum_{i=1}^{t} a_i u^{i(2^k-1)} = F(u)$, as $y^{i(2^k-1)} = 1$ for any y because $y \in K^*$. This means that F is constant on any coset uK^* which makes the analysis much easier.

Exercise 4.5.2 *(Semi-hard) Show the item (2) above by using the fact that F is vectorial bent if and only if $W_F(\lambda,\sigma) = \pm 2^k$ for any $\lambda \in K^*$ and any $\sigma \in L$. Here, $W_F(\lambda,\sigma) = \sum_{x\in L} (-1)^{Tr_1^k(\lambda F(x)) + Tr_1^k(\sigma x)}$. Use the representation $x = u\gamma$ for the elements in L^*, and that $F(u\gamma) = F(u)$ for any $\gamma \in K^*$. Thus $W_F(\lambda,\sigma)$ can be therefore written (using $F(0) = 0$) as $1 + \sum_{u\in U} \sum_{\gamma\in K^*} (-1)^{Tr_1^k(\lambda F(u\gamma)) + Tr_1^k(\sigma u\gamma)}$ ...*

We conclude this part by mentioning that there exist various generalizations of the concept of bent functions, for instance one may naturally define $F : \mathbb{F}_p^n \to \mathbb{F}_p^n$, for prime $p \neq 2$, but this requires a modification of the main cryptographic notions.

4.6 Graph theoretic aspects of Boolean functions

Let G be a multiplicative group of order v. A k-subset D of G is a (v,k,λ,μ) partial difference set (PDS) if each non-identity element in D can be represented as gh^{-1} $(g,h \in D, g \neq h)$ in exactly λ ways, and each non-identity element in $G \setminus D$ can be represented as gh^{-1} $(g,h \in D, g \neq h)$ in exactly μ ways. We shall always assume that the identity element 1_G of G is not contained in D. Using the language of group ring algebra $R[G]$,

a k-subset D of G with $1_G \notin D$ is a (v,k,λ,μ)-PDS if and only if the following equation holds:

$$DD^{(-1)} = (k-\mu)1_G + (\lambda-\mu)D + \mu G, \tag{4.16}$$

where in $R[G]$ we denote $D = \sum_{g\in G} d_g g$ and $D^{(t)} = \sum_{g\in G} d_g g^t$, for $d_g \in R$. Combinatorial objects associated with partial difference sets are strongly regular graphs. A graph Γ with v vertices is called a (v,k,λ,μ) strongly regular graph (SRG) if each vertex is adjacent to exactly k other vertices, any two adjacent vertices have exactly λ common neighbours, and any two non-adjacent vertices have exactly μ common neighbours. Given a group G of order v and a k-subset D of G with $1_G \notin D$ and $D^{-1} = D$, the graph $\Gamma = (V,E)$ is called the Cayley graph generated by D in G and is defined as follows:

1. The vertex set V is G;
2. Two vertices g, h are joined by an edge if and only if $gh^{-1} \in D$.

The following result links together the notions of partial difference set and the property of a graph being strongly regular.

Theorem 4.6.1 [13] *Let Γ be the Cayley graph generated by a k-subset D of a multiplicative group G with order v. Then Γ is a (v,k,λ,μ) strongly regular graph if and only if D is a (v,k,λ,μ)-PDS with $1_G \notin D$ and $D^{-1} = D$.*

Note that in the binary case, when Boolean functions $f : \mathbb{F}_2^n \to \mathbb{F}_2$ are considered, the Cayley graph is induced with respect to a subset of the elementary additivie Abelian 2-group $\mathbb{F}_2^n$. Since the condition that for $d \in D$ we must have $-d \in D$, any $D \subseteq \mathbb{F}_{2^n}$ will define the Cayley graph (each element is its own additive inverse) so that there is an edge between g and h if and only if $h \oplus g \in D$. The Cayley graph $\Gamma_f = (\mathbb{F}_2^n, E_f)$ associated to a Boolean function f is defined by selecting $D = \{x \in \mathbb{F}_2^n : f(x) = 1\}$ (D is called the support set of f) and defining the set of edges as,

$$E_f = \{(u,w) \in \mathbb{F}_2^n \times \mathbb{F}_2^n \mid f(\mathbf{u} \oplus \mathbf{w}) = 1\},$$

where for convenience we use the boldface to denote the elements of $\mathbb{F}_2^n$ so that $\mathbf{u} = (u_1, \ldots, u_n)$. The operation $\oplus$ over $\mathbb{F}_2^n$ is of course the componentwise modulo 2 addition. Furthermore, we specify the elements of $\mathbb{F}_2^n$ by using the decimal representation of their indices, thus $\mathbf{u}_0 = (0,\ldots,0)$, $\mathbf{u}_1 = (1,\ldots,0)$, …, $\mathbf{u}_{2^n-1} = (1,\ldots,1)$.

A graph is called *r*egular of degree (valency) r if every vertex has degree (valency) r, that is, the number of edges incident to it is r. The Cayley graph Γ_f associated to any Boolean function f is obviously D regular. On the other hand, such a graph with parameters $(\mathbb{F}_2^n, D, d, e)$ is called *s*trongly regular graph (SRG) if there exist nonnegative integers e, d such that for all vertices u, v the number of vertices adjacent to both u and v is e if u, v are adjacent, respectively, this number is d if u, v are nonadjacent. An easy

counting argument shows that $D(D-d-1)=e(v-D-1)$. Notice that in general strongly regular graphs appear to be difficult to investigate.

The adjacency matrix A_f of size $2^n \times 2^n$ is the matrix whose entries are $A_{i,j} = f(\mathbf{u}_i \oplus \mathbf{u}_j)$, thus $A_{i,j} = 1$ if and only if $\mathbf{u}_i$ and $\mathbf{u}_j$ are connected. Given a graph Γ_f and its adjacency matrix A_f the *s*pectrum $Spec(\Gamma_f)$ is the set of eigenvalues of A_f. The following result specifies the eigenvalues in terms of Walsh coefficients and vice versa.

Theorem 4.6.2 *Let $f : \mathbb{F}_2^n \to \mathbb{F}_2$, and let λ_i, $0 \le i \le 2^n - 1$ be the eigenvalues of its associated graph Γ_f. Then $\lambda_i = W_f(\mathbf{b}_i)$, for any i.*

PROOF. The eigenvectors of the Cayley graph Γ_f are the characters $Q_{\mathbf{w}}(x) = (-1)^{\mathbf{w}\cdot x}$ of $\mathbb{F}_2^n$. Moreover, the i-th eigenvalue of A_f, corresponding to the eigenvector $Q_{\mathbf{b}_i}$ is given by $\lambda_i = \sum_{x\in\mathbb{F}_2^n}(-1)^{\mathbf{b}_i\cdot x} f(x) = W_f(\mathbf{b}_i)$. ■

It is known that a connected r-regular graph is strongly regular if and only if it has exactly three distinct eigenvalues $\lambda_0 = r$ (or $\lambda_0 = D$ in our notation) and λ_1, λ_2. Furthermore, we have the following $e = r+\lambda_1\lambda_2+\lambda_1+\lambda_2$ and $d = r+\lambda_1\lambda_2$. It can be shown that bent functions, thus n is even, are the only Boolean functions whose associated Cayley graph is a strongly regular graph with $e = d$. In particular, for bent functions we have $\lambda_2 = -\lambda_1 = 2^{n/2-1}$ and $\lambda_0 = D = 2^{n-1} \pm 2^{n/2-1}$.

Exercise 4.6.3 *For $n = 4$ verify that $f(x_1,\ldots,x_4) = x_1x_2+x_3x_4$ is a bent function. Compute the parameters $e = d$.*

An additional property of bent functions is related to the notion of the triangle-free property. In other words, a graph is triangle-free if there are no paths of the form $xyzx$, where the vertices x, y, z are distinct. It can be shown that if Γ_f is triangle-free then f cannot be bent. But this property cannot be used for distinguishing the bent property of Boolean functions since the converse is not true. That is, there are functions whose graphs contain (many) triangles but they are not bent.

4.7 References

[1] C. Carlet, *B*oolean Functions for Cryptography and Error Correcting Codes. Cambridge University Press, 2010.

[2] J. Dillon, APN polynomials: An update. In *F*q9, the 9th International Conference on Finite Fields and Applications, 2009.

[3] J. F. Dillon, Elementary Haddamard Difference Sets. Ph. D. thesis, University of Maryland, U.S.A., 1974.

[4] F. J. MacWilliams and N. J. A. Sloane, *T*he Theory of Error-Correcting Codes. North-Holland, Amsterdam, 1977.

[5] J. L. Massey, Shift-register synthesis and BCH decoding. *I*EEE Trans. on Inform. Theory, IT-15(1):122–127, 1969.

[6] A. Menezes, P. van Oorschot, and S. Vanstone, *H*andbook of Applied Cryptography. CRC Press, Boca Raton, 1997.

[7] A. Muratovic-Ribic, E. Pasalic, and S. Bajrić, An analysis of multiple output trace bent functions with nonlinear Niho exponents using symmetric polynomials. *I*EEE Trans. on Inform. Theory, IT-60(2):1337–1347, 2014.

[8] K. Nyberg, Perfect nonlinear S-boxes. In *A*dvances in Cryptology—EUROCRYPT'91, volume LNCS 547, pages 378–385. Springer-Verlag, 1991.

[9] C. E. Shannon, A mathematical theory of communication. *B*ell System Technical Journal, Vol. 27:379–423 (Part I) and 623–656 (Part II), 1948.

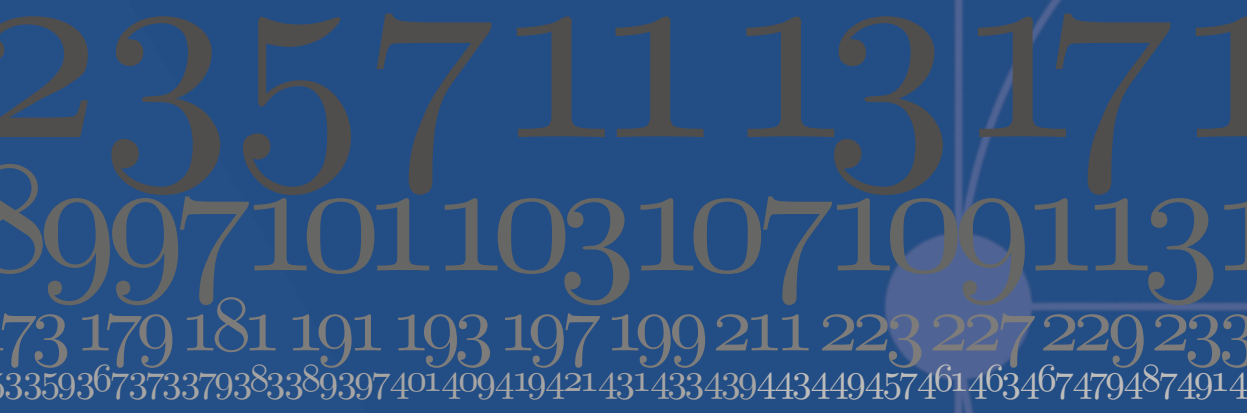
[10] C. E. Shannon, Communication theory of secrecy systems. *B*ell System Technical Journal, Vol. 27:656–715, 1949.

[11] T. Siegenthaler, Correlation-immunity of nonlinear combining functions for cryptographic applications. *I*EEE Trans. on Inform. Theory, IT-30:pages 776–780, 1984.

[12] G. J. Simmons, *C*ontemporary Cryptology. Wiley-IEEE Press, New York, 1999.

[13] G-Z. Xiao and J. L. Massey, A spectral characterization of correlation-immune combining functions. *I*EEE Trans. on Inform. Theory, IT-34:569–571, 1988.

[14] A. M. Youssef and G.. Gong, Hyper-bent functions. In *A*dvances in Cryptology—EUROCRYPT 2001, volume LNCS 2045, pages 406–419. Springer-Verlag, 2001.



Založba Univerze na Primorskem
Titov trg 4, SI-6000 Koper
www.hippocampus.si
zalozba@upr. si

Not for sale

9 789616 832922

REPUBLIKA SLOVENIJA
MINISTRSTVO ZA IZOBRAŽEVANJE, ZNANOST IN ŠPORT

Naložba v vašo prihodnost
OPERACIJO DELNO FINANCIRA EVROPSKA UNIJA
Evropski socialni sklad

Operacijo delno financira Evropska unija, in sicer iz Evropskega socialnega sklada. Projekt se izvaja v okviru Operativnega programa razvoja človeških virov 2007-2013, razvojne prioritete 3: "Razvoj človeških virov in vseživljenjskega učenja"; prednostne usmeritve 3.3 "Kakovost, konkurenčnost in odzivnost visokega šolstva".