MATROIDS AND CANONICAL FORMS: THEORY AND APPLICATIONS

Gregory F Henselman

A DISSERTATION

in

Electrical and Systems Engineering

Presented to the Faculties of the University of Pennsylvania in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy

2017

Supervisor of Dissertation

Robert W Ghrist, Andrea Mitchell University Professor of Mathematics and Electrical and Systems Engineering

Graduate Group Chairperson

Alejandro Ribeiro, Rosenbluth Associate Professor of Electrical and Systems Engineering

Dissertation Committee:
Chair: Alejandro Ribeiro, Rosenbluth Associate Professor of Electrical and Systems Engineering
Supervisor: Robert W Ghrist, Andrea Mitchell University Professor of Mathematics and Electrical and Systems Engineering
Member: Rakesh Vohra, George A. Weiss and Lydia Bravo Weiss University Professor of Economics and Electrical and Systems Engineering

MATROIDS AND CANONICAL FORMS: THEORY AND APPLICATIONS

For Linda

DEDICATION

ΓΙΑΤΗΓΙΑΓΙΑΜΟΥΠΟΥΜΟΥΕΔΩ
ΣΕΤΟΜΥΑΛΟΓΙΑΤΟΝΠΑΠΠΟΥΜ
ΟΥΠΟΥΜΟΥΕΔΩΣΕΤΟΠΝΕΥΜΑ
ΤΟΥΓΙΑΤΗΜΗΤΕΡΑΜΟΥΠΟΥΜΟ
ΥΕΔΩΣΕΤΗΑΓΑΠΗΤΗΣΓΙΑΤΟΝΠ
ΑΤΕΡΑΜΟΥΠΟΥΜΟΥΕΔΩΣΕΤΗ
ΝΕΛΠΙΔΑΤΟΥΓΙΑΤΗΝΑΔΕΛΦΗΜ
ΟΥΠΟΥΚΡΑΤΑΕΙΤΟΝΔΡΟΜΟΜΑ
ΣΚΑΙΓΙΑΤΟΝΘΑΚΟΟΠΟΙΟΣΦΩΤ
ΙΖΕΙΤΑΜΑΤΙΑΜΑΣΓΙΑΚΑΙΤΛΙΝΤ
ΟΥΟΠΟΙΟΥΗΦΙΛΙΑΕΙΝΑΙΗΚΑΛΥ
ΤΕΡΗΜΟΥΣΥΜΒΟΥΛΗΓΙΑΘΑΝΕ
ΤΗΣΟΠΟΙΑΣΤΟΕΡΓΟΕΙΝΑΙΑΥΤΟ

ACKNOWLEDGEMENTS

I would like to thank my advisor, Robert Ghrist, whose encouragement and support made this work possible. His passion for this field is inescapable, and his mentorship gave purpose to a host of burdens.

A tremendous debt of gratitude is owed to Chad Giusti, whose sound judgement has been a guiding light for the past nine years, and whose mathematical insights first led me to consider research in computation.

A similar debt is owed to the colleagues who shared in the challenges of the doctoral program: Jaree Hudson, Kevin Donahue, Nathan Perlmutter, Adam Burns, Eusebio Gardella, Justin Hilburn, Justin Curry, Elaine So, Tyler Kelly, Ryan Rogers, Yiqing Cai, Shiying Dong, Sarah Costrell, Iris Yoon, and Weiyu Huang. Thank you for your friendship and support at the high points and the low.

This work has benefitted enormously from interaction with mathematicians who took it upon themselves to help my endeavors. To Brendan Fong, David Lipsky, Michael Robinson, Matthew Wright, Radmila Sazdanovic, Sanjeevi Krishnan, Paweł Dlotko, Michael Lesnick, Sara Kalisnik, Amit Patel, and Vidit Nanda, thank you.

Many faculty have made significant contributions to the content of this text, whether directly or indirectly. Invaluable help came from Ortwin Knorr, whose instruction is a constant presence in my writing. To Sergey Yuzvinsky, Charles Curtis, Hal Sadofsky, Christopher Phillips, Dev Sinha, Nicholas Proudfoot, Boris Botvinnik, David Levin, Peng Lu, Alexander Kleshchev, Yuan Xu, James Isenberg, Christopher Sinclair, Peter Gilkey, Santosh Venkatesh, Matthew Kahle, Mikael Vejdemo-Johansson, Ulrich Bauer, Michael Kerber, Heather Harrington, Nina Otter, Vladimir Itskov, Carina Curto, Jonathan Cohen, Randall Kamien, Robert MacPherson, and Mark Goresky, thank you. Special thanks to Klaus Truemper, whose text opened the world of matroid decomposition to my imagination.

Special thanks are due, also, to my dissertation committee, whose technical insights continue to excite new ideas for the possibilities of topology in complex systems, and whose coursework led directly to my work in combinatorics.

Thank you to Mary Bachvarova, whose advice was decisive in my early graduate education, and whose friendship offered shelter from many storms.

Finally, this research was made possible by the encouragement, some ten years ago, of my advisor Erin McNicholas, and of my friend and mentor Inga Johnson. Thank you for bringing mathematics to life, and for showing me the best of what it could be. Your knowledge put my feet on this path, and your faith was the reason I imagined following it.

ABSTRACT

MATROIDS AND CANONICAL FORMS: THEORY AND APPLICATIONS

Gregory F Henselman
Robert W Ghrist

This document introduces a combinatorial theory of homology, a topological descriptor of shape. The past fifteen years have seen a steady advance in the use of techniques and principles from algebraic topology to address problems in the data sciences. This new subfield of *Topological Data Analysis* [TDA] seeks to extract robust qualitative features from large, noisy data sets. A primary tool in this new approach is the *homological persistence module*, which leverages the categorical structure of homological data to generate and relate shape descriptors across scales of measurement. We define a combinatorial analog to this structure in terms of *matroid canonical forms*. Our principle application is a novel algorithm to compute persistent homology, which improves time and memory performance by up to several orders of magnitude over current state of the art. Additional applications include new theorems in discrete, spectral, and algebraic *Morse theory*, which treats the geometry and topology of abstract space through the analysis of critical points, and a novel paradigm for matroid representation, via abelian categories. Our principle tool is *elementary exchange*, a combinatorial notion that relates linear and categorical duality with matroid complementarity.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Motivation: computational homology

In the past fifteen years, there has been a steady advance in the use of techniques and principles from algebraic topology to address problems in the data sciences. This new subfield of *Topological Data Analysis* [TDA] seeks to extract robust qualitative features from large, noisy data sets. At the simplest and most elementary level, one has *clustering*, which returns something akin to connected components. There are, however, many higher-order notions of global features in connected components of higher-dimensional data sets that are not describable in terms of clustering phenomena. Such "holes" in data are quantified and collated by the classical tools of algebraic topology: *homology*, and the more recent, data-adapted, parameterized version, *persistent homology* [15, 24, 36]. Homology and persistent homology will be described in Chapter 9. For the moment, the reader may think of homology as an enumeration of "holes" in a data set, outfitted with the structure of a sequence of vector spaces whose bases identify and enumerate the "essential" holes in a data set.

This work has as its motivation and primary application, the efficient computation of homology, persistent homology, and higher variants (such as cellular sheaf cohomology [18]) for application to TDA. Computational homology is an intensely active area of research with a rich literature [44, 26, 25]. For this introductory summary, it suffices to outline a little of the terminology without delving into detail. Homology takes as its input a sequence of vector spaces $C = (C_k)$ and linear transformations $\partial_k \colon C_k \to C_{k-1}$, collectively known as a *chain complex* that, roughly, describes how simple pieces of a space are assembled.

$$C \;=\; \cdots \longrightarrow C_k \xrightarrow{\partial_k} C_{k-1} \xrightarrow{\partial_{k-1}} \cdots \xrightarrow{\partial_2} C_1 \xrightarrow{\partial_1} C_0 \xrightarrow{\partial_0} 0\;. \tag{1.0.1}$$

The chain complex is the primal object in homological algebra, best seen as the higher-dimensional analogue of a graph together with its adjacency matrix. This "algebraic signal" is compressed to a homological core through the standard operations of linear algebra: kernels and images of the boundary maps ∂.

The standard algorithm to compute homology of a chain complex is to compute the

Smith normal form of the aggregated boundary map $\partial : C \to C$, where one concatenates the individual terms of (1.0.1) into one large vector space. This graded boundary map has a block structure with zero blocks on the block-diagonal (since $\partial_k \colon C_k \to C_{k-1}$) and is nonzero on the superdiagonal blocks. The algorithm for computing Smith normal form is a slight variant of the ubiquitous Gaussian elimination, with reduction to the normal form via elementary row and column operations. For binary field coefficients this reduction is easily seen to be of time-complexity $O(n^3)$ in the size of the matrix, with an expected run time of $O(n^2)$.

This is not encouraging, given the typical sizes seen in applications. One especially compelling motivation for computing homology comes from a recent set of breakthrough applications in neuroscience by Giusti et al. [37], which uses homology to extract network structure from a collection of neurons based solely on a correlation matrix. For a typical experimental rig of 250 neurons, the resulting 250-by-250 matrix leads to chain complexes whose net dimension (the size of the resulting boundary matrix) is in the hundreds of billions. Such values of n frustrates the usual algorithms: see [68] for benchmarks, which as of January 2017 state that with best available software over a 1728-core Sandybridge cluster, the largest complex tested has net dimension $n \approx 3 \times 10^9$. The motivation for and direct outcome of this thesis is an algorithm for the efficient computation of homology in larger systems, immediately useful in TDA for Neuroscience and more.

Approach: three ingredients

In order to achieve a breakthrough in computational speed and memory management, this thesis turns to increased abstraction as the means of ascent. There are three ingredients that, though all classical and well-known in certain sectors, are synthesized in a novel way here. These ingredients are as follows.

1. **Matrix Factorization:** This first ingredient is the most familiar and least surprising. As homology computation in field coefficients is little more than Gaussian reduction, one expects the full corpus of matrix factorization methods to weigh heavily in any approach. One novelty of this thesis is the reconciliation of matrix factorization with the more abstract (category-theoretic) approaches to homology, as well as to the preprocessing/reduction methods of [44].

2. **Matroid Theory:** The hero of this story is matroid theory, the mathematical fusion of combinatorial geometry and optimization theory [11, 69, 80]. Matroids have a rich history in combinatorics and combinatorial optimization, but are largely absent in the literature on computational homology. This thesis introduces the language and methods of matroids to computational algebraic topology, using this to relate combinatorial optimization to homology computation. As a result, a novel and wholly combinatorial approach to homology is derived, using filtrations (and bifiltrations) of matroids as an extension of the notion of a chain complex (1.0.1) above.

3. **Homological Algebra:** This core branch of Mathematics is simply the algebra of diagrams [35]. In its simplest emanation, one works with, first, sequences of vector

spaces and linear transformations [chain complexes], generalizing to more intricate diagrams. The key functional tools of linear algebra — kernels, images, cokernels, coimages, and the like — yield inferential engines on diagrams, the simplest of which are homologies. One quickly sees that, as with combinatorial structures leading to matroids, diagrams of vector spaces rely little on the actual details of linear algebra: only the core constructs count. This prompts the usual ascension to *abelian categories* with vector-space-like objects, and transformations with suitable notions of kernels and cokernels. This thesis synthesizes homological algebra with matroid and matrix factorization methods.

As a simple example of the intersection of these three subfields, consider the classical, essential operation of *pivoting* in Gaussian reduction. The correspondences alluded to above yield a reinterpretation of pivoting as (1) a manifestation of the *Exchange Lemma* in a particular matroid, and (2) an instance of the *Splitting Lemma* in the homological algebra of short exact sequences. Though not in itself a deep result, this observation points the way to deeper generalizations of matrix methods in homological and combinatorial worlds.

The Gaussian pivot, the exact splitting, and the matroid basis exchange, are a single event. Iteration of this step is the natural progression and, in these domains, points to the common thread connecting them with each other and with computational homology. This mutual intersection is *Discrete Morse Theory* [DMT], a fairly novel generalization [32, 48] of the classical Morse Theory on manifolds [57]. Discrete Morse Theory has been the basis for some of the most effective compression schemes for simplicial and cellular complexes, leading to novel algorithms [60] and software. This thesis unearths a previously unknown connection between Discrete Morse Theory and (1) the Schur complement in matrix factorization; (2) minimal basis representations and greedy optimization in matroid theory; and (3) exact seqeunces in homological algebra.

The technical details of how these subjects merge and react to yield dramatic improvements in homology computation are not difficult. Indeed, they are presaged in a canonical result seen by every undergraduate student of Mathematics that acts as a microcosm, combining elements of combinatorics, matrix factorization, and (hidden) homological algebra. This key foreshadowing is the *Jordan Form*.

Primal example: Jordan bases

Consider the (abelian) category of finite-dimensional $\mathbf{C}$-linear vector spaces and $\mathbf{C}$-linear maps. The Jordan bases of a complex operator are well-known. Less well-known is the implicit matroid-theoretic structure of Jordan bases: this has not, as far as the author knows, appeared in published literature. On a formal level the connection is fundamental: matroid theory is built on the study of *minimal bases* (read: bases subordinate to a flag of vector spaces) and the study of Jordan forms centers on bases for flags stabilized by a linear operator. That this connection is deep can be demonstrated by an application to the problem of computing Jordan bases using minimal bases, which is simple enough to describe with no matroid-theoretic language at all, though some notation will be required.

The story begins with a standard reduction: to describe the Jordan bases of an arbitrary complex operator, it suffices to describe those of a nilpotent one, since to every operator corresponds a nilpotent with identical Jordan bases. Therefore fix a T such that $T^n = 0$, for some n. The approach will hinge on the canonical projection map, q, from the base space of T to its quotient by the image of T. A subset of the base space is *q-independent* if it maps to an independent set under q. Likewise, sets that map to bases are *q-bases.* These definitions are slightly nonstandard, but the reader who continues through the background section will see where they fit into ordinary combinatorial terminology.

For every real-valued function $\mathcal{K}$ on the domain of T and every finite subset E, define the *$\mathcal{K}$-weight* of E to be $\sum_{e \in E} \mathcal{K}(e)$. A q-basis B has *minimum weight* if its weight is minimal among q-bases. The specific weight function that will occupy our attention is the one uniquely defined by the condition $\mathrm{Ker}(T^m) = \{\mathcal{K} \leq m\}$, for all nonnegative integers m.

For convenience, let the *orbit* of E be the set of all nonzero vectors that may be expressed in form $T^m e$, for some e in E, and the *support* of a linear combination $\sum_{e \in E} \alpha_e e$ be the set of all e such that α_e is nonzero. We may now describe the relationship between q-bases and Jordan bases precisely.

Proposition 1.0.1. *The Jordan bases of a nilpotent operator T are exactly the orbits of minimum-weight q-bases.*

Proof. Let B be a q-basis of minimum weight, and let E be its orbit. We will first show that E is independent.

If some nontrivial linear combination in E evaluates to zero then, by factoring out as many powers of T as possible, the same can be expressed as $T^m \sigma$, where σ is a linear combination in E whose support contains at least one element of B. Let S denote the support of σ, and assume, without loss of generality, that no element of S vanishes under T^m.

Since $q(\sigma)$ lies in the span of $q(S \cap B)$, there is an s in $S \cap B$ for which $q(S \cap B)$ and $q(S \cap B - \{s\} \cup \{\sigma\})$ have equal spans. Evidently, $B - \{s\} \cup \{\sigma\}$ is a q-basis. Since T^m vanishes on σ and not on s, however, this new basis has weight strictly less than B. This is impossible, given our starting hypothesis, so E must be independent.

To see that E has full rank, let U denote the quotient of the base space by the span of E. Our operator induces a nilpotent map on U, and the cokernel of that map is trivial if and only if U is trivial. By the First Isomorphism Theorem, the same cokernel may be identified with the quotient of the base space by the span of $E \cup \mathrm{Im}(T)$. The latter vanishes, and so does U. Thus E is a basis. Evidently, it is a Jordan basis.

This establishes that the orbit of every minimum-weight q-basis is a Jordan basis. For the converse, consider a bijection between Jordan bases and certain of the q-bases. In the forward direction, this map sends J to $J - \mathrm{Im}(T)$. In reverse, it sends $J - \mathrm{Im}(T)$ to its orbit. Thanks to uniqueness of the Jordan decomposition, every q-basis in the image of this map will have equal weight. At least one such will have minimum weight – for we have shown that at least one Jordan basis can be expressed as the orbit of a minimum-weight q-basis – and therefore all do. Thus every Jordan basis is the orbit of a minimum-weight q-basis. □

Conveniently, minimum weight q-bases have a simple description. Let K_m denote the kernel of T^m, and put $B_m = B \cap K_m$

Proposition 1.0.2. *A q-basis B has minimum weight if and only if $q(B_m)$ spans $q(K_m)$, for all m.*

Proof. Fix any q-basis B, and suppose $q(B_m)$ does not span $q(K_m)$, for some m. Then there exists $\sigma \in K_m$ such that $q(\sigma)$ lies outside the span of $q(B_m)$, and one can construct a new q-basis by replacing an element of $B - B_m$ with σ. The new basis will weigh strictly less than B, so B is not minimal. □

The following corollary is an immediate consequence. To save unnecessary verbiage, let us say that a set *represents* a basis in a quotient space if it maps to one under the canonical projection map.

Corollary 1.0.3. *A set B is a minimum-weight q-basis if and only if there exist $I_m \subseteq K_m$ such that*

$$B = I_1 \cup \cdots \cup I_n$$

and I_m represents a basis in $q(K_m)/q(K_{m-1})$, for all m.

Proof. If B is a minimum-weight basis then we may take $I_m = B_m - B_{m-1}$, by Proposition 1.0.2. The converse is simple reversal. □

An immediate application of these results is the ease with which one can now compute Jordan bases: For each m, collect enough vectors from K_m to represent a basis in $q(K_m)/q(K_{m-1})$. Then, take their union.

A second application regards the "constructive" proof of the Jordan form in linear algebra texts. A noted source of discomfort with this approach is that its construction procedure can be shown to work, but there is no clear sense *why* it works. Here, too, discrete optimization can shed some light. As examples we take three clean constructive arguments.

1. The first, by Tao [78], begins with an arbitrary basis of the base space. In general, the orbit of this set will span the space but fail to be linearly independent. It is shown that each linear dependence relation reveals how to augment one of the original basis vectors so as to shorten its orbit, in essence replacing $\{v, Tv, \ldots, T^n v\}$ with $\{v + u, T(v + u), \ldots, T^k(v + u)\}$, for some k less than n. Since the union of orbits grows strictly smaller after each iteration, the process terminates. The set that remains is a linearly independent union of orbits that spans the space. That is, a Jordan basis.

2. The second, by Wildon [81], inducts on the dimension of the base space. The inductive hypothesis provides a Jordan basis for the restriction of T to its image, which for a niltpotent operator lies properly inside the domain. Since this basis lies in the image, every orbit $\{v, Tv, \ldots, T^n v\}$ is contained in that of some vector u such that $v = Tu$. Moreover, some vectors in the basis vanish under T, and we may extend these to form a basis for the kernel. It is shown via algebraic operations that the set composed of the orbits of the u, *plus* all the kernel elements, is linearly independent. Dimension counting then shows the set to be a basis.

3. The third and perhaps the most direct comes from Baker [1], who argues inductively that the linear span of every maximum-length orbit has a T-invariant complement. Splitting off maximal-length orbits from successively smaller complements gives the desired decomposition.

What do these approaches have in common? With the benefit of hindsight, each computes a minimum-weight q-basis. In fact, each implements one of two classical algorithms in combinatorial optimization. In reference to q-bases, the last of these algorithms begins with an empty set, B_0, and so long as B_k is properly contained in a q-independent set of cardinality $|B_k|+1$, chooses from among these a 1-element extension of minimum weight, assigning this the label B_{k+1}. We will show in later section that, because q-bases realize the structure of a matroid, this process returns a minimum-weight basis. The common name for this procedure is the *greedy algorithm* for matroid optimization. The first algorithm begins with a complete q-basis B. So long as some element s in B can be exchanged for some element t to form a new q-basis of lesser weight, the algorithm does so. Again, because q-bases realize the structure of a matroid – and, more generally, of an M-convex set – the output is an optimal basis. This is called *gradient descent.*

Clearly, the algorithm of Tao implements gradient descent. That of Baker is a formal dual to the matroid greedy algorithm we have just described. In this formulation the weight function of interest is the length of a Jordan chains, hence the focus on maximal cycles. That of Wildon implements a quotient construction which we will discuss in later sections. Pleasingly, it takes much less work to prove and understand these observations than it did to prove our first sample proposition. Indeed, the only reason we have worked so hard in the fist place was to avoid the use of formal language.

Outline and Contributions of the Thesis

The principal arc of our story is a dramatic simplification of the example described above regarding Jordan bases. The idea is to de-clutter the anecdote by rising in abstraction and removing excess structure.

- In place of q-bases, we turn to simpler matroid bases.
- In place of linear operators, we generalize to morphisms in an Abelian category.

This lift in level of abstraction yields increased generality. This is leveraged into the following specific contributions:

1. The principal outcome of the thesis is a novel algorithm for computing homology and persistent homology of a complex. This is incarnated in the *Eirene* software platform, which is benchmarked against leading competitors and shown to give roughly two order-of-magnitude improvement in speed (elapsed time) and memory (max heap) demands on the largest complexes computed in the literature. This can be found in §9.4.
2. In §5.3 we give a novel definition of *combinatorial homology* in terms of matroid bifiltrations. This generalizes homology of chain complexes and persistent homology and permits the merger of greedy algorithms with more classical homological algebra.

3. A novel relationship between the Schur complement (in linear algebra), discrete Morse Theory (in computational homology), and minimal bases (in combinatorial optimization) is developed in §9 and §10; this is used as a key step in building the algorithm for *Eirene*.

These outcomes are the product of careful distilling of the notion of minimal bases from the foreshadowed Jordan form story above. The steps are as follows:

- Chapter 3 is a self-contained introduction to the tools from matroid theory here needed, with an emphasis on *exchange* as the key concept.
- Chapter 4 reviews modularity in matroids as the precursor for generating minimal bases.
- Chapter 5 introduces a formal notion of a nilpotent operator on a matroid, and classifies its canonical forms. As special cases, we derive combinatorial generalizations of homology and persistent homlogy.
- Chapter 6 provides formulae relating the combinatorial operation of exchange with the algebraic operation of matrix factorization. The main technical insight is an lifting of the *Möbius Inversion Formula* to the natural setting of homological algebra, abelian categories.
- Chapter 8 classifies the LU and Jordan decompositions of classical matrix algebra combinatorially. We describe combinatorial algorithms to obtain such decompositions by greedy optimization. The key idea in this formalism is the matroid theoretic application of *elementary exchange*.
- Chapter 9 applies the algorithms of Chapter 8 to the problem of efficient homology computation. Our main observation is a three-way translation between the topology, algebra, and combinatorics of a cellular space.
- Chapter 10 posits a new foundation for spectral and algebraic Morse theory. This approach is simpler and more general than the published results in either subject of which we are aware. The main idea is to lift the notion of a Morse complex to that of a Jordan complement.

Chapter 2

Notation

Although a linear-algebraic sensibility is the default mode of this thesis, some conventions are category-theoretic in nature. For instance, we employ a superscript symbol *op* to emphasize the symmetry between certain pairs of objects, such as D and D^{op} below.

To each function f we associate a domain $\mathrm{D}(f)$, an image $\mathrm{I}(f)$, and, optionally, a *codomain* $\mathrm{D}^{op}(f)$. We will write $f : A \to B$ to declare that $A = \mathrm{D}(f)$ and $B = \mathrm{D}^{op}(f)$. The *identity* function on A is the unique map $1_A : A \to A$ so that $1_A(a) = a$ for all $a \in A$. A map *into* W is a map with codomain W, and a map *out of* W is a map with domain W. We write

$$f(S) = \{f(s) : s \in S\} \qquad\qquad f^{-1}(T) = \{a \in A : f(a) \in T\}$$

for any $S \subseteq A$ and $T \subseteq B$.

The terms *collection* and *family* will be used interchangeably with the term *set*. An A*-indexed family in* B is a set function $f : A \to B$. We will sometimes write f_a for $f(a)$ when f is regarded as an index function, and denote f by $(f_a)_{a\in A}$. A *sequence in* B is an I-indexed family in B, where $I = \{n \in \mathbb{Z} : a \leq n \leq b\}$ for some $a, b \in \mathbb{Z} \cup \{-\infty, +\infty\}$. Given an indexed family f and a collection $\mathcal{I}$ of unindexed sets, we write $f \in \mathcal{I}$ when f is injective and $\mathrm{I}(f) \in \mathcal{I}$.

Several mathematical operations accept indexed families as inputs, for example sum, product, union, and intersection. By convention,

$$\sum_{a\in\emptyset} f_a = 0 \qquad \prod_{a\in\emptyset} f_a = 1 \qquad \bigcup_{a\in\emptyset} f_a = \emptyset \qquad \bigcap_{a\in\emptyset} f_a = E,$$

where $E = \cup_{a\in A} f_a$. Arguments will be dropped from the expression $\square_{s\in S} f_s$ where context leaves no room for confusion; for example, $\sum f = \sum_{a\in A} f_a$.

An *unindexed family in* B is a subset $S \subseteq B$. To every unindexed family corresponds a canonical indexed family, 1_S. By abuse of notation, we will use S and 1_S interchangeably as inputs to the standard operations, for instance writing $\sum S$ for $\sum_{s\in S} s$.

Given any relation $\sim$ on B, we write $f_{\sim b}$ for the set $\{a \in A : f(a) \sim b\}$. If f is real-valued, for example, then $f_{\leq t} = \{a \in A : f(a) \leq t\}$. Similarly, given any $C \subseteq A$, we will write $C_{f\sim b}$ for $\{c \in C : f(c) \sim b\}$. By extension we set

$$f_{\sim a} = f_{\sim f(a)} \qquad\qquad \textit{and} \qquad\qquad C_{\sim a} = C_{\sim f(a)}.$$

The *characteristic function* of a subset $J \subseteq I$ is the zero-one set function χ_J with domain I defined by

$$\chi_J(i) = \begin{cases} 1 & i \in J \\ 0 & \textit{else.} \end{cases}$$

We write $\mathbf{R}$ and $\mathbf{C}$ for the fields of real and complex numbers, respectively. Given a coefficient field $\mathbf{k}$, we write $\mathbf{k}^I$ for the set of functions $I \to \mathbf{k}$, regarded as $\mathbf{k}$-linear vector space under the usual addition and scalar multiplication. We write $\mathrm{K}(T)$ for the null space of a linear map T, and $\mathrm{Hom}(U, V)$ for the space of all linear maps $U \to V$. If V is any vector space and I is any set, then the *support* of a function $f : I \to V$ is

$$\mathrm{Supp}(f) = \{i \in I : f(i) \neq 0\}.$$

The *product* of an indexed family of $\mathbf{k}$-linear spaces $(V_i)_{i\in I}$ is the space of all maps $f : I \to \cup V$ such that $f_i \in V_i$ for all i, equipped with the usual addition and scalar multiplication

$$(f + g)_i = f_i + g_i \qquad\qquad (\alpha \cdot f)_i = \alpha \cdot f_i.$$

The *coproduct* of $(V_i)_{i\in I}$ is the subspace $\{f \in \times_i V_i : |\mathrm{Supp}(f)| < \infty\}$. The product an coproduct constructions will coincide in most cases of interest.

Example 2.0.1. By convention $\mathbf{R}^n = \mathbf{R}^{\{1,\dots,n\}}$. Thus if $V_1 = \cdots = V_n = \mathbf{R}$, then $\times V = \mathbf{R}^n$. Since every element of $\mathbf{R}^n$ has finite support, one has $\oplus V = \mathbf{R}^n$, also.

Given any family of maps $f : I \to \mathrm{Hom}(W, V_i)$, we write $\times f$ for the map $W \to \times V$ that assigns to each $w \in W$ the function $(\times f)(w) : I \to \cup V$ such that

$$(\times f)(w)_i = f_i(w).$$

Dually, given any family $f : I \to \mathrm{Hom}(V_i, W)$, we write $\oplus f$ for the map $\oplus V \to W$ that assigns to each $v \in \oplus V$ the sum $\sum_i f_i(v)$.

Example 2.0.2. Suppose $W = V_i = \mathbf{R}$ for $i \in I = \{1, 2\}$, so that $\times V = \oplus V = \mathbf{R}^2$. If f_1 and f_2 are linear maps $\mathbf{R} \to \mathbf{R}$, then

$$\times f : \mathbf{R} \to \mathbf{R}^2 \qquad\qquad (\times f)(x) = (f_1(x), f_2(x))$$

and

$$\oplus f : \mathbf{R}^2 \to \mathbf{R} \qquad\qquad (\oplus f)(x, y) = f_1(x) + f_2(y).$$

Part I

Canonical Forms

Chapter 3

Background: Matroids

This chapter gives a comprehensive introduction to the elements of matroid theory used in this text. It is suggested that the reader skim §3.1-3.4 on a first pass, returning as needed for the small number of applications that require greater detail. Thorough treatments of any of the subjects introduced in this summary may be found in any one of several excellent texts, e.g. [11, 80].

3.1 Independence, rank, and closure

An *independence system* is a mathematical object determined by the data of a set E, called the *ground set*, and a family of subsets of E that is closed under inclusion. The subsets of E that belong to this family are called *independent*, and those that do not are called *dependent*. One requires the empty set to be independent. An independence system satisfies the *Steinitz Exchange Axiom* if for every pair of independent sets I and J such that $|I| < |J|$, there is at least one $j \in J$ such that $I \cup \{j\}$ is independent. The systems that satisfy this property have a special name.

Definition 3.1.1. A *matroid* is an independence system that satisfies the Steinitz Exchange Axiom.

In keeping with standard convention, we write $|\mathcal{M}|$ for the ground set of a matroid $\mathcal{M} = (E, \mathcal{I})$, and $\mathcal{I}(\mathcal{M})$ for the associated family of independent sets.

The pair $(E, \mathcal{I})$, where E is a subset of a vector space and $\mathcal{I}$ is the family of all linearly independent subsets of E, is an archetypal matroid. The fact that $\mathcal{I}$ satisfies the Steinitz Exchange Axiom is the content of the eponymous Steinitz exchange lemma. A closely related example is the family of all sets that *map* to independent sets under an index function $r : E \to V$. Matroids realized in this fashion are called *linear* or *representable*, and the associated index functions are called *linear representations*. Representations that index the columns of a matrix give rise to the term "matroid."

Remark 3.1.2. Every representation functions $r : E \to \mathbf{k}^I$ determines an array $[r] \in I \times E$ such that $[r](i, e) = r(e)_i$. We call this the *matrix representation* of r.

A proper subset of the linear matroids is the family of graphical matroids. Every (finite) undirected graph $G = (V, E)$, where V is the set of vertices and E the set of edges, determines an independence system $(E, \mathcal{I})$, where $\mathcal{I}$ is the family of forests. That this system satisfies the Exchange Axiom follows from the fact that a set of edges forms a forest if and only if the corresponding columns of the node-incidence matrix are linearly independent over the two element field. A matroid is *graphical* if it is isomorphic to one of this form.

Graphical matroids are a field of study in their own right, but also provide helpful intuition at all levels of matroid theory. At the introductory level, especially, recasting the statement a general theorem in terms of a small graphical example, e.g. the complete graph on four vertices, has a useful tendency to demystify abstract results. Several terms used generally throughout matroid theory have graph-theoretic etymologies: an element e of the ground set is a *loop* if the singleton $\{e\}$ is dependent. Two non-loops are *parallel* if the two-element set that contains them both is dependent. A matroid is *simple* if it has no loops and no parallel pairs, i.e., no dependent sets of cardinality strictly less than three.

Some familiar constructions on vector spaces have natural counterparts in matroid theory. To begin with dimension, let us say that an independent set is *maximal* if it is properly contained in no other independent set. We call such sets *bases*, and denote the family of all bases $\mathcal{B}(\mathcal{M})$. It is clear from the Exchange Axiom that every basis has equal cardinality, and this common integer is called the *rank* of the matroid. Since the pair $\mathcal{M}|_S = (S, \{I \in \mathcal{I} : I \subseteq S\})$ is itself a matroid for every $S \subseteq E$, we may extend the notion of rank to that of a *rank function* ρ on the subsets of E.

In linear algebra the span of a subset $S \subseteq V$, denoted $\mathrm{span}(S)$, may be characterized as the set of vectors v such that any maximal independent subset of S is also maximal in $S \cup \{v\}$. The analog for matroids is the collection of all t such that $\rho(S) = \rho(S \cup \{t\})$. We call this the *closure* of S, denoted $\mathrm{cl}(S)$. A subspace of V is a subset that equals its linear span, and, by analogy, a *flat* of a matroid is defined to be a subset that equals its closure. Flats are also called *closed sets*.

3.2 Circuits

A set ζ is called *minimally dependent* if it contains no proper dependent subset, or, equivalently, if every subset of order $|\zeta| - 1$ is independent. Minimally dependent sets are called *circuits*. While the combinatorial definition of a circuit may seem a bit alien at first glance, circuits themselves are quite familiar to students of graph theory and linear algebra. It is quickly shown, for example, that a set C in a graphical matroid forms a circuit if and only if the corresponding edges form a simple cycle in the underlying graph, since any one edge may be dropped from such a set to form a tree. For a fixed basis B of a finite dimensional vector space V, moreover, every vector may be be expressed as a unique linear combination in B. The set

$$\zeta_B(v) = \{v\} \cup \mathrm{supp}(\alpha)$$

where α is the indexed family of scalars such that $v = \sum_{b \in B} \alpha_b$, is called the *fundamental circuit* of v with respect to B. It is simple to check that this is a bona fide circuit in the associated matroid, since its rank and cardinality differ by one, and each $u \in \zeta_B(v)$ lies in the linear span of $\zeta_B(v) - \{u\}$.

More generally, given any element i of a matroid with basis B, one can define the *fundamental circuit of i with respect to B*, denoted either $\zeta_B(i)$ to be the set containing v and those $j \in B$ for which $B - \{j\} \cup \{i\}$ is a basis. It is simple to check that this, also, is a bona fide circuit, since every subset of cardinality $|\zeta_B(i)| - 1$ extends to a basis.

Exercise 3.2.1. Suppose that E is the family of edges in an undirected graph, and $\mathcal{I}$ the family of forests. If B is a spanning forest and $e \in E - B$, then the matroid-theoretic fundamental circuit of e in B agrees with the graph-theoretic fundamental circuit.

It will be useful to have a separate notation for the portion of $\zeta_B(i)$ that lies in B. We call this the *support of i with respect to B*, denoted $\mathrm{supp}_B(i) = \zeta_B(i) - \{i\}$. Let us further set $\mathrm{supp}_B(b) = b$ for any b in B and $\mathrm{supp}_B(S) = \cup_{s \in S} \mathrm{supp}_B(s)$. This notation is not standard, so far as we are aware, but it is highly convenient.

Example 3.2.2. Let $\mathbf{k}$ be the two-element field and N be the 3×6 matrix over $\mathbf{k}$ shown below. Here, for visual clarity, zero elements appear only as blank entries. Let $\mathcal{M}$ be the matroid on ground set $\{a_1, a_2, a_3, b_1, b_2, b_3\}$ for which a subset I is independent iff it indexes a linearly independent set of column vectors. The fundamental circuit of a_1 with respect to basis $B = \{b_1, b_2, b_3\}$ is $\{a_1, b_2, b_3\}$. Its combinatorial support is $\{b_2, b_3\}$. The fundamental circuit of a_3 is $\{a_3, b_1, b_3\}$, and its combinatorial support is $\{b_1, b_3\}$.

$$
\begin{array}{cccccc}
a_1 & a_2 & a_3 & b_1 & b_2 & b_3 \\
\end{array}
$$
$$
\left[\begin{array}{ccc|ccc}
 & 1 & 1 & 1 & & \\
1 & 1 & & & 1 & \\
1 & 1 & 1 & & & 1
\end{array}\right]
$$

The reader will note that in the matrix N of Example 3.2.2, the matroid-theoretic support of each column coincides with its linear support as a vector in $\mathbf{k}^n$, if one identifies the unit vectors b_1, b_2, b_3 with the rows that support them.

In general, any representation that carries $B \subseteq E$ bijectively onto a basis of unit vectors in $\mathbf{k}^B$ is called *B-standard*. Standard representations play a foundational role in the study of linear matroids, due in part to the correspondence between linear and combinatorial supports observed above. This correspondence, which we have observed for one standard representation, evidently holds for all.

One elementary operation in matroid theory is the rotation of elements into and out of a basis. Returning to N, let us consider the operation swapping a_2 for b_2 in B, producing a new basis $C = B - \{b_2\} \cup \{a_2\}$. We can generate a C-standard representation of M by multiplying N on the left with an invertible 3×3 matrix Q, while leaving the column labels unchanged. This multiplication should map a_2 to the second standard unit vector and leave the remaining unit vectors unchanged. These conditions determine that Q should have form

$$
\left[\begin{array}{ccc}
1 & 1 & \\
 & 1 & \\
 & 1 & 1
\end{array}\right]
$$

over the two element field. The resulting representation, M, has form

$$\begin{array}{cccccc} a_1 & a_2 & a_3 & b_1 & b_2 & b_3 \\ \hline 1 & & 1 & 1 & 1 & \\ 1 & 1 & & & 1 & \\ & & 1 & & 1 & 1 \end{array}$$

Note that the fundamental circuit of b_2 with respect to C agrees with the fundamental circuit of a_2 with respect to B.

We define the *support matrix* of a basis B, denoted $\text{Supp}_B \in \{0,1\}^{B \times A}$ by the condition $\text{Supp}_B(i,j) = 1$ iff $i \in \text{supp}_B(j)$. The literature refers to this with the lengthier name *B-fundamental circuit incidence matrix* [**?**]. As we have seen, if $[r] \in \mathbf{k}^{B \times |\mathcal{M}|}$ is the matrix realization of a B-standard representation $r : |\mathcal{M}| \to \mathbf{k}^B$, then $\text{Supp}_B(|\mathcal{M}|)$ is the matrix obtained by switching all nonzero entries of $[r]$ to 1's.

3.3 Minors

A *minor* of a matroid $\mathcal{M}$ is a matroid obtained by a sequence of two elementary operations, called *deletion* and *contraction*. A *deletion minor* of $\mathcal{M}$ is a pair $(S, \mathcal{I}|_S)$, where S is a subset of E and $\mathcal{I}_S = \{I \in \mathcal{I} : I \subseteq S\}$. We call this the *restriction* of $\mathcal{M}$ to S or the minor obtained by *deleting* $E - S$.

The notion of a contraction minor is motivated by the following example. Suppose are given a vector space V, a subspace W, and a surjection $r : V \to U$ with kernel W. We may regard r as the index function of a linear matroid on ground set V, declaring $S \subseteq V$ to be independent if and only if $r(S)$ is independent in U. If we follow this action by deleting S, the resulting matroid $\mathcal{N}$ *contraction minor* of V. The general operation of contraction is an elementary generalization of this construction.

Of note, a subset I is independent in $\mathcal{N}$ if and only if one (respectively, all) of the following three conditions are met: (i) $I \cup J$ is independent, for some basis J of W, (ii) $I \cup J$ is independent for *every* basis of W, and (iii) $I \cup J$ is independent in V for every independent $J \subseteq W$.

Let us see if we can equate these conditions combinatorially. Equivalence of (ii) and (iii) is clear from the definition of an independence system. For (i) and (ii), posit a basis S of W such that $I \cup S$ is linearly independent in V. Any basis T of W may then be extended to a maximal independent subset of $I \cup S \cup T$, by adding some elements of $I \cup S$. None of the elements in this extension can come from S, since otherwise T would fail to be maximal in W. Thus the extended basis includes into $T \cup I$. By dimension counting, it is exactly $T \cup I$. Therefore the union of I with *any* basis of W is independent in V, and all three conditions are equivalent.

This argument does not depend on the fact that W is a subspace of V, or on any algebraic structure whatsoever. For any $W \subseteq E$, then, and any matroid $(E, \mathcal{I})$, we may unambiguously define $\mathcal{I}/W$ to be the family of all I that satisfy one or all of (i), (ii), and (iii). This family is closed under inclusion, and since only elements from I could possibly be transferred from a set of form $S \cup I$ to a set of form $S \cup J$ while preserving independence, satisfaction of the Steinitz exchange axiom by $\mathcal{I}/W$ follows from $\mathcal{I}$. Thus the pair $(E - W, \mathcal{M}/W)$ is a bona fide matroid, the *contraction minor* of $\mathcal{M}$ by $E - W$.

Remark 3.3.1. We will make frequent use of a related matroid, $\mathcal{M} /\!\!/ W = (E, I/W)$, in later sections.

It will be useful to record the following two identities for future reference, while the material is fresh. For any U and W, one has

$$(\mathcal{I}/U)/W = \mathcal{I}/(U \cup W) \qquad \text{and} \qquad \mathcal{I}/U = \mathcal{I}/\mathrm{cl}(U). \tag{3.3.1}$$

The first assertion follows easily from the definition of contraction. One can verify the second is by noting that I is independent in $\mathcal{I}/U$ if and only if $\rho(U \cup I) = \rho(U) + |I|$. This suffices, since $\rho(U \cup I) = \rho(\mathrm{cl}(U) \cup I)$.

A matroid obtained by some sequence of deletion and contraction operations is called a *minor*. It can be shown that distinct sequences will produce the same minor, provided that the each deletes (respectively, contracts) the same set of elements in total. Thus any minor may be expressed in form $(\mathcal{M}/T)|_S$. Where context leaves no room for confusion, we will abbreviate this expression to S/T. More generally, we write S/T to connote the minor $(\mathcal{M}/(T \cap E))|_{(S \cap T \cap E)}$ for sets S and T not contained in E. This will be convenient, for example, when "intersecting" two minors of the same matroid.

3.4 Modularity

Given any pair of subsets $S, T \subseteq E$, one can in general extend a basis B for $S \cap T$ to either a basis $B \cup I$ of S or a basis $B \cup J$ of T. Since $B \cup I \cup J$ then generates $S \cup T$, it follows that

$$\rho(S \cup T) \leq \rho(S) + \rho(T) - \rho(S \cap T). \tag{3.4.1}$$

A real-valued function on the power set of E that satisfies (3.4.1) for any S and T is called *submodular*.

The language of submodularity is motivated by the theory of lattices. If V is a finite-dimensional vector space and $\mathcal{L}$ the associated lattice of linear subspaces (read: matroid flats), then the restriction of ρ to $\mathcal{L}$ coincides exactly with the height function of this lattice. By definition, a lattice with height function function h is *semimodular* if $r(s \vee t) \leq r(s) + r(t) - r(s \wedge t)$ for all s and t, and *modular* if strict equality holds. By way of extension, we say the pair (S, T) is modular with respect to ρ if strict equality holds in (3.4.1). More generally, given any families of sets $\mathcal{S}$ and $\mathcal{T}$, we say that $(\mathcal{S}, \mathcal{T})$ is modular if every $(S, T) \in \mathcal{S} \times \mathcal{T}$ is modular.

It will be useful to have one alternative interpretation of (3.4.1). Since $\rho(S/T) = \rho(S \cup T) - \rho(S)$ and $\rho(T/S \cap T) = \rho(T) - \rho(S \cap T)$, this inequality may be rewritten

$$\rho(S/T) \leq \rho(S/S \cap T). \tag{3.4.2}$$

Strict equality holds if and only if (S, T) is modular. In fact, with modularity more can be said.

Lemma 3.4.1. *If (S, T) is a modular pair, then $S/T = S/(S \cap T)$.*

Proof. Let B be a basis for $S \cap T$, and let I, J be bases for $S/(S \cap T)$ and $T/(S \cap T)$, respectively. Then $B \cup I$ is a basis for S and $B \cup J$ is a basis for T. The union $B \cup I \cup J$ contains a basis for $S \cup T$, and by modularity $|B \cup I \cup J| = \rho(S \cup T)$. Thus $B \cup I \cup J$ is a *basis* for $S \cup T$, so $I \in \mathcal{I}(S/T)$. It follows that the independent sets of $S/(S \cap T)$ are also independent in S/T. Since the converse holds as well, the desired conclusion follows. $\square$

3.5 Basis Exchange

A *based* matroid is a pair $(\mathcal{M}, B)$, where $\mathcal{M} = (E, \mathcal{I})$ is a matroid and B is a basis. A fundamental operation on based matroids is *basis* exchange: the substitution of B in $(\mathcal{M}, B)$ with a basis of form $C = B - I \cup J$.

A *based representation* (with coefficients in field $\mathbf{k}$) of a based matroid is a linear representation $r : E \to \mathbf{k}^B$ such that $r(b) = \chi_b$ for all $b \in B$. We call such representations *B-standard.* Basis exchange induces a canonical transformation of based representations, sending each $r : E \to \mathbf{k}^B$ to the composition $T \circ r$, where T is the linear map $\mathbf{k}^B \to \mathbf{k}^C$ such that $(T \circ r)(c) = \chi_c$ for each element $c \in C$.

In general, if we regard r as an element of $\mathbf{k}^{B \times E}$ with block decomposition

$$\begin{array}{c|cc|} & J & E-J \\ \hline I & \alpha & \beta \\ B-I & \gamma & \delta \\ \hline \end{array}$$

then T may be understood as an element of $\mathbf{k}^{C \times B}$ with block form

$$\begin{array}{c|cc|} & I & B-I \\ \hline J & \alpha^{-1} & 0 \\ B-I & -\gamma\alpha^{-1} & 1 \\ \hline \end{array}$$

where, 1 denotes an identity submatrix of appropriate size. Representation $T \circ r$ will then have form

$$\begin{array}{c|cc|} & J & E-J \\ \hline J & 1 & \alpha^{-1}\beta \\ B-I & 0 & \sigma \\ \hline \end{array}$$

where, by definition, $\sigma = \delta - \gamma\alpha^{-1}\beta$ is the *Schur complement* of α in r. In the special case where $I = \{i\}$ and $J = \{j\}$ are singletons, multiplication on the left with T is precisely the clearing operation that transforms column j to the unit vector supported on i (and subsequently relabels this row as j). By analogy we may regard composition with T as the "block clearing" operation that transforms all the columns of J to unit vectors (and relabels the corresponding rows). This perspective will be developed in later sections.

Base, Rank, and Closure

Based matroids owe their importance in part to the special relationship between the elements of a distinguished basis and those of the ground set generally. It is very easy, for example, to describe the closure of any $I \subseteq B$: it is the collection of all $e \in E$ for which $I \cup \{e\}$ is dependent. If a B-standard representation is available, then we may alternately characterize this set as the collection of all e for which $\mathrm{supp}(r(e)) \subseteq I$. Simultaneously, if we wish to find a basis for I we may take I itself, and to calculate its rank we may take $|I|$.

The situation is more delicate with sets not contained in B. This is one place where elementary exchange offers some help. If, for example, we wish to know the rank of an arbitrary subset S, we may rotate as many elements as possible from S into the standard basis B. This may be done element-by-element via singleton exchange operations, or in sets of elements simultaneously via the "block-clearing" operation. One is forced to stop moving elements of S into B if and only if one of following two equivalent stopping conditions are met (i) the current basis, C, contains a basis for S, (ii) the closure of $S \cap C$ contains S. Once either condition holds, $\rho(S) = |S \cap C|$. This procedure also admits the calculation of $\mathrm{cl}(S) = \mathrm{cl}(S \cap C)$, via the support heuristic used for $\mathrm{supp}(I)$.

Contraction

Elementary exchange offers a means to calculate contraction minors, as well. Recall that the independent sets of $\mathcal{M} /\!\!/ S$ are those I for which $I \cup J$ is independent, for some (equivalently, every) basis J of S. Let us suppose that I is given, together with a basis B containing I and a B-standard representation r. Since r carries I to a collection of unit vectors, it is easy to see that $q \circ r$ represents $\mathcal{M} /\!\!/ I$, where $q : \mathbf{k}^B \to \mathbf{k}^{B-I}$ is the canonical deletion operator.

Given r, can in general leverage the fact that

$$\mathcal{M} /\!\!/ S = \mathcal{M} /\!\!/ \mathrm{cl}(I) = \mathcal{M} /\!\!/ I$$

for $I \in \mathcal{B}(S)$ to compute a representation of $\mathcal{M} /\!\!/ S$, for any S. To do so, simply rotate an $I \in \mathcal{B}(S)$ into the standard basis by elementary exchange, and apply the deletion operator q. To obtain a representation for the formal contraction $\mathcal{M}/C$, restrict the resulting representation to $E - S$.

Deletion

The role of deletion in standard representations is in many ways dual to that of contraction. The simplest case holds when one wishes to delete a set I that is *disjoint from* the standard basis B. In this case one may simply restrict the representation r to $E - I$. To delete an arbitrary S requires slightly more care, as deletion of elements from the standard basis may result in a non-standard representation.

To address this issue, first rotate as many elements of S as possible *out* of the standard basis. There is only one condition under which some elements of S will remain in the resulting basis, C, namely that the corresponding representation r has the following block matrix form

$$
\begin{array}{c|ccc|}
 & S \cap C & S - C & E - C - S \\
\hline
S \cap C & 1 & * & 0 \\
C - S & 0 & * & * \\
\hline
\end{array}
$$

where asterisks denote blocks of indeterminate form. Consequently, if $q : \mathbf{k}^C \to \mathbf{k}^{C-S}$ is the standard deletion operator, then $(q \circ r)|_{E-S}$ will represent $\mathcal{M} - S$ with respect to the standard basis $C - S$.

3.6 Duality

The *dual* of a matroid $\mathcal{M} = (E, \mathcal{I})$, denoted $\mathcal{M}^* = (E, \mathcal{I}^*)$, is the matroid on ground set E for which a subset S is independent if and only if the complement of S in E contains a basis. Bases of the dual matroid are called *cobases* of $\mathcal{M}$.

Duality is integral to the study of matroids general, and to our story in particular. We will introduce terminology for this structure as needed throughout the text.

Chapter 4

Modularity

4.1 Generators

Recall from Section 3.4 that the rank function ρ of a matroid $\mathcal{M} = (E, \mathcal{I})$ is *submodular*, meaning

$$\rho(S \cup T) + \rho(S \cap T) \leq \rho(S) + \rho(T) \tag{4.1.1}$$

for any $S, T \subseteq E$. We say that an unordered pair $\{S, T\}$ is *modular* if strict equality holds in (4.1.1). More generally a family $\mathcal{S}$ is modular if $\{S, T\}$ is modular for all $S, T \in \mathcal{S}$.

Remark 4.1.1. Every pair of linear subspaces $S, T \subseteq W$ form a modular pair in the matroid associated to W. This pair is defined to be *transverse* if

$$\rho(W) = \rho(S) + \rho(T) - \rho(S \cap T).$$

The tameness conditions imposed by modularity on set families is analogous and indeed closely linked to the tameness conditions imposed by transversality in bundle theory and geometric topology[39].

A set I *spans* a set S if $S \subseteq \mathrm{cl}(I)$, and *generates* S if $I \cap S$ spans S. If in addition I is independent, we say I *freely generates* S. By extension, I spans (respectively, generates, freely generates) a family $\mathcal{S}$ if I spans (respectively, generates, freely generates) each $S \in \mathcal{S}$. Free generation has a basic relation to modularity.

Proposition 4.1.2. *Freely generated set families are modular.*

Proof. If S and T are any two sets generated by an independent set I, and if I_U denotes $I \cap U$ by for arbitrary U, then $|I_S| + |I_T| - |I_{S \cup T}| = |I_{S \cap T}|$. As the sets on the left hand side are bases of S, T, and $S \cup T$, respectively, one has

$$\rho(S) + \rho(T) - \rho(S \cup T) = |I_{S \cap T}| \leq \rho(S \cap T).$$

Submodularity implies that the opposite estimate holds as well, so strict equality holds throughout. $\square$

The converse to Proposition 4.1.2 is in general false: if u, v, and w are points in the Euclidean plane and none is a scalar multiple of another, then the family $\{\{u\},\{v\},\{w\}\}$ is modular, but cannot be generated by an independent set.

Let us describe some additional conditions necessary for free generation. Fix any freely generated set family $\mathcal{S}$ and, to avoid pathologies, assume $\mathcal{S}$ to be finite. The proof of Proposition 4.1.2 shows that $\mathcal{S}\cup\{S\cap T\}$ is freely generated for $S,T\in\mathcal{S}$, so we may assume that $\mathcal{S}$ is closed under intersection. It is evidently closed under union, so by hypothesis there exists an independent I for which $I_{\cap_{T\in\mathcal{T}}T} - I_{\cup_{U\in\mathcal{U}}U}$ freely generates

$$(\cap_{T\in\mathcal{T}}T)/(\cup_{U\in\mathcal{U}}U) \tag{4.1.2}$$

for arbitrary $\mathcal{T},\mathcal{U}\subseteq\mathcal{S}$. Let $\mathcal{M}_{\mathcal{T}}$ denote the matroid minor (4.1.2) with $\mathcal{U}=\mathcal{S}-\mathcal{T}$.

Proposition 4.1.3. *A set family $\mathcal{S}$ is freely generated if and only if the matroid union $\cup_{\mathcal{T}\subseteq S}\mathcal{M}_{\mathcal{T}}$ has rank equal to $\mathcal{M}$.*

Proof. To each element e of $\mathcal{M}$ corresponds exactly one $\mathcal{T}$ such that $e\in\mathcal{M}_{\mathcal{T}}$, namely $\mathcal{T}=\{T\in\mathcal{S}: e\in T\}$. The family of all $E_{\mathcal{T}}$, where $E_{\mathcal{T}}$ is the ground set of $\mathcal{M}_{\mathcal{T}}$, has several helpful properties. First, it forms a pairwise disjoint partition of the ground set of $\mathcal{M}$. Second, each $\mathcal{M}_{\mathcal{T}}$ may be expressed as as a minor of form $E_{\mathcal{T}}/(\cup_{\mathcal{U}\neq\mathcal{T}}E_{\mathcal{U}})$ in $\mathcal{M}$, so, by definition of contraction, any union of independent sets drawn from $\{\mathcal{M}_{\mathcal{T}}:\mathcal{T}\subseteq\mathcal{S}\}$ will be independent in $\mathcal{M}$. In symbols,

$$\mathcal{I}(\cup_{\mathcal{T}}\mathcal{M}_{\mathcal{T}})\subseteq\mathcal{I}(\mathcal{M}). \tag{4.1.3}$$

Third and finally, each $S\in\mathcal{S}$ may be expressed as the union of some $E_{\mathcal{T}}$'s. Let such an S be given. Inclusion (4.1.3) implies that the rank of S is at least as high in $\mathcal{M}$ as it is in $\cup_{\mathcal{T}}\mathcal{M}_{\mathcal{T}}$, so

$$\rho((\cup_{\mathcal{T}}\mathcal{M}_{\mathcal{T}})/S)\geq\rho(\mathcal{M}/S). \tag{4.1.4}$$

On the other hand, if $\mathcal{N}=\mathcal{M}\mathbin{/\!\!/}S$ then $(\cup_{\mathcal{T}}\mathcal{M}_{\mathcal{T}})/S=\cup_{\mathcal{T}}\mathcal{N}_{\mathcal{T}}$. Since $\mathcal{I}(\cup_{\mathcal{T}}\mathcal{N}_{\mathcal{T}})\subseteq\mathcal{I}(\mathcal{N})$, it follows that strict equality holds in (4.1.4). Consequently, when $\mathcal{M}$ and $\mathcal{M}_{\mathcal{T}}$ have equal rank, S has equal rank in both matroids. In particular, any basis of $\cup_{\mathcal{T}}\mathcal{M}_{\mathcal{T}}$ will freely generate $\mathcal{S}$. Conversely, if there exists a basis $B\in\mathcal{B}(\mathcal{M})$ that freely generates $\mathcal{S}$, then, as argued above, B is independent in $\cup_{\mathcal{T}}\mathcal{M}_{\mathcal{T}}$. In fact, B is a basis in $\cup_{\mathcal{T}}\mathcal{M}_{\mathcal{T}}$, by (4.1.3). Thus the two matroids have equal rank. $\square$

The matroid union in Proposition 4.1.3 will reappear sufficiently often to warrant special notation: for any set family $\mathcal{S}$, we will set

$$\mathcal{M}/\mathcal{S}=\cup_{\mathcal{T}\subseteq\mathcal{S}}\mathcal{M}_{\mathcal{T}}.$$

Let us consider this object for two specific types of $\mathcal{S}$.

Filtrations

A (finite) *filtration* $\mathcal{S}$ is nested sequence of sets $S_1\subseteq\cdots\subseteq S_m$, where S_m is the ground set. The nonempty minors that make up $\mathcal{M}/\mathcal{S}$ are those of form S_p/S_{p-1}. Since the rank of S_p/S_{p-1} equals $\rho(S_p) - \rho(S_{p-1})$, the ranks of $\mathcal{M}$ and $\mathcal{M}/\mathcal{S}$ agree for every filtration $\mathcal{S}$. Thus filtrations are both modular and freely generated, though this is easy enough to show directly – one can always build a basis that generates $\mathcal{S}$ by first finding a basis for S_1, extending to a basis for S_2, et cetera.

There is a one to one correspondence between length m filtrations and functions $\mathcal{F} : E \to \{1, \ldots, m\}$. In one direction this correspondence sends $\mathcal{F}$ to the sequence $(\mathcal{F}_{\leq p})_{p=1}^{m}$, where

$$\mathcal{F}_{\leq p} = \{i : \mathcal{F}(i) \leq p\}.$$

In the opposite direction one maps $\mathcal{S}$ to the unique integer-valued function so that $S_p = \mathcal{F}_{\leq p}$. It will be convenient to identify functions with filtrations under this correspondence, so that one may speak, for example, of the value of a filtration $\mathcal{S}$ on an element e, and of the matroid $\mathcal{M}/\mathcal{F}$ without ambiguity. Where context leaves no room for confusion we will accordingly write $\mathcal{F}_p$ for $\mathcal{F}_{\leq p}$.

Bifiltrations

A *bifiltration* is a union of form $\mathcal{S} = \mathcal{F} \cup \mathcal{G}$, where $\mathcal{F}$ and $\mathcal{G}$ are filtrations. The minors that make up $\mathcal{M}/\mathcal{S}$ are those of form

$$(\mathcal{F}_p \cap \mathcal{G}_q)/(\mathcal{F}_{p-1} \cup \mathcal{G}_{q-1}).$$

Such unions need not be modular or freely generated. Unlike arbitrary families, however, they must be both if they are either.

Proposition 4.1.4. *A bifiltration is freely generated if and only if it is modular.*

Proof. One implication has already been established by Proposition 4.1.2. For the converse, suppose $\mathcal{S}$ to be modular. Recall from the background section on modularity that $\mathcal{F}_p/\mathcal{G}_{q-1} = \mathcal{F}_p/(\mathcal{F}_p \cap \mathcal{G}_{q-1})$, so that, in particular,

$$(\mathcal{F}_p \cap \mathcal{G}_q)/\mathcal{G}_{q-1} = (\mathcal{F}_p \cap \mathcal{G}_q)/(\mathcal{F}_p \cap \mathcal{G}_{q-1}).$$

The rank of the righthand side is $\rho(\mathcal{F}_p \cap \mathcal{G}_q) - \rho(\mathcal{F}_p \cap \mathcal{G}_{q-1})$, so a sum over q telescopes to $\rho(\mathcal{F}_p)$. The left hand side is isomorphic to the restriction of $\mathcal{N}_q = \mathcal{G}_q/\mathcal{G}_{q-1}$ to $\mathcal{F}_p \cap |\mathcal{N}_q|$, so $\mathcal{F}_p$ has the same rank in $\mathcal{M}$ that it has in $\mathcal{M}/\mathcal{G} = \cup_q \mathcal{N}_q$. Any basis that generates $\mathcal{F}$ in the latter matroid will therefore generate both filtrations in the former. □

N-filtrations

It is a simple matter to show that a union of $N \geq 2$ filtrations may be modular without being freely generated, even in the special case where each is a linear filtration on a vector space. Given three distinct lines ℓ_1, ℓ_2, ℓ_3 in the Euclidean plane, for example, the family $\{\ell_1, \ell_2, \ell_3\}$ is modular, but cannot be generated by an independent set.

4.2 Minimal Bases

An object of basic interest in the study of matroids is the weighted basis. Given any real-valued function $\mathcal{F}$ on E, we may define the $\mathcal{F}$*-weight* of a finite subset $S \subseteq E$ to be the sum $\sum_S \mathcal{F}(s)$. A basis is $\mathcal{F}$*-minimal* (respectively, $\mathcal{F}$-maximal) if its weight is no greater (respectively, no less) than that of any other basis. Here and throughout the remainder of this text, we will assume that weight functions take finitely many values, hence minimal bases will always exist.

It is an important property of minimal bases that they determine a matroid independence system in their own right. By extension of the notation introduced in the preceding section, let us identify $\mathcal{F}$ with the family of sublevel sets

$$\{\mathcal{F}_{\leq \varepsilon} : \varepsilon \in \mathbf{R}\}$$

so that $\mathcal{M}/\mathcal{F} = \cup_{\varepsilon \in \mathbf{R}} \mathcal{F}_\varepsilon / \mathcal{F}_{<\varepsilon}$.

Lemma 4.2.1. *The $\mathcal{F}$-minimal bases of $\mathcal{M}$ are the bases of $\mathcal{M}/\mathcal{F}$. Equivalently, they are the bases that generate $\mathcal{F}$.*

Proof. Fix a basis B, and suppose there exists an ε for which $B_{\mathcal{F}_{\leq \varepsilon}}$ is not a basis of $\mathcal{F}_{\leq \varepsilon}$. There exists at least one e in $\mathcal{F}_{\leq \varepsilon}$ such that $B_{\mathcal{F}_{\leq \varepsilon}} \cup \{e\}$ is independent, and this union may be extended to a basis by adding some elements of $B_{\mathcal{F}_{>\varepsilon}}$. The new basis will be identical to B, except that one element of $B_{\mathcal{F}_{>\varepsilon}}$ will be replaced by $e \in \mathcal{F}_{\leq \varepsilon}$. The old basis evidently outweighs the new, so B is not minimal. Therefore $B_{\mathcal{F}_{\leq \varepsilon}}$ freely generates $\mathcal{F}_{\leq \varepsilon}$, whenever B is minimal. The converse is immediate, since every basis that satisfies this criterion has identical weight. $\square$

Chapter 5

Canonical Forms

This chapter introduces a combinatorial abstraction of the idea of a Jordan basis for a nilpotent operator. This is not the first combinatorial model of nilpotent canonical forms. A closely related notion of nilpotent Jordan basis for complete $\vee$-homomorphisms on atomistic lattices was introduced by J. Szigeti in [77]. The idea of Szigeti was to understand operators via their action on the lattice of subspaces of the domain. In this formulation, orbits of vectors are replaced by orbits of one-dimensional subspaces, or more generally *atoms*, and bases are replaced by families of atoms that join irredundantly to the maximum element of the lattice.

While the language of posets dramatically increases the reach of traditional canonical forms, many of the properties one would typically wish for break down in this more general context. Of particular significance, distinct bases in a lattice that lacks the Jordan-Hölder property may have different cardinalities. Moreover, while a sufficient condition for the existence of Jordan bases is given in [77], no structural characterization is given. Our main result states that in any lattice where bases satisfy the Exchange Axiom, an exact condition for existence can be given, and the corresponding Jordan bases may be characterized combinatorially.

5.1 Modular Filtrations

In Lemma 5.1.1 and Proposition 5.1.2 below, we assume a pair of filtrations $\mathcal{F}$ and $\mathcal{G}$ on the ground set of a matroid $\mathcal{M}$. We say that a basis is $\mathcal{F}$-$\mathcal{G}$ minimal if it is minimal with respect to both $\mathcal{F}$ and $\mathcal{G}$.

Lemma 5.1.1. *A basis B is $\mathcal{F}$-$\mathcal{G}$ minimal if and only if*

$$B \cap (\mathcal{F}_i \cap \mathcal{G}_j) \in \mathcal{B}(\mathcal{F}_i \cap \mathcal{G}_j) \tag{5.1.1}$$

for all i, j.

Proof. If B satisfies (5.1.1) for all i, j then minimality with respect to $\mathcal{F}$ follows from an application of Lemma 4.2.1 to the family of intersections $B \cap \mathcal{F}_i \cap E$. Minimality with respect to $\mathcal{G}$ follows likewise. If on the other hand $S = B \cap (\mathcal{F}_i \cap \mathcal{G}_j) \notin \mathcal{B}(\mathcal{F}_i \cap \mathcal{G}_j)$ for some i, j, then there exists $s \in \mathcal{F}_i \cap \mathcal{G}_j$ such that $S + s \in \mathcal{I}(M)$. The fundamental circuit of s with respect to B intersects $B - S$ nontrivially, hence $B - b + s \in \mathcal{B}(M)$ for some $b \in B - S$. Either $\chi_{\mathcal{F}}(s) < \chi_{\mathcal{F}}(b)$ or $\chi_{\mathcal{G}}(s) < \chi_{\mathcal{G}}(b)$, so B is not $\mathcal{F}$-$\mathcal{G}$-minimal. □

Proposition 5.1.2. *The union $\mathcal{F} \cup \mathcal{G}$ is modular if and only if there exists an $\mathcal{F}$-$\mathcal{G}$-minimal basis of $\mathcal{M}$.*

Proof. The "if" portion follows from Lemma 4.2.1 and Proposition 4.1.2. Therefore assume $(\mathcal{F}, \mathcal{G})$ is modular, and for each i fix a $\chi_{\mathcal{G}}$-minimal $B_i \in \mathcal{B}(\mathcal{F}_i/\mathcal{F}_{i-1})$.

The union $B = \cup_i B_i \in \mathcal{B}(\mathcal{M})$ forms a $\mathcal{F}$-minimal basis in $\mathcal{M}$, so it suffices to show B is minimal with respect to $\mathcal{G}$. Since $|B \cap G_j| \leq \rho(G_j)$, we may do so by proving $|B \cap \mathcal{G}_j| \geq \rho(\mathcal{G}_j)$ for all j. Therefore let i and j be given, fix $S_{i-1} \in \mathcal{B}(\mathcal{F}_{i-1} \cap \mathcal{G}_j)$ and extend S_{i-1} to a basis S_i of $\mathcal{F}_i \cap \mathcal{G}_j$.

Modularity provides the first identity below: for all i and j,

$$\rho(\mathcal{G}_j/\mathcal{F}_i) = \rho(\mathcal{G}_j/(\mathcal{F}_i \cap \mathcal{G}_j)) = \rho(\mathcal{G}_j/S_i) \leq \rho(\mathcal{G}_j/(S_i \cup \mathcal{F}_{i-1})) \leq \rho(\mathcal{G}_j/\mathcal{F}_i).$$

As the left and right hand sides are identical, strict equality holds throughout. Hence the second identity below.

$$\begin{aligned}\rho(\mathcal{G}_j/S_{i-1}) - \rho(S_i/S_{i-1}) &= \rho(\mathcal{G}_j/S_i) \\ &= \rho(\mathcal{G}_j/(S_i \cup \mathcal{F}_{i-1}) = \rho(\mathcal{G}_j/\mathcal{F}_{i-1}) - \rho(S_i/\mathcal{F}_{i-1}).\end{aligned}$$

A comparison of left- and right-hand sides shows $\rho(S_i/S_{i-1}) = \rho(S_i/\mathcal{F}_{i-1})$. Since the set $T = S_i - S_{i-1}$ forms a basis in S_i/S_{i-1}, one has

$$|T| = \rho(S_i/S_{i-1}) = \rho(S_i/\mathcal{F}_{i-1}) = \rho(T/\mathcal{F}_{i-1})$$

and therefore $T \in \mathcal{I}(\mathcal{F}_i/\mathcal{F}_{i-1})$. Thus $|T| \leq |B_i \cap \mathcal{G}_j|$. A second and third application of modularity provide the second and third equalities below,

$$\rho(\mathcal{G}_j/S_{i-1}) - \rho(\mathcal{G}_j/S_i) = \rho(\mathcal{G}_j/\mathcal{F}_{i-1}) - \rho(\mathcal{G}_j/\mathcal{F}_i) = \rho(\mathcal{G}_j \cap \mathcal{F}_i) - \rho(\mathcal{G}_j \cap \mathcal{F}_{i-1}).$$

Since the left hand side agrees with $|T|$, it follows that $|B_i \cap \mathcal{G}_j|$ dominates $\rho(\mathcal{G}_j \cap \mathcal{F}_i) - \rho(\mathcal{G}_j \cap \mathcal{F}_{i-1})$. Summing over i yields $|B \cap \mathcal{G}_j| \geq \rho(\mathcal{G}_j)$, which was to be shown. □

5.2 Nilpotent Canonical Forms

We say a set function $T : E \to E$ is $\vee$-*complete* if $T(\mathrm{cl}(S)) \subseteq \mathrm{cl}(TS)$ for every $S \subseteq E$. If $\mathcal{M}$ is a simple matroid, then T is $\vee$-complete if and only if the rule

$$F \mapsto \mathrm{cl}(TF)$$

determines a $\vee$-complete homomorphism on the associated lattice of flats.

Several properties of $\vee$-complete maps are immediate from the definition. First, T is $\vee$-complete iff T^m is $\vee$-complete for all nonnegative m, since one can "pass" sequential copies of T across the closure operator to form an increasing sequence of sets ranging from $T^m\mathrm{cl}(S)$ to $\mathrm{cl}(T^m S)$, for any S.

Second, suppose that $\mathcal{M}$ contains a loop 0, and define the *kernel* of T by $\mathrm{K}(T) = \{e \in E : T(e) \in \mathrm{cl}(0)\}$. This is the natural analog to the notion of a kernel introduced in [77]. One may argue that

$$\rho(TS) \leq \rho(S/\mathrm{K}(T)) \tag{5.2.1}$$

for any subset S, as follows. Extend a basis, J, of $\mathrm{K}(T)$ to a basis $I \cup J$ of $S \cup \mathrm{K}(T)$. One then has $\rho(TS) = \rho(T(S \cup \mathrm{K}(T))) = \rho(TI)$. The righthand side is bounded by $|I| = \rho(S/\mathrm{K}(T))$, hence (5.2.1). We will say that T is *complementary* if strict equality holds for every subset S.

In the special case where S is the entire ground set, (5.2.1) implies that

$$\rho(\mathrm{K}(T)) + \rho(\mathrm{I}(T)) \leq \rho(\mathcal{M}). \tag{5.2.2}$$

Strict equality holds in (5.2.2) when T is complementary, and in fact, the converse holds as well. Why? If I is a basis for $\mathrm{K}(T)$, $I \cup J$ is a basis for $I \cup S$, and $I \cup J \cup B$ is a basis for M, then equality in (5.2.2) implies the middle identity in

$$\rho(T(J \cup B)) = \rho(\mathrm{I}(T)) = \rho(\mathcal{M}) - |I| = |J \cup B|.$$

Thus J is independent, so $\rho(TS) = |J| = \rho(S/\mathrm{K}(T))$.

We will say that T is *nilpotent* if $T^n \subseteq \mathrm{cl}(0)$ for some n. The *orbit* of an element e under a nilpotent operator T is the set (possibly empty) of all nonzero elements that may be expressed in form $T^m e$, for some nonnegative m. A basis is *Jordan* with respect to T if it may be expressed as the disjoint union of some T-orbits.

A reasonable point of departure for the study of Jordan bases is the relation between kernels, images, and nilpotent maps. Provided $T^n = 0$ one may define an increasing sequence of kernels

$$\mathcal{K}: \quad \mathrm{K}(T^0) \subseteq \cdots \subseteq \mathrm{K}(T^n)$$

and one of images

$$\mathcal{I}: \quad \mathrm{I}(T^n) \subseteq \cdots \subseteq \mathrm{I}(T^0),$$

each terminating with the ground set. We call these the *kernel* and *image* filtrations, respectively, of T.

Remark 5.2.1. Our notation for the image filtration is slightly unfortunate, as it conflicts with the universal convention that $\mathcal{I}$ should denote a family of independent sets. The benefit of this minot transgression is the clarity it lends to certain duality results, c.f. Corollary 5.2.6.

Proposition 5.2.2. *Every Jordan basis of a nilpotent $\vee$-complete operator T freely generates the kernel and image filtrations of T.*

Proof. Suppose J is a Jordan basis of T. Since $J_{\mathcal{K} \leq m}$ and $T^m J_{\mathcal{K} > m}$ are independent subsets of $\mathrm{K}(T^m)$ and $\mathrm{I}(T^m)$, respectively, inequality (5.2.2) implies that $J_{\mathcal{K} \leq m}$ generates $\mathrm{K}(T^m)$, and $T^m J_{\mathcal{K} > m}$ generates $\mathrm{I}(T^m)$. □

Proposition 5.2.3. *A nilpotent $\vee$-complete operator T has a Jordan basis if and only if T^m is complementary for every nonnegative integer m.*

Proof. The "if" portion follows from Proposition 5.2.4 below. The "only if" follows from proof of Proposition 5.2.2, where we showed that every Jordan basis may be partitioned into two disjoint subsets, one generating the kernel of T^m and the other its image. $\square$

Proposition 5.2.4. *If T^m is complementary for all nonnegative m, then the Jordan bases of T are exactly the orbits of $\mathcal{K}$-minimal basis in $\mathcal{M}/\mathrm{I}(T)$.*

Proof. That every Jordan basis may be expressed as the orbit of a $\mathcal{K}$-minimal basis in $\mathcal{M}/\mathrm{I}(T)$ follows from Lemma 4.2.1, which characterizes the minimal bases of a filtration as those that freely generate it, and Proposition 5.2.2. For the converse, let us suppose that E is the orbit a $\mathcal{K}$-minimal basis B in $\mathcal{M}/\mathrm{I}(T)$, and show E to be a basis.

To establish independence, assume for a contradiction that E contains a circuit ζ. Fix an integer m and a subset $\omega \subseteq E$ such that $\zeta = T^m\omega$ and $\omega \cap B$ is nonempty. Since $z \in \zeta$ lies in the closure of $\zeta - \{z\}$, complementarity implies that $w \in \omega$ lies in the closure of $\omega - \{w\}$ in $\mathcal{M}/\mathcal{K}_{\leq m}$. Since $\mathcal{K}$ takes values strictly greater than m on ω, it follows *a fortiori* that w lies in the closure of $\omega - \{w\}$ in $\mathcal{M}/(\mathcal{K}_{\leq w} \cup \mathrm{I}(T))$. However, if we take $w \in \omega \cap B$ to have maximum $\mathcal{K}$-weight this implies a contradiction, since w includes into a $\mathcal{K}$-basis in $\mathcal{M}/\mathrm{I}(T)$. Thus E is independent.

To see that E spans $\mathcal{M}$, let $\mathcal{N}$ be the matroid obtained by introducing a unique zero element into the minor $\mathcal{M}/\mathrm{cl}(E)$. Evidently, T induces a nilpotent map Q on $\mathcal{N}$ that sends e to Te if Te lies outside $\mathrm{cl}(E)$, and to zero otherwise. Since $\mathrm{cl}_{\mathcal{N}}(S) = \mathrm{cl}(E \cup S) - E$ for any S, an element $j \in \mathcal{M} - E$ belongs to $T\mathrm{cl}_{\mathcal{N}}(S)$ if and only if it lies in

$$T(\mathrm{cl}(E \cup S)) \subseteq \mathrm{cl}(T(E \cup S)) \subseteq \mathrm{cl}(E \cup TS).$$

Inclusion in the righthand side is equivalent to membership in $\mathrm{cl}_{\mathcal{N}}(TS)$, for j outside E, so $Q\mathrm{cl}_{\mathcal{N}}(S) \subseteq \mathrm{cl}_{\mathcal{N}}(QS)$. In particular, Q is $\vee$-complete.

As Q is evidently nilpotent, it follows that either $\rho(Q\mathcal{N}) < \rho(\mathcal{N})$ or $\mathcal{N}$ has rank zero. If the former holds, then $\mathcal{N}/Q\mathcal{N}$ will have positive rank. This is impossible, since the independent sets of $\mathcal{N}/Q\mathcal{N}$ are exactly those of $\mathcal{M}/(E \cup T\mathcal{M})$, and the latter has rank zero. Therefore $\mathcal{N}$ has rank zero, whence E is a basis. $\square$

Jordan bases admit a natural dual characterization. Let us say that $\{e, \ldots, T^m e\}$ is a *preorbit* of $T^m e$ if there exists if

$$e \in T^m\mathcal{M} - T^{m+1}\mathcal{M}.$$

A preorbit of a set S is a union of form $\cup_{s \in S} J_s$, where for each $s \in S$ the set J_s is a preorbit of s. The proof of the following observation is entirely analogous to that of Proposition 5.2.4. The details are left as an exercise to the reader.

Proposition 5.2.5. *If T^m is complementary for all nonnegative m, then the Jordan bases of T are the preorbits of $\mathcal{I}$-maximal basis in* $\mathrm{K}(T)$*.*

In summary, we have the following.

Corollary 5.2.6. *If $\mathcal{M}$ is a matroid and*

$$T : E \to E,$$

is a nilpotent $\vee$-complete operator on the ground set of $\mathcal{M}$, then following are equivalent.

1. *T has a Jordan basis.*
2. *T^m is complementary, for all m.*
3. *The Jordan bases of T are the orbits of $\mathcal{K}$-minimal bases in $\mathcal{M} \,/\, \mathrm{I}(T)$.*
4. *The Jordan bases of T are the preorbits of $\mathcal{I}$-maximal bases in $\mathcal{M} \mid \mathrm{K}(T)$.*

The fourth and final characterization is encountered quite often in practice. The procedure for finding a nilpotent Jordan basis outlined in §8.2, for example, may be understood as concrete application of the classical greedy algorithm for matroid optimization to the problem of finding a $\mathcal{I}$-maximal basis in $\mathrm{K}(T)$. The elements of this argument are not new. The basic elements were recorded at least as early as 1956 [72], and have been revisited frequently over the following decades, e.g. [1, 27, 28, 50, 73], though to our knowledge none has recognized that the problem being solved was one of matroid optimization.

Let us say that an orbit I is *maximal with respect to inclusion* if there exists no orbit J such that $I \subseteq J$ and $I \neq J$.

Corollary 5.2.7 (Uniqueness)**.** *Suppose that $(I_1, \ldots, I_m)$ and $(J_1, \ldots, J_n)$ are pairwise disjoint families of maximal orbits for which*

$$I_1 \cup \cdots \cup I_m \qquad \text{and} \qquad J_1 \cup \cdots \cup J_n$$

are Jordan bases. Then there exists a bijection $\varphi : \{1, \ldots, m\} \to \{1, \ldots, n\}$ such that $|I_p| = |J_{\varphi(p)}|$ for all p.

Proof. If $\cup_p I_p$ is a Jordan basis then

$$I_p = \mathrm{Orb}(\psi(p)) \qquad p \in \{1, \ldots, m\}$$

for some $\mathcal{K}$-minimal basis B in $\mathcal{M}/\mathrm{I}(T)$ and some bijection $\psi : \{1, \ldots, m\} \to B$. Thus the number of orbits of given length in each Jordan basis is uniquely determined. □

5.3 Graded Canonical Forms

A $\mathbb{Z}$-*grading* on a matroid $\mathcal{M}$ is a function $\mathcal{H}$ that assigns a flat of $\mathcal{M}$ to each integer p, subject to the condition that

$$\mathrm{cl}\left(\bigcup_p \mathcal{H}_p\right) = \mathcal{M} \qquad \text{and} \qquad \sum_p \rho(\mathcal{H}_p) = \rho(\mathcal{M}).$$

Recalling that $\mathrm{cl}(0)$ denotes the possibly empty family of *loops* in $\mathcal{M}$, we write Γ for the integer-valued function on $\cup_p \mathcal{H}_p - \mathrm{cl}(0)$ such that

$$\Gamma(e) = p \qquad e \in \mathcal{H}_p - \mathrm{cl}(0).$$

Example 5.3.1. If $\mathcal{M} = (V, \mathcal{I})$, where V is a $\mathbf{k}$-linear vector space and $\mathcal{I}$ is the family of $\mathbf{k}$-linearly independent subsets of V, then the $\mathbb{Z}$-gradings of $\mathcal{M}$ are the $\mathbb{Z}$-indexed families of subspaces $\mathcal{H}$ such that V is the internal direct sum of $\{\mathcal{H}_p : p \in \mathbb{Z}\}$.

A map $T : \mathcal{M} \to \mathcal{M}$ is *graded of degree k* if

$$T\mathcal{H}_p \subseteq \mathcal{H}_{p+k}$$

for all p. Unless otherwise indicated, we will write *graded* for graded of degree one. Here and throughout the remainder of this discussion we will assume that $\mathcal{M}$ has finite rank, so that all graded graded maps on $\mathcal{M}$ are nilpotent.

A subset $I \subseteq \mathcal{M}$ is *graded* if $I \subseteq \cup_p \mathcal{H}_p$. To every orbit J_q in a graded Jordan basis J corresponds an integer interval,

$$\mathrm{Supp}(J_q) = \{p : J_q \cap \mathcal{H}_p \neq \emptyset\}.$$

The associated multiset

$$\{\mathrm{Supp}(J_q) : q = 1, \ldots, m\},$$

where $J_1, \ldots, J_m$ are the orbits that compose J, is the *barcode* of J.

Proposition 5.3.2 states that the barcode of a graded Jordan basis for T is uniquely determined by T.

Proposition 5.3.2. *If $I_1, \ldots, I_m$ and $J_1, \ldots, J_n$ are the orbits that compose two graded Jordan bases, then there exists a bijection $\varphi : \{1, \ldots, m\} \to \{1, \ldots, n\}$ such that*

$$\mathrm{Supp}(I_p) = \mathrm{Supp}(J_{\varphi(p)})$$

for each p in $\{1, \ldots, m\}$.

Proof. Every Jordan basis is the orbit of a $\mathcal{K}$-minimal basis of $\mathcal{M}/\mathrm{I}(T)$. A graded Jordan basis, therefore, is the orbit of a $\mathcal{K}$-minimal basis in

$$(\cup_p \mathcal{H}_p)/\mathrm{I}(T) = \cup_p(\mathcal{H}_p/\mathrm{I}(T)).$$

The orbits of any two such bases will determine identical barcodes. □

The remainder of this section will be devoted to a class of graded nilpotent maps with a particularly simple combinatorial structure. Fix filtrations $\mathcal{Z}$, $\mathcal{B}$ on a finite-rank matroid $\mathcal{M}$. Assume that the elements of these filtrations are closed, and that $\mathcal{B}_p \subseteq \mathcal{Z}_p$ for $p \in \mathbb{Z}$. Define

$$\mathcal{H}_p = \mathcal{Z}_p \mathbin{/\!\!/} \mathcal{B}_p,$$

and let $\mathcal{N}$ be the matroid union $\cup_p \mathcal{H}_p$. For each e in the ground set of $\mathcal{M}$, write $e^{(p)}$ for the copy of e in $\mathcal{H}_p \subseteq \mathcal{N}$. Finally, let $T : \mathcal{N} \to \mathcal{N}$ be the function sending $e^{(p)}$ to $e^{(p+1)}$. Given this data, it is natural to ask when T engenders a Jordan basis, or equivalently, when the powers of T are complementary.

Lemma 5.3.3. *If m is a nonnegative integer, then T^m is complementary if and only if $(\mathcal{Z}_p, \mathcal{B}_{p+m})$ is modular, for every p.*

Proof. Let us write $\mathcal{Z}^{(p)}$ and $\mathcal{B}^{(p)}$ for the filtrations on $\mathcal{H}_p$ engendered by $\mathcal{Z}$ and $\mathcal{B}$. Since the sublevel sets of $\mathcal{Z}$ and $\mathcal{B}$ are closed, one has

$$\mathrm{K}(T^m) = \bigcup_p \left(\mathcal{Z}_p^{(p)} \cap \mathcal{B}_{p+m}^{(p)} \right) \qquad\qquad \mathrm{I}(T^m) = \bigcup_p \mathcal{Z}_p^{(p+m)}.$$

The ranks of $\mathrm{K}(T^m)$ and $\mathrm{I}(T^m)$ in $\mathcal{N}$ are therefore given by the left and right-hand sums below.

$$\sum_p \rho\left((\mathcal{Z}_p \cap \mathcal{B}_{p+m})/\mathcal{B}_p \right) \qquad\qquad \sum_p \rho\left(\mathcal{Z}_p/\mathcal{B}_{p+m} \right).$$

The identity below follows from the inclusion of $\mathcal{B}_p$ into the intersection of $\mathcal{Z}_p$ and $\mathcal{B}_{p+m}$. The subsequent estimate is a consequence of submodularity, and holds with strict equality if and only if $(\mathcal{Z}_p, \mathcal{B}_{p+m})$ is modular.

$$\begin{aligned} \rho((\mathcal{Z}_p \cap \mathcal{B}_{p+m})/\mathcal{B}_p) &= \rho(\mathcal{Z}_p \cap \mathcal{B}_{p+m}) - \rho(\mathcal{B}_p) \\ \rho(\mathcal{Z}_p/\mathcal{B}_{p+m}) &\leq \rho(\mathcal{Z}_p) - \rho(\mathcal{Z}_p \cap \mathcal{B}_{p+m}). \end{aligned}$$

Since the rank of $\mathcal{N}$ is $\sum_p(\rho(\mathcal{Z}_p) - \rho(\mathcal{B}_p))$, complementarity holds if and only if strict equality holds in both estimates, for all p. $\square$

Since $(\mathcal{Z}_{p+m}, \mathcal{B}_p)$ is trivially modular for every nonnegative m, we have shown the following.

Proposition 5.3.4. *Operator T is Jordan if and only if $\mathcal{Z} \cup \mathcal{B}$ is modular.*

In light of the preceding observation, it is reasonable to suppose that a basis that generates $\mathcal{Z}$ and $\mathcal{B}$ may bear some special relation to the Jordan bases of T. For convenience, define the *orbit* of a subset $S \subseteq E$ to be the orbit of $\psi(S)$, where ψ is the map that sends each $e \in \mathcal{M}$ to $e^{(\mathcal{Z}(e))}$ in $\mathcal{N}$.

Proposition 5.3.5. *The Jordan bases of T are the orbits in $\mathcal{N}$ of the $\mathcal{Z}$-$\mathcal{B}$-minimal bases in $\mathcal{M}$.*

Proof. If B freely generates $\mathcal{Z}$ and $\mathcal{B}$ in $\mathcal{M}$, and if J is the orbit of B, then $J \cap \mathcal{H}_p = B^{(p)}_{\mathcal{Z} \leq p < \mathcal{B}}$ freely generates $\mathcal{H}_p = \mathcal{Z}_p \,/\!\!/\, \mathcal{B}_p$ for all p. Therefore J is a Jordan basis.

Conversely, suppose that J is a Jordan basis, and let

$$B = \bigcup_p \{e : e^{(p)} \in J\}.$$

Since $\mathcal{N}$ is the matroid union of the $\mathcal{H}_p$, J freely generates both $\mathrm{I}(T^m) \cap \mathcal{H}_p$ and $\mathrm{K}(T^m) \cap \mathcal{H}_p$ for all m and p. Thus B freely generates $\mathcal{Z}$ and $\mathcal{B}$ on the minor $\mathcal{Z}_p/\mathcal{B}_p$. Consequently, if $\mathcal{Z}'$ and $\mathcal{B}'$ are the restrictions of $\mathcal{Z}$ and $\mathcal{B}$, respectively, to $\mathcal{Z}_p - \mathcal{B}_p$, then B freely generates $(\mathcal{Z}'_{q+1} \cap \mathcal{B}'_{p+1})/\mathcal{B}_p$ and $\mathcal{Z}'_q/\mathcal{B}_p$ for all q. By modularity, it generates $(\mathcal{Z}'_{q+1} \cap \mathcal{B}'_{p+1})/(\mathcal{Z}'_q \cup \mathcal{B}_p)$. When $q < p$ one has $\mathcal{Z}'_{q+1} \cap \mathcal{B}'_{p+1} = \mathcal{Z}_{q+1} \cap \mathcal{B}_{p+1}$ and $\mathcal{Z}'_q = \mathcal{Z}_q$, so this minor agrees with

$$(\mathcal{Z}_{q+1} \cap \mathcal{B}_{p+1})/(\mathcal{Z}_q \cup \mathcal{B}_p). \tag{5.3.1}$$

When $p \leq q$ the ground set of (5.3.1) contains only elements e for which $\mathcal{B}(e) = p+1 \leq q+1 = \mathcal{Z}(e)$. Since $e^{(p)}$ is a loop for every such e, it follows that B intersects the ground set of (5.3.1) trivially when $p \leq q$.

In conclusion, B is the disjoint union of some independent sets in the matroid $\mathcal{M}/(\mathcal{Z} \cup \mathcal{B})$, and may thus be extended to a basis that generates $\mathcal{Z}$ and $\mathcal{B}$. The orbit of this set will be a Jordan basis containing J. □

In closing, let us return to the linear regime. Suppose that $\mathcal{Z}_p$ and $\mathcal{B}_p$ are linear filtrations on a vector space W, and let V_p denote the linear quotient space $\mathcal{Z}_p/\mathcal{B}_p$. The inclusion maps $\mathcal{Z}_p \to \mathcal{Z}_{p+1}$ induce linear maps $V_p \to V_{p+1}$, which collectively determine a graded linear operator on $\oplus_p V_p$. Let us denote this map by Q. There is a canonical set function $\phi : \cup_p \mathcal{H}_p \to \cup_p V_p$, which may be described as the rule that assigns $e \in \mathcal{H}_p$ to the equivalence class of e in V_p. For each p there is a commutative square

$$\begin{tikzcd}
\mathcal{H}_p \arrow[r, "T"] \arrow[d] & \mathcal{H}_{p+1} \arrow[d] \\
V_p \arrow[r, "Q"] & V_{p+1}
\end{tikzcd}$$

whose vertical maps are given by ϕ. Since $S \subseteq \mathcal{H}_p$ is independent in $\mathcal{N}$ iff $\varphi(S)$ is linearly independent in V_p, it follows that $J \subseteq \mathcal{N}$ is a matroid theoretic Jordan basis of T if and only if $\varphi(J)$ is a linear Jordan basis of Q. Thus the following.

Proposition 5.3.6. *The graded Jordan bases of Q are the orbits of the $\mathcal{Z}$-$\mathcal{B}$ minimal bases in W.*

5.4 Generalized Canonical Forms

Let $\mathbf{k}[x]$ denote the ring of polynomials over ground field $\mathbf{k}$. Recall that this object consists of a $\mathbf{k}$-linear vector space freely generated by a basis of formal symbols $\{1, x, x^2, \ldots\}$, and a binary operation $\mathbf{k}[x] \times \mathbf{k}[x] \to \mathbf{k}[x]$ sending (p, q) to $p \cdot q$, where $\cdot$ is the standard multiplication of polynomials. A polynomial is *irreducible* over $\mathbf{k}$ if it cannot be expressed as the product of two polynomials of positive degree. We will write (p) for the *ideal* generated by p, which may be formally expressed $\{q \cdot p : q \in \mathbf{k}[x]\}$.

Remark 5.4.1. As (p) is a linear subspace of $\mathbf{k}[x]$ we may form the quotient space $\mathbf{k}[x]/(p)$ by the usual coset construction. It is typical, in any such construction, to write $[p]$ for the equivalence class of a vector p in the quotient space. In order to avoid an excessive number of brackets, however, we will bypass this convention, writing p for $[p] \in \mathbf{k}[x]/(p)$.

Since left-multiplication with x carries every element of (p) to a polynomial in the same set, the linear map $p \mapsto x \cdot p$ determines an operator T on the quotient space $\mathbf{k}[x]/(q)$. If $q = a_0 + a_1 x + \cdots + a_d x^d$ then $\{1, x, \ldots, x^{d-1}\}$ is a basis for $\mathbf{k}[x]/(q)$, as is simple to check. The matrix representation of T with respect to this basis has form

$$
\mathrm{Com}(q) = \begin{pmatrix} 0 & 0 & \cdots & 0 & -a_0 \\ 1 & 0 & \cdots & 0 & -a_1 \\ 0 & 1 & \cdots & 0 & -a_2 \\ & & \vdots & & \\ 0 & 0 & \cdots & 0 & -a_{d-2} \\ 0 & 0 & \cdots & 1 & -a_{d-1} \end{pmatrix}
$$

If q is irreducible then this matrix has full rank, since a_0 does not vanish. If a is any positive integer, then

$$
\{q^{a-1}, xq^{a-1}, \ldots, x^{d-1}q^{a-1}, q^{a-2}, xq^{a-2}, \ldots, x^{d-1}q^{a-2}, \ldots, 1, x, \ldots x^{d-1}\}.
$$

is a basis for $\mathbf{k}[x]/(q^a)$, and in this basis multiplication with x has the following matrix form

$$
J(q,a) = \begin{pmatrix} \mathrm{Com}(q) & M & 0 & \cdots & 0 & 0 \\ 0 & \mathrm{Com}(q) & M & \cdots & 0 & 0 \\ & & & \vdots & & \\ 0 & 0 & 0 & \cdots & \mathrm{Com}(q) & M \\ 0 & 0 & 0 & \cdots & 0 & \mathrm{Com}(q) \end{pmatrix}
$$

where M is the square matrix with 1 in the top right entry and zeros elsewhere. We call this array a *generalized Jordan block*. Since $\mathrm{Com}(x + a_0) = (-a_0)$, the notion of a generalized Jordan block specializes to that of a classical Jordan block, when q is a linear polynomial. The following is a standard in linear algebra.

Theorem 5.4.2 (Generalized Jordan Canonical Form)**.** *If T is a linear operator on a finite dimensional vector space V over an arbitrary field $\mathbf{k}$, then there exists a basis of V with respect to which T has matrix form*

$$
\mathrm{diag}(J(p_1, a_1), \ldots, J(p_m, a_m)),
$$

where $p_1, \ldots, p_m$ are polynomials irreducible over $\mathbf{k}$, and $a_1, \ldots, a_m$ are positive integers. Such a presentation is unique up to a permutation of generalized Jordan blocks.

A block-diagonal matrix of the form $\mathrm{diag}(J(p_1, a_1), \ldots, J(p_m, a_m))$ is said to have *generalized Jordan Canonical Form.* The corresponding basis is a *generalized Jordan basis.* We will refer to the set of basis vectors that index a generalized Jordan block as a *Jordan orbit.* As an application of the classification of nilpotent Jordan bases for matroids, let us classify the generalized Jordan bases of $\mathbf{k}$-linear maps.

To begin, let T be any operator on a finite-dimensional $\mathbf{k}$-linear vector space V. By Theorem 5.4.2, there exist irreducible polynomials $p_1, \ldots, p_m$, positive integers $a_1, \ldots, a_m$ and a linear isomorphism Φ from V to

$$
U = \bigoplus_{i=1}^{m} \mathbf{k}[x]/(p_i^{a^i})
$$

such that $\Phi T = X\Phi$, where X is the linear map on U induced by left-multiplication with x. The generalized Jordan bases of T are exactly those of form $\Phi^{-1}(J)$, where J is a generalized Jordan basis of X, so we may deduce most of what we need to know about general operators from the special case $T = X$.

For convenience, let us denote by p the map $U \to U$ induced by left-multiplication with p. Under this convention

$$\mathbf{K}\left(p^{\dim V}\right) = \oplus_{i \in I} \mathbf{k}[x]/\left(p^{a_i}\right) \tag{5.4.1}$$

where $I = \{i : p_i = p\}$, for each p. Every Jordan orbit is contained in a subspace of form $\mathrm{K}(p^{\dim V})$, and every subspace of this form is invariant under X. Thus every generalized Jordan basis may be expressed as a disjoint union $\cup_p J_p$, where p runs over irreducible polynomials and J_p is a generalized Jordan basis for the restriction of X to $\mathrm{K}(p^{\dim V})$.

Let us fix an irreducible polynomial p with degree d, and assume temporarily that

$$U = \mathrm{K}\left(p^{\dim V}\right).$$

If $E_k = \mathrm{I}(p^k)$, then for each nonnegative integer k the quotient module

$$U_k = E_k/E_{k+1}$$

is a direct sum of simple modules isomorphic to $\mathbf{k}[x]/(p)$. We define a subset $S \subseteq E_k$ to be *independent* if $\pi_k S$ generates a submodule of length $|S|$ in U_k, where π_k is the quotient map $E_k \to U_k$. Let us denote the family of all such sets by $\mathcal{I}_k$.

Lemma 5.4.3. *The pair*

$$\mathcal{N}_k = (E_k, \mathcal{I}_k).$$

is a matroid independence system.

Proof. Both the Steinitz Exchange Axiom and closure under inclusion follow from the Krull-Schmidt theorem for semisimple modules of finite length. □

If $\mathcal{N}$ is the matroid union $\cup_k \mathcal{N}_k$, and $e^{(k)}$ is the copy of polynomial e that lies in the ground set of $\mathcal{N}_k$, then one may define an operator $P : \mathcal{N} \to \mathcal{N}$ by

$$P\left(e^{(k)}\right) = (p \cdot e)^{(k+1)}.$$

As P^m is complementary for all m, it has a Jordan basis. We claim that the generalized Jordan bases of X are exactly the sets of form

$$X^0 J \cup \cdots \cup X^{d-1} J \tag{5.4.2}$$

where J is a Jordan basis of P.

To see that this is the case, let J be such a basis, and for convenience put $\pi_k j = 0$ for all $j \notin E_k$. A set $I \subseteq E_k$ freely generates $\mathcal{N}_k$ as a matroid iff $\pi_k I$ freely generates U_k as a module, so the nonzero elements of $\cup_k \pi_k J$ freely generate $\oplus_k U_k$ as a module. For each j and k so that $\pi_k j \neq 0$, the submodule generated by $\pi_k j$ in U_k is the linear span of $\{\pi_k X^0 j, \ldots, \pi_k X^{d-1} j\}$, so the nonzero elements of

$$\bigcup_k \bigcup_{m=0}^{d-1} \pi_k X^m J$$

freely generate $\oplus_k U_k$ as a $\mathbf{k}$-linear vector space. It follows easily that (5.4.2) is a basis for U. Evidently, it is a generalized Jordan basis.

To describe the generalized Jordan bases of an arbitrary operator T, one need only synthesize across irreducible polynomials. For a given operator T on vector space V, let us define $U^{(p)} \subseteq V$ to be the module on ground set $p^{\dim(V)}(T)$ on which x acts by

$$x \cdot v = p(T) \cdot v.$$

Set $E_k^{(p)} = x^k U^{(p)} = p^k(T) U^{(p)}$, and put

$$U_k^{(p)} = E_k^{(p)} / E_{k+1}^{(p)}.$$

Let $\mathcal{I}_k^{(p)}$ be the family of independent sets on $E_k^{(p)}$ given by the module structure of $U_k^{(p)}$, as described above, and $\mathcal{N}_k^{(p)} = (E_k^{(p)}, \mathcal{I}_k^{(p)})$ be the corresponding matroid. Write $P^{(p)}$ for the operator on $\mathcal{N}^{(p)} = \cup_k \mathcal{N}_k^{(p)}$ that sends each $e \in \mathcal{N}_k^{(p)}$ to the element $x \cdot e \in \mathcal{N}_{k+1}^{(p)}$.

Finally, let $\mathcal{N} = \cup_p \mathcal{N}^{(p)}$ be the matroid union of $\mathcal{N}_k^{(p)}$ over all irreducible polynomials p, and P be the operator on $\mathcal{N}$ determined by the maps $P^{(p)}$. If the *generalized Jordan orbit* of a vector $i \in U^{(p)}$ is the union $T^0 I \cup \cdots \cup T^{\deg(p)-1} I$, where I is the usual orbit of i under x in $U^{(p)}$, then we have shown the following.

Proposition 5.4.4. *If T, $\mathcal{N}$, and P are as above, then the generalized Jordan bases of T are the generalized orbits of the $\mathcal{K}$-minimal bases of $\mathcal{N}/P\mathcal{N}$, where $\mathcal{K}$ is the unique integer-valued weight function so that $\mathcal{K}_{\leq m} = \mathrm{K}(P^m)$.*

Part II

Algorithms

Chapter 6

Algebraic Foundations

This chapter lifts several classical ideas from the domain of matrix algebra and linear matroid representations to that of *abelian* and *preadditive categories*. For readers unfamiliar with the language of category theory, the terms *object* and *morphism* may be replaced by **k**-*linear vector space* and **k**-*linear map*, respectively. The term *map* is occasionally used in place of morphism. A *monomorphism* is an injection and an *epimorphism* is a surjection. An *endomorphism* on W is a morphism $W \to W$, and an *automorphism* is an invertible endomorphism. With these substitutions in place, the phrases *in a preadditive category* and *in an abelian category* may be stricken altogether.

6.1 Biproducts

A *product structure* on an object W is a family λ of maps $f : W \to \mathrm{D}^{op}(f)$, with the property that $\times\lambda : W \to \times_{f\in\lambda}\mathrm{D}^{op}(f)$ is invertible. A *coproduct structure* W is a family υ of maps $f : \mathrm{D}(f) \to W$, such that $\oplus\upsilon : \oplus_{f\in\upsilon}\mathrm{D}(f) \to W$ is an isomorphism.

To every product structure λ corresponds a unique *dual* coproduct structure $\lambda^\flat$ and bijection $\flat : \lambda \to \lambda^\flat$ such that $f^\flat g = \delta_{fg}$. Likewise, to every coproduct structure υ corresponds a unique dual product structure $\upsilon^\sharp$ and bijection $\sharp : \upsilon \to \upsilon^\sharp$ such that $fg^\sharp = \delta_{fg}$. We refer to an unordered pair consisting of a product structure and its dual as a *biproduct structure*.

A *complement* to a family λ of morphisms out of W is a map g such that $\lambda \cup \{g\}$ is a product structure. A *complement* to a family υ of morphisms into W is a morphism g such that $\upsilon \cup \{g\}$ is a coproduct structure.

Remark 6.1.1. The following mnemonic may be helpful in recalling the distinction between $\flat$ and $\sharp$. The former denotes a gravitational drop in tone, with $\flat$ connoting maps directed into W. Conversely, sharp denotes a rise in tone, suggesting upward, outward motion, with $\sharp$ connoting maps leaving W.

Examples

Example 6.1.2. Let $\mathcal{U}$ be any finite family of **k**-linear spaces. For each $V \in \mathcal{U}$, let π_V denote the projection $\oplus_{\mathcal{U}}U \to V$ that vanishes on $\oplus_{U\neq V}U$, and let ι_V denote the inclusion

$V \to \oplus_{\mathcal{U}} U$. The pair of $\mathcal{U}$-indexed families $\{\pi, \iota\}$ is the *canonical biproduct structure* on $\oplus_{\mathcal{U}} U$ generated by $\mathcal{V}$.

Example 6.1.3. Suppose that the set $\mathcal{U}$ in the preceding example is a complementary family of subspaces in some $\mathbf{k}$-linear space W. The structure defined by replacing $\oplus_{\mathcal{U}} U$ with W in the description of π_V and ι_V is called the *canonical biproduct structure* on W generated by $\mathcal{V}$.

Example 6.1.4 (Primal). An arbitrary family of injective linear maps into W is a coproduct structure iff $\{\mathrm{I}(f) : f \in \lambda\}$ is a complementary family of subspaces in W.

Example 6.1.5 (Dual). If λ is a family of surjective linear maps out of W, then λ is a product structure iff the family of kernels $\mathcal{K} = \{\mathrm{K}(f) : f \in \lambda\}$ is *cocomplementary*, meaning that $\cap_{\mathcal{K}} L = 0$ and $\cap_{\mathcal{U}} L \neq 0$ when $\mathcal{U} \subsetneq \mathcal{K}$.

Example 6.1.6. Let λ be the unique index function $\lambda : I \to \mathrm{Hom}(\mathbf{k}, \mathbf{k}^I)$ such that

$$\lambda_i(1) = \chi_i$$

for each $i \in I$. Then λ is a coproduct structure on $\mathbf{k}^I$. The dual product structure is the function $\lambda^\sharp : I \to \mathrm{Hom}(\mathbf{k}^I, \mathbf{k})$ such that

$$\lambda_i^\sharp(w) = w_i$$

for $w \in \mathbf{R}^n$. We call λ the *canonical indexed coproduct structure* on $\mathbf{k}^I$. The set $\{\lambda_i : i \in I\}$ is the canonical *unindexed* coproduct structure. By a slight abuse of language, we will use the term *canonical coproduct structure* to refer to either of these constructions.

6.2 Idempotents

An endomorphism $e : W \to W$ is *idempotent* if $e^2 = e$. Equivalently, e is an idempotent iff

$$e|_{\mathrm{I}(e)} = 1_{\mathrm{I}(e)} \qquad\qquad e|_{\mathrm{K}(e)} = 0.$$

Given complementary subobjects K and I, we write e_I^K for the idempotent with kernel K and image I. We refer to this morphism as *projection onto I along K*.

If ω is a (co)product structure, then to each $f \in \omega$ we associate the endomorphism

$$e_f = f^\flat f^\sharp.$$

Evidently, $e_f = e_{\mathrm{I}(f^\flat)}^{\mathrm{K}(f^\sharp)}$. To emphasize dependence on ω, we will sometimes write e_f^ω for e_f. If $\omega = \{f, g\}$ is a coproduct structure, then $e_f^\omega = e_{\mathrm{I}(f)}^{\mathrm{I}(g)}$. Similarly, if ω is a product structure then $e_f^\omega = e_{\mathrm{K}(f)}^{\mathrm{K}(g)}$.

Proposition 6.2.1. *If ω is a coproduct structure, then*

$$\sum_{f \in \omega} e_f = 1. \tag{6.2.1}$$

Proof. For convenience, assume ω to be a product structure. Since $\times\omega$ is an isomorphism,

$$(\times\omega)a = (\times\omega)b$$

iff $a = b$. Equivalently, $a = b$ iff $ga = gb$ for all $g \in \omega$. Since $g\left(\sum e_f\right) = g = g1$ for all $g \in \omega$, the desired conclusion follows. $\square$

The chief application of idempotents, for us, will be to generate new complements from old. The instrument that executes such operations is Lemma 6.2.2, which says that projection along a subspace L yields a linear isomorphism between any two of its complements. See Examples 1 and 2 for illustration.

Lemma 6.2.2. *Let $\omega^{op} = \{f^{op}, g^{op}\}$ and $\omega = \{f, g\}$ be product and coproduct structures on W^{op} and W, respectively.*

1. *If $h : \mathrm{D}(h) \to W$ and either of the two families*

$$\{f, h\} \qquad \textit{and} \qquad \{f, e^{\omega}_{g} h\}$$

is a coproduct structures, then so is the other.

2. *If $h^{op} : W^{op} \to \mathrm{D}^{op}(h^{op})$ and either of the two families*

$$\{f^{op}, h^{op}\} \qquad \textit{and} \qquad \{f^{op}, e^{\omega^{op}}_{g^{op}} h^{op}\}$$

is a product structure, the so is the other.

Example 1: Complements in $\mathbf{R}^2$

Let $\lambda = \{\lambda_1, \lambda_2\} \subseteq \mathrm{Hom}(\mathbf{R}, \mathbf{R}^2)$ be the canonical unindexed coproduct structure on $\mathbf{R}^2$. Recall from Example 6.1.6 that this means λ_1, λ_2 are the morphisms $\mathbf{R} \to \mathbf{R}^2$ such that

$$\lambda_1(1) = (1, 0) \qquad\qquad \lambda_2(1) = (0, 1).$$

The idempotents e_{λ_1} and e_{λ_2} are then the orthogonal projection maps onto the first and second coordinate axes, respectively.

If $\upsilon = \{\upsilon_1, \upsilon_2\} \subseteq \mathrm{Hom}(\mathbf{R}, \mathbf{R}^2)$ is defined by

$$\upsilon_1(1) = (1, 0) \qquad\qquad \upsilon_2(1) = (1, 1).$$

then

$$e_{\lambda_2}\upsilon_2 = \lambda_2$$

complements λ_1. Thus Lemma 6.2.2 confirms what was clear from the outset: that the family $\upsilon = \{\lambda_1, \upsilon_2\}$ is a coproduct structure.

Example 2: Complements in $\mathbf{R}^n$

If $\lambda = \{\lambda_1, \lambda_2\}$ is any coproduct structure on $\mathbf{R}^n$ then the subspaces

$$L_1 = \mathrm{I}(\lambda_1) \qquad \textit{and} \qquad L_2 = \mathrm{I}(\lambda_2)$$

are complementary. Suppose that U_2 is a subspace of $\mathbf{R}^n$ complementary to L_1, and let υ_2 denote the inclusion $U_2 \to \mathbf{R}^n$. Since $\mathrm{K}(e_{\lambda_2}) = \mathrm{I}(\lambda_1)$ intersects U_2 trivially, one may deduce

$$L_2 = e_{\lambda_2} U_2$$

by dimension counting. Thus $\{\lambda_1, e_{\lambda_2}\upsilon_2\}$ is a coproduct structure. Alternatively, one could have arrived at this conclusion by noting that $\{\lambda_1, \upsilon_2\}$ is a coproduct structure, and applying Lemma 6.2.2.

6.3 Arrays

Let T be any map $V \to W$. We will frequently compose elements of (co)product structures with maps of this form, and in such cases musical symbols may be dropped without ambiguity. For instance, if υ is a product structure on V and $f \in \upsilon$, then one may write

$$Tf$$

without ambiguity, since $Tf^\flat$ is well defined and $Tf^\sharp$ is not. This convention has the beneficial consequence of reducing unwieldy notation.

If α is a family of maps out of W and β is a family of maps into V, the *matrix representation* of T with respect to (α, β) is an $\alpha \times \beta$ array, also denoted T, defined by

$$T(a, b) = aTb.$$

In keeping with convention, when α and β are (co)products, we write $T(\alpha, \beta)$ for $T(\alpha^\sharp, \beta^\flat)$.

There is a natural multiplication on matrix representations, which agrees with composition. Recall that to every (co)product υ on the domain of T corresponds a υ-indexed family of idempotents $e_f = f^\flat f^\sharp$ such that $\sum_\upsilon e_f = 1$. Thus for any composable T and U one has

$$(TU)(a, b) = \sum_f aTe_fUb = \sum_f T(a, f)U(f, b).$$

When λ and υ are composed of linear maps $\mathbf{k} \to \mathrm{D}^{op}(T)$ and $\mathbf{k} \to \mathrm{D}(T)$, respectively, $T(\lambda, \upsilon)$ is the standard matrix representation of T with respect to bases $\{f(1) : f \in \lambda\}$ and $\{g(1) : g \in \upsilon\}$.

Example 1: Endomorphisms on $\mathbf{R}^2$

Let $\lambda = \{\lambda_1, \lambda_2\} \subseteq \mathrm{Hom}(\mathbf{R}, \mathbf{R}^2)$ be the canonical unindexed coproduct structure on $\mathbf{R}^2$. Recall from Example 6.1.6 that this means λ_1, λ_2 are the morphisms $\mathbf{R} \to \mathbf{R}^2$ such that

$$\lambda_1(1) = (1, 0) \qquad \lambda_2(1) = (0, 1).$$

Let $\upsilon = \{\upsilon_1, \upsilon_2\} \subseteq \mathrm{Hom}(\mathbf{R}, \mathbf{R}^2)$ be any other coproduct structure, and define (u_{ij}) by

$$\upsilon_1(1) = (u_{11}, u_{21}) \qquad\qquad \upsilon_2(1) = (u_{12}, u_{22}).$$

If we identify $\mathrm{Hom}(\mathbf{R}, \mathbf{R})$ with $\mathbf{R}$ via $f \leftrightarrow f(1)$, then

$$1(\lambda, \upsilon) \quad = \quad \begin{array}{c|cc|} & \upsilon_1 & \upsilon_2 \\ \cline{2-3} \lambda_1^\sharp & u_{11} & u_{12} \\ \lambda_2^\sharp & u_{21} & u_{22} \\ \cline{2-3} \end{array}$$

Likewise, if $(u^{ij}) = (u_{ij})^{-1} \in GL_2(\mathbf{R})$, then

$$1(\lambda, \upsilon) \quad = \quad \begin{array}{c|cc|} & \lambda_1 & \lambda_2 \\ \cline{2-3} \upsilon_1^\sharp & u^{11} & u^{12} \\ \upsilon_2^\sharp & u^{21} & u^{22} \\ \cline{2-3} \end{array}$$

Thus if $u^i = (u^{i1}, u^{i2})$, then

$$\upsilon_i^\sharp(w) = \langle u^i, w \rangle,$$

where $\langle \cdot\,, \cdot \rangle$ is the standard inner product on $\mathbf{R}^2$.

Example 2: Endomorphisms on $\mathbf{R}^n$

The observations in Example 1 extend naturally to $\mathbf{R}^n$. Let λ be the canonical unindexed coproduct structure on $\mathbf{R}^n$. Let $\upsilon = \{\upsilon_1, \ldots, \upsilon_n\} \subseteq \mathrm{Hom}(\mathbf{R}, \mathbf{R}^n)$ be a second coproduct structure, and put

$$u_q = \upsilon_q(1) = (u_{1q}, \ldots, u_{nq}).$$

If one defines $(u^{pq}) = (u_{pq})^{-1}$, then

$$1(\lambda, \upsilon) = (u_{pq}) \qquad \textit{and} \qquad 1(\upsilon, \lambda) = (u^{pq}).$$

The righthand identity states $\upsilon_p^\sharp \lambda_q = u^{pq}$ for all $p, q \in \{1, \ldots, n\}$. Since

$$w = (w_1\lambda_1 + \cdots + w_n\lambda_n)(1)$$

for all $w \in \mathbf{R}^m$, it follows that

$$\upsilon_p^\sharp(w) = \langle u^p, w \rangle$$

for all $p \in \{1, \ldots, n\}$, where $u^p = (u^{p1}, \ldots, u^{pn})$.

6.4 Kernels

A morphism k is *kernel* to a morphism f if the following are equivalent for every g such that fg is well defined:

$$\begin{pmatrix} - & u^1 & - \\ - & u^2 & - \\ & \vdots & \\ - & u^n & - \end{pmatrix} \qquad\qquad \begin{pmatrix} | & | & & | \\ u_1 & u_2 & \cdots & u_n \\ | & | & & | \end{pmatrix}$$

Figure 6.1: Arrays associated to the coproduct structures λ and υ. *Left:* $1(\upsilon, \lambda)$. The pth row of this array is the unique tuple u^p such that $\langle u^p, w \rangle = \upsilon_p^\sharp(w)$ for all $w \in \mathbf{R}^n$. *Right:* $1(\lambda, \upsilon)$. The pth column of this array is the tuple $u_p = \upsilon_p(1)$. The matrix products $1(\upsilon, \lambda)1(\lambda, \upsilon) = 1(\upsilon, \upsilon)$ and $1(\lambda, \upsilon)1(\upsilon, \lambda) = 1(\lambda, \lambda)$ are Dirac delta functions on $\upsilon \times \upsilon$ and $\lambda \times \lambda$, respectively.

1. $fg = 0$.
2. There exists exactly one morphism h so that $kh = g$. Equivalently, there exists exactly one h such that the following diagram commutes.

$$\begin{tikzcd} \bullet \arrow[dr, "k"'] & & \bullet \arrow[ll, dotted, "h"'] \arrow[dl, "g"] \\ & \bullet & \end{tikzcd}$$

Lemma 6.4.1. *A morphism k is kernel to f iff k is a monomorphism and*

$$\mathrm{I}(k) = \mathrm{K}(f).$$

The notion of a kernel has a natural dual: we say that $k^{op} : \mathrm{D}^{op}(f^{op}) \to \mathrm{D}^{op}(k^{op})$ is *cokernel* to f^{op} if the following are equivalent for every g^{op} such that $g^{op} f^{op}$ is well defined.

1. $g^{op} f^{op} = 0$.
2. There exists exactly one morphism h^{op} so that $g^{op} = k^{op} h^{op}$. Equivalently, there exists exactly one h^{op} such that the following diagram commutes.

$$\begin{tikzcd} \bullet \arrow[rr, dotted, "h^{op}"] & & \bullet \\ & \bullet \arrow[ul, "k^{op}"] \arrow[ur, "g^{op}"'] & \end{tikzcd}$$

Lemma 6.4.2. *A morphism k^{op} is cokernel to f^{op} iff k^{op} is an epimorphism and*

$$\mathrm{I}(f^{op}) = \mathrm{K}(k^{op}).$$

Biproducts provide natural examples of kernel and cokernel maps.

Lemma 6.4.3. *If $\{f, g\}$ is a coproduct structure, then*

1. *g is kernel to $f^\sharp$.*

2. *$g^\sharp$ is cokernel to f.*

Proof. The first assertion follows from Lemma 6.4.1, since g is a monomorphism and $\mathrm{I}(g) = \mathrm{K}(f^\sharp)$. The second assertion follows from Lemma 6.4.2, similarly. □

6.4.1 The Splitting Lemma

The relationship between kernel morphisms and complements is essential to the current discussion. This relationship, as formalized by Lemma 6.4.4 (the splitting lemma), represents a bridge between the matroid-theoretic notion of *complementarity* (read: matroid duality) and the category-theoretic notion of a biproduct (read: linear duality).

While the formulation of Lemma 6.4.4 is specially suited to the task at hand, several others are standard in various branches of mathematics. One familiar to students of linear algebra is the fact that

$$V \cong \mathrm{I}(T) \oplus \mathrm{K}(T)$$

for every linear $T : V \to W$. A second is the splitting lemma of homological algebra, which states that for any sequence of maps and objects

$$0 \longrightarrow A \xrightarrow{k} B \xrightarrow{f} C \longrightarrow 0$$

where k is kernel to f, any right-inverse to f is a complement to k. Each of these related results is implied by Lemma 6.4.4. The proof is omitted.

Lemma 6.4.4 (Splitting Lemma)**.** *Suppose that*

$$f^{op} f = 1,$$

where f^{op} and f are composable morphisms in a preadditive category.

1. *If k is kernel to f^{op}, then $\{f, k\}$ is a coproduct structure.*

2. *If k^{op} is cokernel to f, then $\{f^{op}, k^{op}\}$ is a product structure.*

Remark 6.4.5. Since $\mathrm{K}(h) = \mathrm{K}(\phi h)$ for any isomorphism ϕ, the condition $f^{op} f = 1$ may be replaced in the preceding lemma by the condition that $f^{op} f$ be invertible.

6.4.2 The Exchange Lemma

If the purpose of the splitting lemma is to understand duality, then the purpose of duality, for us, is to understand *exchange*. An *exchange* is an operation of form $\omega \mapsto \omega - \alpha \cup \beta$, where ω and $\omega - \alpha \cup \beta$ are product (respectively, coproduct) structures, and

$$\alpha \subseteq \alpha \cap \omega, \qquad\qquad \beta \cap \omega \subseteq \alpha. \tag{6.4.1}$$

In practice the two conditions encoded in (6.4.1) may be safely ignored; we impose them only to avoid degenerate cases, e.g. where the sets $(\omega - \alpha)$ and β intersect nontrivially, hence where β "adds" elements to ω that are already present.

A pair (α, β) that meets all the given criteria as an *exchange pair* for ω. We refer to an exchange operation of form

$$\omega \mapsto \omega - \{a\} \cup \{b\}$$

as an *elementary exchange*, and to the pair (a, b) as an *elementary* exchange pair, or simply as an exchange pair, where context leaves no room for confusion.

Exchange operations are fundamental both to modern methods in matrix algebra and to matroid theory. One of the simplest questions surrounding this subject is the following: Given ω, when is (α, β) an exchange pair? This question reduces determining when (a, b) is an *elementary* exchange pair, since, for example, if ω is a coproduct then $\omega - \alpha \cup \beta$ is a coproduct iff

$$(\omega - \alpha) \cup \{\oplus\beta\}$$

is a coproduct.

The question of determining exchange pairs is therefore answered exhaustively by Lemma 6.4.6, the *Exchange Lemma.* This result is mathematically equivalent to the splitting lemma, and retains much of its flavor. Like the splitting lemma, it has several variations, the Steinitz Exchange Lemma prominent among them. This correspondence outlines the foundational overlap between homological algebra matroid theory. As with the splitting lemma, the proof is left as an exercise to the reader.

Lemma 6.4.6 (Exchange Lemma)**.** *Let $\{f, g\}$ and $\{f^{op}, g^{op}\}$ be product and coproduct structures, respectively.*

1. *An unordered pair $\{f, h\}$ is a product iff $g^\sharp h$ is invertible.*

2. *An unordered pair $\{f^{op}, h^{op}\}$ is a coproduct iff $h^{op} (g^{op})^\flat$ is invertible.*

To gain some familiarity with kernel maps, as well as the splitting and exchange lemmas, let us compute some examples.

Example 1: Kernels in $\mathbf{R}^2$

Let λ be the canoncial indexed coproduct structure on $\mathbf{R}^2$, and let

$$T : \mathbf{R}^2 \to \mathbf{R}^2$$

be any endomorphism. If $a_{ij} = \lambda_i^\sharp T \lambda_j$, then

$$T(\lambda, \lambda) \quad = \quad \begin{array}{c|cc|} & \lambda_1 & \lambda_2 \\ \hline \lambda_1^\sharp & a_{11} & a_{12} \\ \lambda_2^\sharp & a_{21} & a_{22} \\ \hline \end{array} \;.$$

Suppose that $a_{11} \neq 0$, and let k be the map $\mathbf{R} \to \mathbf{R}^2$ such that

$$1(\lambda, k) \quad = \quad \begin{array}{c|c|} & k \\ \cline{2-2} \lambda_1^\sharp & -a_{11}^{-1}a_{12} \\ \lambda_2^\sharp & 1 \\ \cline{2-2} \end{array} \; .$$

Since

$$T(\lambda, \lambda)1(\lambda, k) \quad = \quad \begin{array}{c|c|} & k \\ \cline{2-2} \lambda_1^\sharp & 0 \\ \lambda_2^\sharp & a_{22} - a_{21}a_{11}^{-1}a_{12} \\ \cline{2-2} \end{array}$$

one has, in particular, that

$$\lambda_1^\sharp T k = 0.$$

As the kernel of $\lambda_1^\sharp T$ has dimension 1, it follows that

$$\mathrm{I}(k) = \mathrm{K}(\lambda_1^\sharp T).$$

Since, in addition, k is a monomorphism, we have that k is kernel to $\lambda_1^\sharp T$.

Example 2: Preadditive kernels

The construction described in Example 1 is in fact very general. Let W and W^{op} be any two objects with coproduct structures $\lambda = \{\lambda_1, \lambda_2\}$ and $\lambda^{op} = (\lambda_1^{op}, \lambda_2^{op})$, respectively, and fix

$$T : W \to W^{op}.$$

For convenience, let $a_{ij} = \lambda_i^{op} T \lambda_j$.

Proposition 6.4.7. *If a_{11} is invertible, then in any preadditive category*

$$k = \lambda_2 - \lambda_1 a_{11}^{-1} a_{12}$$

is kernel to $\lambda_1^{op} T$.

Proof. The matrix representation of T with respect to λ^{op} and λ is

$$T(\lambda^{op}, \lambda) \quad = \quad \begin{array}{c|cc|} & \lambda_1 & \lambda_2 \\ \cline{2-3} \lambda_1^{op} & a_{11} & a_{12} \\ \lambda_2^{op} & a_{21} & a_{22} \\ \cline{2-3} \end{array}$$

by definition, and it is simple to check that

$$1(\lambda, k) \quad = \quad \begin{array}{c|c|} & k \\ \cline{2-2} \lambda_1^\sharp & -a_{11}^{-1}a_{12} \\ \lambda_2^\sharp & 1 \\ \cline{2-2} \end{array} \; .$$

As in Example 1, we have

$$T(\lambda^{op}, \lambda)1(\lambda, k) \quad = \quad \begin{array}{cc} & k \\ \begin{array}{c} \lambda_1^\sharp \\ \lambda_2^\sharp \end{array} & \boxed{\begin{array}{c} 0 \\ a_{22} - a_{21}a_{11}^{-1}a_{12} \end{array}} \end{array} .$$

In particular, $\lambda^{op}Tk = 0$. In the category finite-dimensional $\mathbf{k}$-linear spaces and maps between them, one can confirm that k is a kernel to $\lambda^{op}T$ by checking

$$\mathrm{I}(k) = \mathrm{K}(\lambda_1^{op}T),$$

via dimension counting. To establish the result for arbitrary preadditive categories, however, we will check the definition directly.

Recall that $e_f = f^\flat f^\sharp$ is the idempotent projection operator generated by a morphism f in a (co)product structure ω, and that $\sum_{f\in\omega} e_f = 1$. If h is any morphism such that $0 = \lambda_1^{op}Th$, then

$$0 = \lambda_1^{op}T(e_{\lambda_1} + e_{\lambda_2})h = a_{11}\lambda_1^\sharp h + a_{12}\lambda_2^\sharp h,$$

whence

$$\lambda_1^\sharp h = -a_{11}^{-1}a_{12}\lambda_2^\sharp h,$$

and therefore

$$h = (\lambda_2 - \lambda_1 a_{11}^{-1}a_{12})\lambda_2^\sharp h.$$

To wit, the diagram

$$\begin{tikzcd} \bullet \arrow[dr, "k"'] & & \bullet \arrow[ll, dotted, "\lambda_2^\sharp h"'] \arrow[dl, "h"] \\ & \bullet & \end{tikzcd}$$

commutes. Therefore h factors through k when $\lambda_1^{op}Th = 0$. Since $\lambda_2^\sharp k = 1$, we have that k is a monomorphism. Thus the factorization is unique. The desired conclusion follows. $\square$

Example 3: Cokernels in $\mathbf{R}^2$

As in Example 1, let λ be the canonical unindexed coproduct structure on $\mathbf{R}^2$, and fix

$$T : \mathbf{R}^2 \to \mathbf{R}^2$$

If $a_{ij} = \lambda_i^\sharp T \lambda_j$, then

$$T(\lambda, \lambda) \quad = \quad \begin{array}{c|cc|} & \lambda_1 & \lambda_2 \\ \cline{2-3} \lambda_1^\sharp & a_{11} & a_{12} \\ \lambda_2^\sharp & a_{21} & a_{22} \\ \cline{2-3} \end{array} \; .$$

Suppose that $a_{11} \neq 0$, and let k^{op} be the map $\mathbf{R} \to \mathbf{R}^2$ such that

$$1(k^{op}, \lambda) \quad = \quad \begin{array}{c|cc|} & \lambda_1 & \lambda_2 \\ \cline{2-3} k^{op} & -a_{21}a_{11}^{-1} & 1 \\ \cline{2-3} \end{array} \; .$$

Since

$$1(k^{op}, \lambda)T(\lambda, \lambda) \quad = \quad \begin{array}{c|cc|} & \lambda_1 & \lambda_2 \\ \cline{2-3} k^{op} & 0 & a_{22} - a_{21}a_{11}^{-1}a_{12} \\ \cline{2-3} \end{array} \; .$$

one has, in particular, that $k^{op} T \lambda_1 = 0$. As the image of $T\lambda_1$ has dimension 1, it follows that

$$\mathrm{I}(T\lambda_1) = \mathrm{K}(k^{op}).$$

Since, in addition, k is an epimorphism, we have that k^{op} is a cokernel to $T\lambda_1$.

Example 4: Preadditive cokernels

Like the kernel in Example 1, the cokernel constructed in Example 3 has a natural analog for arbitrary two-element corpoducts. Let W and W^{op} be any two objects with coproduct structures $\lambda = \{\lambda_1, \lambda_2\}$ and $\lambda^{op} = (\lambda_1^{op}, \lambda_2^{op})$, respectively, and fix

$$T : W \to W^{op}.$$

As before, set $a_{ij} = \lambda_i^{op} T \lambda_j$.

The proof of Proposition 6.4.8 entirely dual to that of Proposition 6.4.7. The details are left as an exercise to the reader.

Proposition 6.4.8. *If a_{11} is invertible, then in any preadditive category*

$$k^{op} = \lambda_2^{op} - a_{11}^{-1} a_{21} \lambda_1^{op}$$

is cokernel to $T\lambda_1$.

6.4.3 Idempotents

The maps k and k^{op} described in the preceding sections bear a simple description in terms of idempotents. For a first example, take morphism k. By design, the image of k is the kernel of $\lambda_1^{op} T$. There is a canonical projection operator that acts as identity on this subspace, and that vanishes on the image of λ_1. We will show (Proposition 6.4.9) that k is merely the composition of λ_2 with this projection.

Dually, k^{op} is a morphism that vanishes on the image of $T\lambda_1$. There is canonical projection that vanishes on this subspace, and that acts as identity on the kernel of λ_1^{op}. We claim that k^{op} is merely the composition of λ_2^{op} with this operator. Proposition 6.4.9 formulates these assertions symbolically.

Proposition 6.4.9. *Let k and k^{op} be the kernel and cokernel maps constructed in Examples 2 and 4. Then*

$$k = e^{\mathrm{I}(\lambda_1)}_{\mathrm{K}(\lambda_1^{op}T)}\lambda_2 \qquad \textit{and} \qquad k^{op} = \lambda_2^{op} e^{\mathrm{I}(T\lambda_1)}_{\mathrm{K}(\lambda_1^{op})}.$$

Proof. We will argue the first identity; the proof of the second is essentially dual. From Example 2 we have that

$$1(\lambda, \{\lambda_1, k\}) \quad = \quad \begin{array}{c|cc|} & \lambda_1 & k \\ \hline \lambda_1^\sharp & 1 & -a_{11}^{-1}a_{12} \\ \lambda_2^\sharp & 0 & 1 \\ \hline \end{array} \; .$$

Since $1(\lambda, \{\lambda_1, k\})1(\{\lambda_1, k\}, \lambda) = 1(\lambda, \lambda)$, it follows that

$$1(\{\lambda_1, k\}, \lambda) \quad = \quad \begin{array}{c|cc|} & \lambda_1 & \lambda_2 \\ \hline \lambda_1^{\sharp\sharp} & 1 & a_{11}^{-1}a_{12} \\ k^{\sharp\sharp} & 0 & 1 \\ \hline \end{array}$$

where $\sharp\sharp$ is the sharp operator on $\{\lambda_1, k\}$, as distinguished from that of λ. Thus

$$(kk^{\sharp\sharp})(\lambda, \lambda) \quad = \quad \begin{array}{c|cc|} & \lambda_1 & \lambda_2 \\ \hline \lambda_1^\sharp & 0 & a_{11}^{-1}a_{12} \\ \lambda_2^\sharp & 0 & 1 \\ \hline \end{array} \; .$$

In particular we have

$$(kk^{\sharp\sharp})\lambda_2 = \lambda_2 - \lambda_1 a_{11}^{-1}a_{12} = k.$$

Since $kk^{\sharp\sharp} = e_k^{\{\lambda_1,k\}} = e^{\mathrm{I}(\lambda_1)}_{\mathrm{K}(\lambda_1^{op}T)}$, the desired conclusion follows. $\square$

6.4.4 Exchange

Let us consider the dual structures $\lambda^\sharp$ and $(\lambda^{op})^\flat$. Since by hypothesis

$$(\lambda_1^{op}T)\lambda_1 = a_{11} = \lambda_1^{op}(T\lambda_1)$$

is an isomorphism, the exchange lemma provides that

$$\upsilon = (\lambda_1^{op} T, \lambda_2^{\sharp}) \qquad \textit{and} \qquad \upsilon^{op} = (T\lambda_1, \lambda_2^{op\,\flat})$$

are ordered product and coproduct structures, respectively. Since k was designed as a kernel to $\lambda_1^{op} T$ and k^{op} was designed as a kernel to $T\lambda_1$, we naturally have

$$\upsilon_1 k = 0 \qquad \textit{and} \qquad k^{op}\upsilon_1^{op} = 0.$$

Since moreover

$$\upsilon_2 k = \lambda_2^{\sharp}\left(\lambda_2 \; - \lambda_1 a_{11}^{-1} a_{12}\right) = 1 \qquad k^{op}\upsilon_2^{op} = (\lambda_2^{op} - a_{11}^{-1} a_{21} \lambda_1^{op})\lambda_2^{op\,\flat} = 1$$

if follows that

$$k = \upsilon_2^{\flat} \qquad \textit{and} \qquad k^{op} = \upsilon_2^{op\,\sharp}.$$

To wit, k and k^{op} are *dual morphisms* in the biproduct structures generated by

$$\lambda^{\sharp} - \{\lambda_1^{\sharp}\} \cup \{\lambda_1^{op} T\} \qquad \textit{and} \qquad \lambda^{op\,\flat} - \{\lambda_1^{op\,\flat}\} \cup \{T\lambda_1\},$$

respectively.

6.5 The Schur Complement

Let A be an array with block structure

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}.$$

We allow entries a_{ij} to be either block submatrices with coefficients in a ground field $\mathbf{k}$, or maps of form $\lambda_i T \upsilon_j$, where λ and υ are (co)product structures on the (co)domain of a morphism T.

If a_{11} is invertible, then the *column clearing operation* on a_{11} is the operation $A \mapsto AU$, where

$$U = \begin{pmatrix} 1 & -a_{11}^{-1} a_{12} \\ 0 & 1 \end{pmatrix}.$$

Similarly, a *column pivot* on a_{11} is an operation of form $A \mapsto A\tilde{U}$, where

$$\tilde{U} = \begin{pmatrix} a_{11}^{-1} & -a_{11}^{-1} a_{12} \\ 0 & 1 \end{pmatrix}.$$

Dually, the *row clearing operation* on a_{11} is the operation $A \mapsto LA$, where

$$L = \begin{pmatrix} 1 & 0 \\ -a_{21} a_{11}^{-1} & 1 \end{pmatrix}$$

and the *row pivot* is an operation of form $A \mapsto \tilde{L}A$, where

$$\tilde{L} = \begin{pmatrix} a_{11}^{-1} & 0 \\ -a_{21}a_{11}^{-1} & 1 \end{pmatrix}.$$

We group row pivots and row clearing operations under the common heading of a *row operation*. Likewise, we refer to both column clearing operations and column pivots as *column operations*.

The name "clearing operation" finds motivation in the following matrix identities, where $*$ serves as a placeholder for arbitrary block submatrices.

$$AU = \begin{pmatrix} a_{11} & 0 \\ a_{21} & a_{22} - a_{21}a_{11}^{-1}a_{12} \end{pmatrix} \quad A\tilde{U} = \begin{pmatrix} 1 & 0 \\ * & a_{22} - a_{21}a_{11}^{-1}a_{12} \end{pmatrix}$$

$$LA = \begin{pmatrix} a_{11} & a_{12} \\ 0 & a_{22} - a_{21}a_{11}^{-1}a_{12} \end{pmatrix} \quad \tilde{L}A = \begin{pmatrix} 1 & * \\ 0 & a_{22} - a_{21}a_{11}^{-1}a_{12} \end{pmatrix}$$

It is clear that each row (respectively, column) operation on a_{11} produces an array with an invertible block in position $(1, 1)$. Thus any finite sequence of row and column operations may be performed on the upper-lefthand block of A.

If this sequence includes at least one row, one column, and one pivot operation, then the resulting array will have form

$$\begin{pmatrix} 1 & 0 \\ 0 & a_{22} - a_{21}a_{11}^{-1}a_{12} \end{pmatrix}.$$

If the sequence includes at least one row and one column operation but no pivots, then the resulting array will have form

$$\begin{pmatrix} a_{11} & 0 \\ 0 & a_{22} - a_{21}a_{11}^{-1}a_{12} \end{pmatrix}.$$

The submatrix that appears in all six of the preceding arrays, denoted

$$\sigma = a_{22} - a_{21}a_{11}^{-1}a_{12}$$

has a special name in matrix algebra: the *Schur complement*.

The Schur complement is an object of fundamental importance in mathematics. It is basic to modern algebra, analysis, geometry, probability, graph theory, combinatorics, and optimization, and plays a commensurate role in most branches of of the sciences and engineering. Specific subfields touched by the Schur complement include functional analysis (via Fredholm operators), conditional independence, matroid theory, geometric bundle theory, convex duality, and semidefinite programing. For historical details see the excellent introductory text *The Schur Complement and its Applications*, by Zhang [84]. Of particular interest to our story, the Schur complement encapsulates LU factorization, hence to a tremendous body of work in computational linear algebra.

We have seen that the Schur complement appears as the result of row and column operations. Let us give several new characterizations. As in Examples 2 and 4 of §6.4, let

$$\lambda = \{\lambda_1, \lambda_2\} \qquad \textit{and} \qquad \lambda^{op} = (\lambda_1^{op}, \lambda_2^{op}),$$

be coproduct structures on objects W and W^{op}, respectively, and fix

$$T : W \to W^{op}.$$

Let k be the kernel to $\lambda_1^{op}T$ obtained by composing λ_2 with projection onto the nullspace along the image of λ_1. In symbols,

$$k = e^{\mathrm{I}(\lambda_1)}_{\mathrm{K}(\lambda_1^{op}T)}\lambda_2.$$

Dually, let k^{op} be the cokernel to $T\lambda_1$ obtained by precomposing λ_2^{op} with projection onto the nullspace of λ_1^{op} along the image. In symbols,

$$k^{op} = \lambda_2^{op} e^{\mathrm{I}(T\lambda_1)}_{\mathrm{K}(\lambda_1^{op})}$$

We have seen (Propositions 6.4.7, 6.4.8, and 6.4.9) that

$$k = \lambda_2 - \lambda_1 a_{12} a_{11}^{-1} \qquad \textit{and} \qquad k^{op} = \lambda_2^{op} - \lambda_1^{op} a_{11}^{-1} a_{21}. \tag{6.5.1}$$

We have also noted that

$$1(\{\lambda_1^{op}, k^{op}\}, \lambda) \quad = \quad \begin{array}{c|cc|} & \lambda_1 & \lambda_2 \\ \hline \lambda_1^{op} & 1 & 0 \\ k^{op} & -a_{21}a_{11}^{-1} & 1 \\ \hline \end{array}$$

and

$$1(\lambda, \{\lambda_1, k\}) \quad = \quad \begin{array}{c|cc|} & \lambda_1 & k \\ \hline \lambda_1^{\sharp} & 1 & -a_{11}^{-1}a_{12} \\ \lambda_2^{\sharp} & 0 & 1 \\ \hline \end{array}\,.$$

whence

$$T(\{\lambda_1^{op}, k^{op}\}, \{\lambda_1, k\}) \quad = \quad \begin{array}{c|cc|} & \lambda_1 & \lambda_2 \\ \hline \lambda_1^{op} & a_{11} & 0 \\ k^{op} & 0 & \sigma \\ \hline \end{array}\,.$$

and

$$T(\lambda^{op}, \{\lambda_1, k\}) \quad = \quad \begin{array}{c|cc|} & \lambda_1 & k \\ \cline{2-3} \lambda_1^{op} & a_{11} & 0 \\ \lambda_2^{op} & a_{21} & \sigma \\ \cline{2-3} \end{array}\,,$$

$$T(\{\lambda_1^{op}, k^{op}\}, \lambda) \quad = \quad \begin{array}{c|cc|} & \lambda_1 & \lambda_2 \\ \cline{2-3} \lambda_1^{op} & a_{11} & a_{12} \\ k^{op} & 0 & \sigma \\ \cline{2-3} \end{array}\,.$$

Thus σ may be characterized as any of

$$k^{op}T\lambda_2 \qquad\qquad k^{op}Tk \qquad\qquad \lambda_2 Tk$$

where k is

$$\lambda_2 - \lambda_1 a_{12} a_{11}^{-1}, \qquad\qquad e^{\mathrm{I}(\lambda_1)}_{\mathrm{K}(\lambda_1^{op}T)}\lambda_2,$$

or the dual to $\lambda_2^\sharp$ in $\{\lambda_1^{op}T, \lambda_2^\sharp\}^\flat$, and k^{op} is

$$\lambda_2^{op} - \lambda_1^{op} a_{11}^{-1} a_{21}, \qquad\qquad \lambda_2^{op} e^{\mathrm{I}(T\lambda_1)}_{\mathrm{K}(\lambda_1^{op})},$$

or the dual to $\lambda_2^{op\,\flat}$ in $\{T\lambda_1, \lambda_2^{op\,\flat}\}^\sharp$.

6.5.1 Diagramatic Complements

The following characterization of σ will not enter our later discussion directly, however it is of basic algebraic interest. Some knowledge of the language of category theory is assumed.

Let D be the category on two objects and three morphisms, exactly one of which is not an endomorphism. For each preadditive category $\mathcal{C}$, let $[D, \mathcal{C}]$ be the preadditive category of diagrams $D \to \mathcal{C}$.

If

$$\upsilon^{op} = (\lambda_1^{op}, k^{op}) \qquad\qquad \textit{and} \qquad\qquad \upsilon = (\lambda_1, k),$$

then it may be checked directly that the diagrams

$$\begin{tikzcd} \bullet & \bullet \arrow[l, "a_{11}"'] \\ \bullet \arrow[u, "\upsilon_1^{op}"] & \bullet \arrow[l, "T"] \arrow[u, "\upsilon_1^\sharp"'] \end{tikzcd} \qquad\qquad \begin{tikzcd} \bullet & \bullet \arrow[l, "\sigma"'] \\ \bullet \arrow[u, "\upsilon_2^{op}"] & \bullet \arrow[l, "T"] \arrow[u, "\upsilon_2^\sharp"'] \end{tikzcd}$$

determine a product structure on the the diagram $D \to \mathcal{C}$ that sends the nonidentity morphism to T. The dual coproduct structure is given by

$$\begin{tikzcd}
\bullet \arrow[d, "\upsilon_1^{op\,\flat}"'] & \bullet \arrow[l, "a_{11}"'] \arrow[d, "\upsilon_1"] \\
\bullet & \bullet \arrow[l, "T"]
\end{tikzcd}
\qquad
\begin{tikzcd}
\bullet \arrow[d, "\upsilon_2^{op\,\flat}"'] & \bullet \arrow[l, "\sigma"'] \arrow[d, "\upsilon_2"] \\
\bullet & \bullet \arrow[l, "T"]
\end{tikzcd}$$

Remark 6.5.1. While there has been at least one category-theoretic treatment of the subject [76] the interpretation of a Schur complement as a literal complement to α in T in the category $[D, \mathcal{C}]$ has, to our knowledge, gone unremarked in the published literature.

6.6 Möbius Inversion

Let A be an array of form $T(\omega, \omega)$, where T is an endomorphism and ω is a (co)product structure on the domain of T. We write $\mathcal{P}^m$ for the set of all sequences of form

$$p : \{0, \dots, m\} \to \omega$$

and $\mathcal{P}$ for $\cup_{m \geq 0} \mathcal{P}^m$. Given any family of sequences $\mathcal{Q}$, we write

$$\mathcal{Q}(i, j) = \{q \in \mathcal{Q} : q(0) = i,\ q(\ell(q)) = j\}$$

where $\ell(q) = |\mathrm{D}(q)|$ is the *length* of q.

We define $A(p) = 1$ when $m = 0$ and

$$A(p) = A(p_0, p_1)A(p_1, p_2) \cdots A(p_{m-1}, p_m)$$

when $m > 0$. Consequently

$$A^m(i, j) = \sum_{\mathcal{P}^m(i,j)} A(p) \tag{6.6.1}$$

for nonnegative m.

Remark 6.6.1. Identity (6.6.1) is vacuous when $m \in \{0, 1\}$. It is the definition of matrix multiplication when $m = 2$.

Remark 6.6.2. Identity (6.6.1) remains valid if one replaces $\mathcal{P}^m(i, j)$ with

$$\mathcal{P}^m_A(i, j) = \{p \in \mathcal{P}^m(i, j) : A(p) \neq 0\}.$$

Given any pair of binary relations $\sim_I$ and $\sim_J$ on I and J, respectively, let us say that a map $p : I \to J$ is *increases monotonically* if

$$i \sim_I j \implies p(i) \sim_J p(j).$$

The following is a conceptually helpful characteristic of $\mathcal{P}^m_A$.

Lemma 6.6.3. *If*

$$p : \{0, \dots, m\} \to I$$

and $A(p) \neq 0$*, then* p *increases monotonically with respect to the transitive closure of* $\mathrm{Supp}(A)$ *and the canonical order on* $\{0, \dots, m\}$*.*

The main result of this section is a generalization of the classical *Möbius inversion formula* of number theory and combinatorics [74]. Lemma 6.6.4 expresses one part of this extension.

Lemma 6.6.4. *Suppose* $A = T(\omega, \omega)$*, where* T *is an endomorphism in a preadditive category and* ω *is a finite (co)product structure on* $\mathrm{D}(T)$*. If* $A = 1 - t$ *and* t *has strictly acyclic support, then*

$$A^{-1}(i, j) = \sum_{\mathcal{P}_t(i,j)} t(p) \tag{6.6.2}$$

for all $i, j \in I$*.*

Proof. Note that $t^{|I|} = 0$, by Lemma 7.1.1. Left multiplication by $1 - t$ therefore shows $(1 - t)^{-1} = \sum_{k=0}^{\infty} t^k$. The desired conclusion follows from (6.6.1). □

Remark 6.6.5. In the special case where $A = \chi_{\preceq}$ for some partial order $\preceq$ on I, identity (6.6.2) recovers the classical Möbius inversion formula.

We are now prepared to state and prove the main result. For convenience, let

$$A[p] = A(p_0, p_0)^{-1} A(p_0, p_1) A(p_1, p_1)^{-1} \cdots A(p_m, p_m)^{-1}.$$

Under the hypotheses of Lemma 6.6.4, one has

$$t(p) = (-1)^{\ell(p)+1} A[p],$$

so Theorem 6.6.6 is a strictly more general.

Theorem 6.6.6. *Let* $A = T(\omega, \omega)$*, where* T *is an automorphism in a preadditive category and* ω *is a finite (co)product structure on* $\mathrm{D}(T)$*. If* A *has acyclic support, then*

$$A^{-1}(i, j) = \sum_{\mathcal{P}_t(i,j)} (-1)^{\ell(p)+1} A[p],$$

where t *is the strictly-triangular part of* A*.*

Proof. Let d and t be the unique diagonal and off-diagonal arrays, respectively, such that $A = d - t$. If $B = d^{-1}A$ and $s = d^{-1}t$, then by definition

$$B(p_{k-1}, p_k) = A(p_{k-1}, p_{k-1})^{-1} A(p_{k-1}, p_k).$$

The terms on the righthand side of $(A^{-1}d)(i, j) = B^{-1}(i, j) = \sum_{\mathcal{P}_t(i,j)} s(p)$ are of the form

$$(-1)^{\ell(p)-1} A(p_0, p_0)^{-1} A(p_0, p_1) \cdots A(p_{m-1}, p_{m-1})^{-1} A(p_{m-1}, p_m).$$

Right-multiplication with d^{-1} yields the desired result. $\square$

Example 6.6.7. Let A be the following matrix 4×4 matrix with coefficients in $\mathbf{R}$,

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

One has

$$\mathcal{P}_A(1,4) = \{(1,4),(1,3,4)\},$$

so $A^{-1}(1,4) = 0$. Likewise

$$\mathcal{P}_A(2,4) = \{(2,4),(2,34)\} \qquad \mathcal{P}_A(1,2) = \emptyset \qquad \mathcal{P}_A(3,4) = \{(3,4)\}$$

whence

$$A^{-1}(2,4) = 0 \qquad A^{-1}(1,2) = 0 \qquad A^{-1}(3,4) = -1.$$

Altogether, A^{-1} has form

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

Chapter 7

Exchange Formulae

This section is devoted to the calculation of certain algebraic and relational identities involved in structure exchange. It is recommended that, on a first pass, the reader skim the definitions and results of §7.1 before passing directly to the following chapter. The facts cataloged in §7.2 may be accessed as needed for later portions of the discussion.

7.1 Relations

A *relation* on a set product $I \times J$ is a subset of $I \times J$. A *binary relation* on I is a relation on $I \times I$. It is standard to write iRj when $(i, j) \in R$.

We say that R is *reflexive* if iRi for all $i \in I$, *antisymmetric* if $i = j$ when both iRj and jRi, and *transitive* if iRk whenever iRj and jRk. A relation with all three properties is a *partial order*. A *linear order* is a partial order for which every pair of elements is comparable, that is, when iRj or jRi for all $i, j \in I$. The *transitive (reflexive) closure* of R is the least transitive (reflexive) relation that contains R.

The following is simple to verify for finite sets. Recall that S *extends* R, or R *extends to* S if $R \subseteq S$. A set that meets any of the conditions in Lemma 7.1.1 is *acyclic*. An acyclic relation is *strict* if it contains no pair of form (i, i).

Lemma 7.1.1. *For any binary relation R, the following are equivalent.*

1. *R extends to a partial order on I.*

2. *R extends to a linear order on I.*

3. *The transitive closure of R is antisymmetric.*

If R is strictly acyclic and i_pRi_{p+1} for $0 \leq p \leq m$, then I has cardinality strictly greater than m.

A *partial matching* on $R \subseteq I \times J$ is a subset $\mathrm{d} \subseteq R$ such that

$$\#\{(i, j) \in \mathrm{d} : i = i_0\} \leq 1 \qquad \textit{and} \qquad \#\{(i, j) \in \mathrm{d} : j = j_0\} \leq 1$$

for all $i_0 \in I$ and all $j_0 \in J$. To every partial matching we may associate a set $\mathrm{d}^{(\sharp)} = \{i : (i,j) \in \mathrm{d}\}$, called the *domain* of d, and a set $\mathrm{d}^{(\flat)} = \{j : (i,j) \in d\}$, called the *image*. A *perfect matching* is a partial matching with $\mathrm{d}^{(\sharp)} = I$ and $\mathrm{d}^{(\flat)} = J$.

Every partial matching determines a unique function $(\flat)$ sending $i \in \mathrm{d}^{(\sharp)}$ to the unique $i^{(\flat)}$ such that $(i, i^{(\flat)}) \in \mathrm{d}$. Symmetrically, there is a unique function $(\sharp)$ such that $(j^{(\sharp)}, j) \in \mathrm{d}$ for all j. These maps are mutually inverse bijections, and we refer to each as a *pairing function*.

Example 7.1.2. The *diagonal* of a set product $I \times I$ is the set

$$\Delta(I \times I) = \{(i,i) : i \in I\}.$$

The diagonal is a partial matching on $I \times I$. Each of the two pairing functions on $\Delta(I \times I)$ is the identity map on I.

A partial matching on R determines a pair of *induced relations*. The induced relation on I, denoted R_{d}, is the transitive reflexive closure of

$$\{(i,j) : (i, j^{(\flat)}) \in \mathrm{d}\}.$$

in $I \times I$. The induced relation on J, denoted R^{d}, is the transitive reflexive closure of

$$\{(i,j) : (i^{(\sharp)}, j) \in \mathrm{d}\}$$

in $J \times J$.

The *composition* of relations R and S is the set $RS = \{(i,k) : iRj,\ jSk\}$. We will be interested, in future discussion, in compositions of form $S = R_{\mathrm{d}} R R^{\mathrm{d}}$.

Proposition 7.1.3. *If $S = R_{\mathrm{d}} R R^{\mathrm{d}}$, then $R_{\mathrm{d}} = S_{\mathrm{d}}$ and $R^{\mathrm{d}} = S^{\mathrm{d}}$.*

Proof. Let us restrict to the first identity, as the second is similar. Relabeling elements as necessary, assume without loss of generality that there exists a set

$$K = \mathrm{d}^{(\sharp)} = \mathrm{d}^{(\flat)}$$

such that $\mathrm{d} = \Delta(K \times K)$. Then R_{d} and R^{d} are the transitive reflexive closures of the left and righthand intersections, respectively, in

$$R \cap (I \times K) \qquad\qquad R \cap (K \times K) \qquad\qquad R \cap (K \times J).$$

Let $R_{\mathrm{d}}^{\mathrm{d}}$ denote the transitive reflexive closure in K of the center intersection.

It is evident that $R_{\mathrm{d}} \subseteq S_{\mathrm{d}}$, so it suffices to show the reverse inclusion. For this, it is enough to establish that R_{d} contains $(R_{\mathrm{d}} R R^{\mathrm{d}}) \cap I \times K$. A pair (i,l) belongs to this set iff there exist j,k so that $i(R_d)j$, jRk, and $k(R^d)l$. Since $l \in K$, one must in fact have $(k,l) \in R_{\mathrm{d}}^{\mathrm{d}} \subseteq R_{\mathrm{d}}$, hence $(j,k) \in R_{\mathrm{d}}$. The desired conclusion follows. □

We say that a partial matching is *acyclic* on R if R^{d} and R_{d} are acyclic. Corollary 7.1.4 will enter later discussion as a simple means to relate the sparsity structures of arrays engendered by an LU decomposition with a variety of partial orders on their respective row and column indices.

Corollary 7.1.4. *A pairing is acyclic on R iff it is acyclic on $R_d R R^d$.*

Proof. This is an immediate consequence of 7.1.3. $\square$

7.2 Formulae

This section is devoted to the calculation of matrix identities involved in structure exchange. It is suggested that the reader skip this content, returning as needed where the various exchange formulae arise.

Relations

Recall that in §7.1 we defined a partial matching on a relation R. For economy, let us now define a partial matching on an *array* A to be a partial matching d on $\mathrm{Supp}(A)$ such that $A(\mathrm{d}^{(\sharp)}, \mathrm{d}^{(\flat)})$ is invertible. By abuse of notation, we will use the same symbols that denote the *sets* $\mathrm{d}^{(\sharp)}$ and $\mathrm{d}^{(\flat)}$ for the *product and coproduct maps* $\times\mathrm{d}^{(\sharp)}$ and $\oplus\mathrm{d}^{(\flat)}$, respectively.

It will be shown in §8.1 that if d is a partial matching and $A = T(\lambda, \upsilon)$ for some some coproduct (not product) structure λ and some product (not coproduct) structure υ, then the left- and right-hand families

$$\lambda - \left(\mathrm{d}^{(\sharp)}\right)^{\flat} \cup T\left(\mathrm{d}^{(\flat)}\right)^{\flat} \qquad\qquad \upsilon - \left(\mathrm{d}^{(\flat)}\right)^{\sharp} \cup \left(\mathrm{d}^{(\sharp)}\right)^{\sharp} T$$

form coproduct and product structures, respectively. Denote these by $\lambda[\mathrm{d}]$ and $\upsilon[\mathrm{d}]$, for brevity.

Recall that in order to avoid an excess of superscripts we sometimes drop the sharp and flat notation when composing maps, for instance writing

$$T\mathrm{d}^{(\flat)}$$

for $T\left(\mathrm{d}^{(\flat)}\right)^{\flat}$. There is no risk of ambiguity in this practice, as $T\left(\mathrm{d}^{(\flat)}\right)^{\sharp}$ is not defined.

If g is the dual to $T\mathrm{d}^{(\flat)}$ in $\lambda[\mathrm{d}]$, then $gT\mathrm{d}^{(\flat)} = 1$ by definition, so every partial matching on $T(\lambda, \upsilon)$ determines a partial matching d_0 on $T(\lambda[\mathrm{d}], \upsilon)$. We call the operation $(\lambda, \upsilon) \mapsto (\lambda[\mathrm{d}], \upsilon[\mathrm{d}_0])$ a *row-first pivot* on *d*. The row-first pivot defined in Section 8.1 is the special case of this operation, when $d = \{(a, b)\}$.

A remark on notation: while the support of $1(\lambda, \lambda[\mathrm{d}])$ is formally a relation on $\lambda \times \lambda[\mathrm{d}]$, it may be naturally realized as an relation on $\lambda \times \lambda$ by means of the associated pairing functions.

Lemma 7.2.1. *If* d *is a partial matching on $A = T(\lambda, \upsilon)$ and $R = \mathrm{Supp}(A)$, then*

1. *R_{d} is the transitive symmetric closure of* $\mathrm{Supp}(1(\lambda, \lambda[\mathrm{d}]))$
2. *R^{d} is the transitive symmetric closure of* $\mathrm{Supp}(1(\upsilon[\mathrm{d}_0], \upsilon))$

under the canonical identification

$$\lambda \times \lambda \leftrightarrow \lambda \times \lambda[d] \qquad\qquad \upsilon \times \upsilon \leftrightarrow \upsilon[\mathrm{d}_0] \times \upsilon$$

given by the pairing functions on d.

Proposition 7.2.2. *Let* d *be a partial matching on* $T(\lambda, \upsilon)$ *and* d_0 *the associated pairing on* $T(\lambda[d], \upsilon)$. *If* d *is a cyclic, then* d_0 *is acyclic, also.*

Proof. That d is a pairing on $T(\lambda[d], \upsilon)$ holds by fiat. To see that d is acyclic, let R denote the support of $T(\lambda, \upsilon)$, and note

$$T(\lambda[d], \upsilon) = 1(\lambda[d], \lambda)T(\lambda, \upsilon) = 1(\lambda, \lambda[d])^{-1}T(\lambda, \upsilon). \tag{7.2.1}$$

The transitive closure of $\mathrm{Supp}(1(\lambda, \lambda[d]))$ is the acyclic induced relation R_d, so the righthand side of (7.2.1) lies in $R_d R$, by Möbius inversion. Proposition 7.1.3 implies that d is acyclic in $R_d R R^d$, and so in (7.2.1) *a foriori.* □

Corollary 7.2.3. *Any transitive relation that extends the support of* $1(\upsilon[d], \upsilon)$ *extends the support of* $1(\upsilon[d_0], \upsilon)$, *also.*

Proof. If R is the support of $T(\lambda, \upsilon)$ then R^d is the transitive reflexive closure of the support of $1(\upsilon[d], \upsilon)$. Proposition 7.1.3 implies that any transitive relation extending R^d extends $(R_d R R^d)^{d_0}$, also. The latter extends $1(\upsilon[d_0], d)$, and the desired conclusion follows. □

Proposition 7.2.4. *An invertible array with acyclic support has exactly one perfect matching.*

Proof. Every invertible acyclic array has a column c supported on a single row, r. This row-column pair belongs to every perfect matching. The rows and columns complementary to r and c index a strictly smaller invertible acyclic array, and it may argued similarly that this, too, has a row-column pair that must be contained in any perfect matching. The desired conclusion follows by a simple induction. □

Proposition 7.2.5. *If* A *is invertible with acyclic support, then the transitive closures of* $\mathrm{Supp}(A)$ *and* $\mathrm{Supp}(A^{-1})$ *are identical.*

Proof. One has $\mathrm{Supp}(A^{-1}) \subseteq \mathrm{Supp}(A)$ by Möbius inversion. The reverse inclusion holds by symmetry. □

Schur Complements

Propositions 7.2.6, 7.2.7, and 7.2.8 are a elementary consequences of the splitting lemma (Lemma 6.4.4). The proofs are left as exercises to the reader.

Proposition 7.2.6. *If* $\lambda_0^{op} T \lambda_0$ *is an isomorphism, then its Schur complement in* $T(\lambda^{op}, \lambda)$ *is the block indexed by* $(\lambda_1^{op}, \lambda_1)$ *in the matrix representation of* T *with respect to* $(\lambda^{op}, \lambda)[\lambda_0^{op}, \lambda_0]$.

Proposition 7.2.7. *Let* d *be a partial matching on* $T(\lambda, \upsilon)$, *and let* f *be any element of* υ *not removed by the operation* $\upsilon \mapsto \upsilon[d_0]$. *Then the dual to* f *in* $\upsilon[d_0]^\flat$ *is*

$$e_k f^\flat$$

where e_k *is idempotent projection onto the null space of* $\mathrm{d}^{(\sharp)}T$ *along* $\mathrm{d}^{(\flat)}$*, and* $h^\sharp$ *is the dual to* f *in* υ*. Likewise, if* f^{op} *is any element of* λ *not removed by* $\lambda \mapsto \lambda[\mathrm{d}]$*, then the dual to* f^{op} *in* $\lambda[\mathrm{d}]^\sharp$ *is*

$$f^{op\,\sharp} e_{k^{op}}$$

where $e_{k^{op}}$ *is the natural opposite to* e_k *(that is, the idempotent so that* $e_{k^{op}} T \mathrm{d}^{(\flat)} = 0$ *and* $h e_{k^{op}} = h$ *iff* h *vanishes on the image of* $T\mathrm{d}^{(\flat)}$*), and where* $f^{op\,\flat}$ *is the dual to* f^{op} *in* λ*. Under these assumptions*

$$f^{op\,\flat} e_{k^{op}} T f^\sharp = f^{op\,\flat} e_{k^{op}} T e_k f^\sharp = f^{op\,\flat} T e_k f^\sharp. \tag{7.2.2}$$

The matrix representation of T *with respect to* $(\lambda, \upsilon)[\mathrm{d}]$ *has form*

$$diag(1, \sigma)$$

where σ *is the Schur complement of the invertible submatrix indexe* $\mathrm{d}^{(\sharp)} \cup \mathrm{d}^{(\flat)}$*. The elements of* σ *are given by* (7.2.2)*.*

Proposition 7.2.8. *Suppose that* $\{f, g, h\}$ *is a coproduct structure on* W*. If* $T^2 = 0$ *and* $g^\sharp T f$ *is invertible, then* (f, g) *is a Jordan pair, and* $e_k h$ *complements* Tf *in* $K(g^\sharp T)$*, where* e_k *is idempotent projection onto the kernel of* $g^\sharp T$ *along* f*. With respect to the ordered coproduct structure* $(e_k h, f, Tf)$*,* T *has matrix representation*

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

where $e_{k^{op}}$ *is the natural dual to* e_k *(that is, the idempotent so that* $e_{k^{op}} T f = 0$ *and* $h e_{k^{op}} = h$ *iff* h *vanishes on the image of* Tf*),* $h^\sharp$ *is the dual to* h *in* $\{f, g, h\}$*, and* τ *is any of the following*

$$h^\sharp e_{k^{op}} T h \qquad\qquad h^\sharp e_{k^{op}} T e_k h \qquad\qquad h^\sharp T e_k h.$$

That is, τ *is the submatrix indexed by* $(e_k h, e_k h)$ *in the Schur complement of* $g^\sharp T f$*.*

Möbius Inversion

Proposition 7.2.9. *Let* d *be an acyclic partial matching on* $T(\lambda, \upsilon)$*. If* $A = 1(\lambda, \lambda[\mathrm{d}])$ *and* $C = T(\lambda[\mathrm{d}], \upsilon[\mathrm{d}_0])$*, then*

$$C(f, g) = \begin{cases} \delta_{f^{(\flat)}, g} & f \in \mathrm{d}^{(\sharp)} \text{ or } g \in \mathrm{d}^{(\flat)} \\ \sum_{h \in \lambda} \sum_{\mathcal{P}_A(f,h)} (-1)^{\ell(p)+1} A[p] T(h, g) & \text{otherwise.} \end{cases}$$

Proof. The first case holds by fiat. For the second, let $B = T(\lambda[\mathrm{d}], \upsilon)$. It is a simple consequence of the splitting lemma (Lemma 6.4.4) that $C(f, g) = B(f, g)$ when $f \notin \mathrm{d}^{(\sharp)}$ and $g \notin \mathrm{d}^{(\flat)}$. Consequently the second case follows from the observation $B = 1(\lambda[\mathrm{d}], \lambda) T(\lambda, \upsilon) = 1(\lambda, \lambda[\mathrm{d}])^{-1} T(\lambda, \upsilon)$, plus Möbius inversion. $\square$

Let us say that a partial matching d is *Jordan* if it is a proper Jordan pair in the sense of Section 8.2. The proof of the following statement is entirely analogous to that of Proposition 7.2.9.

Proposition 7.2.10. *Suppose* $T^2 = 0$*, and let* d *be an acyclic Jordan pairing on* $T(\lambda, \upsilon)$*. If* $A = 1(\lambda[\mathrm{d}], \lambda)$ *and* C *is the matrix representation of* T *with respect to* $\lambda[[\mathrm{d}]]$*, then*

$$C(f,g) = \begin{cases} \delta_{f^{(\flat)},g} & \{f,g\} \cap (\mathrm{d}^{(\sharp)} \cup \mathrm{d}^{(\flat)}) \neq \emptyset \\ \sum_{h \in \lambda} \sum_{\mathcal{P}_A(f,h)} (-1)^{\ell(p)+1} A[p] T(h,g) & \text{otherwise.} \end{cases}$$

Chapter 8

Exchange Algorithms

8.1 LU Decomposition

Let us recall Lemma 6.4.4 and its logical equivalent, Lemma 6.4.6.

Lemma 6.4.4 (Splitting Lemma)**.** *Suppose that*

$$f^{op} f = 1,$$

where f^{op} and f are composable morphisms in a preadditive category.

1. *If k is kernel to f^{op}, then $\{f, k\}$ is a coproduct structure.*

2. *If k^{op} is cokernel to f, then $\{f^{op}, k^{op}\}$ is a product structure.*

Remark 8.1.1. As noted in §6.4.1, the criterion $f^{op} f = 1$ may be loosened to the condition that $f^{op} f$ be invertible.

As previously observed, the main application of the splitting lemma in this discussion is to clarify the relationship between certain dual structures. Suppose, for example, that $\lambda = \{\lambda_f, \lambda_g\}$ and $\upsilon = \{\upsilon_f, \upsilon_h\}$ are products. How does $\lambda_f^\flat$ relate to $\upsilon_f^\flat$ when $\lambda_f = \upsilon_f$? Recall that every k in a (co)product ω engenders an idempotent $e_k = k^\flat k^\sharp$. It is elementary to show that $\lambda_f^\flat = e_k \upsilon_f^\flat$ when $\omega = \{\lambda_f^\flat, k\}$ and k is kernel to λ_f. It can be shown similarly that $\lambda_g^\flat = \lambda_h^\flat$ when $\lambda_g \lambda_h^\flat = 1$. If λ and υ are coproducts, then $\lambda_f^\sharp = \upsilon_f^\sharp e_{k^{op}}$ whenever k^{op} is cokernel to λ_f; likewise $\lambda_g^\sharp = \lambda_h^\sharp$, when $\lambda_g^\sharp \lambda_h = 1$. These identities represent what are essentially a variety of different access points to the same underlying mathematical structure, and we group all under the common heading of the splitting lemma.

The main application of duality, for this discussion, is to describe exchange. Suppose λ and υ are coproduct and product structures, respectively, on the codomain and domain of an operator T. An *(elementary) exchange pair* for λ and υ is an ordered pair $(a, b) \in \lambda \times \upsilon$ such that $a^\sharp T b^\flat$ is invertible. By the Exchange Lemma, (a, b) is an exchange pair if and only if $\lambda[a, b^\flat] = \lambda - a \cup T b^\flat$ is a coproduct and $\upsilon[a^\sharp, b] = \upsilon - b \cup a^\sharp T$ is a product. We call the operation that produces these structures *elementary exchange*.

Lemma 6.4.6 (Exchange Lemma)**.** *Let $\{f, g\}$ and $\{f^{op}, g^{op}\}$ be product and coproduct structures, respectively.*

1. *An unordered pair $\{f, h\}$ is a product iff $g^\sharp h$ is invertible.*

2. *An unordered pair $\{f^{op}, h^{op}\}$ is a coproduct iff $h^{op}(g^{op})^\flat$ is invertible.*

Individual exchanges may be combined in an important way. Suppose $\sharp$ and $\flat$ are the usual sharp and flat operators on λ and υ, and let $\sharp\sharp$, $\flat\flat$ denote the corresponding operators on $\lambda[a, b^\flat]$ and $\upsilon[a^\sharp, b]$, respectively. The composition $(Tb^\flat)^{\sharp\sharp}T(b^\flat)$ is identity by definition, so the righthand side of

$$\lambda[a, b^\flat] \qquad\qquad \upsilon[(Tb^\flat)^{\sharp\sharp}, b] \tag{8.1.1}$$

is a product. We write $(\lambda, \upsilon)[a, b]$ for the pair (8.1.1), and refer to the operation

$$(\lambda, \upsilon) \mapsto (\lambda, \upsilon)[a, b]$$

as a *row-first pivot* on (a, b). There is a natural mirror to this operation, the *column-first pivot*, which maps υ to $\upsilon[a^\sharp, b]$ and λ to $\lambda[a, (a^\sharp T)^{\flat\flat}]$. Column pivots will not enter our story directly, though an equivalent, antisymmetric story may be told exclusively in terms of these operations.

The following algorithm initiates with a pair $(\alpha, \beta) = (\lambda, \upsilon)$, where υ is a product on the domain of T and λ is a coproduct on the codomain of T. To avoid degeneracies, we stipulate that all elements of λ and υ have rank one, so that $\alpha \times \beta$ fails to contain an exchange pair iff $(\times\alpha^\sharp)T(\oplus\beta^\flat) = 0$.

Algorithm 1 LU Decomposition

```
while α × β contains an exchange pair do
  fix an exchange pair (a, b) ∈ α × β.
  (λ, υ) ← (λ, υ)[a, b]
  (α, β)− = (a, b)
end while
```

Let λ^p and υ^p denote the coproduct and product structures generated on the pth iteration of Algorithm 1, with ∞ denoting the final iteration. For convenience, let us index λ and λ^∞ so that the elements of λ^p may be arranged into a a tuple of form $(\lambda^\infty_1, \ldots, \lambda^\infty_p, \lambda_{p+1}, \ldots, \lambda_m)$ for each p, and order the elements of υ^p into a tuple $(\upsilon^\infty_1, \ldots, \upsilon^\infty_p, \upsilon_{p+1}, \ldots, \upsilon_n)$, similarly.

Remark 8.1.2. One benefit of working with indexed families is that there is no need for the special notation to differentiate between the various sharp (respectively, flat) operators. Whereas previously one had to write double scripts $\sharp\sharp$ and $\flat\flat$, under this convention one simply has $\upsilon^p_p = (\lambda^p_p)^\sharp T$ and $\lambda^p_p = T(\upsilon^{p-1}_p)^\flat$.

Lemma 8.1.3. *For any $r \geq p$, one has $(\upsilon^r_p)^\flat = (\upsilon^{p-1}_p)^\flat$ and $(\lambda^r_p)^\sharp = (\lambda^p_p)^\sharp$. Moreover,*

$$(\lambda^p_q)^\sharp T(\upsilon^p_p)^\flat = \delta_{qp} \qquad\qquad (\lambda^p_p)^\sharp T(\upsilon^p_q)^\flat = \delta_{qp} \tag{8.1.2}$$

for any p, q.

Proof. Since $(\lambda_p^p)^\sharp T(\upsilon_p^{p-1})^\flat = 1$ by definition, the splitting lemma implies $(\upsilon_p^p)^\flat = (\upsilon_p^{p-1})^\flat$. Thus $\lambda_p^p = T(\upsilon_p^p)^\flat$ and $\upsilon_p^p = (\lambda_p^p)^\sharp T$, whence (8.1.2). It follows from these identities that $(\lambda_p^p)^\sharp$ and $(\upsilon_p^p)^\flat$ remain invariant under any increase in upper index, again by the splitting lemma. □

Lemma 8.1.4. *For any p and q one has*

$$\lambda_p^\sharp \lambda_q^\infty = (\lambda_p^{q-1})^\sharp T(\upsilon_q^{q-1})^\flat \qquad\qquad \upsilon_p^\infty \upsilon_q^\flat = (\lambda_p^p)^\sharp T(\upsilon_q^{q-1})^\flat. \tag{8.1.3}$$

The left-hand operator vanishes when $p < q$; the right-hand when $q < p$.

Proof. Let f denote the righthand side of the lefthand identity. When $p < q$ one may increase the upper indices in this expression by one without changing its value, and consequently f vanishes, by (8.1.2). Since $\lambda^{p-1} = (\lambda_1^\infty, \dots, \lambda_{p-1}^\infty, \lambda_p, \dots, \lambda_m)$, it follows from the splitting lemma that $f = \lambda_p^\sharp T(\upsilon_q^{q-1})^\flat$ for any p and q. This suffices for the lefthand identity, as $T(\upsilon_q^{q-1})^\flat = \lambda_q^\infty$ by definition. The righthand identity and may be argued similarly, as may its vanishing properties. □

In the language of arrays, one may interpret Lemmas 8.1.3 and 8.1.4 to say that

$$\begin{aligned} 1(\lambda, \lambda^p) &= [\ \mu_1 \mid \cdots \mid \mu_p \mid \chi_{p+1} \mid \cdots \mid \chi_m\ ] \\ 1(\upsilon^p, \upsilon) &= [\ \nu_1 \ / \ \cdots \ / \ \nu_p \ / \ \chi_{p+1} \ / \ \cdots \ / \ \chi_n\ ] \end{aligned}$$

are lower and upper triangular, respectively, where μ_p is the pth column of $T(\lambda^{p-1}, \upsilon^{p-1})$ and ν_p is the pth row of $T(\lambda^p, \upsilon^{p-1})$. Moreover,

$$T(\lambda^\infty, \upsilon^\infty) = \mathrm{diag}(1, \dots, 1, 0, \dots, 0)$$

with nonzero entries appearing on exchange pairs. Thus

$$T(\lambda, \upsilon) = 1(\lambda, \lambda^\infty) T(\lambda^\infty, \upsilon^\infty) 1(\upsilon^\infty, \upsilon)$$

where $1(\lambda, \lambda^\infty)$ is lower triangular, $1(\upsilon^\infty, \upsilon)$ is upper-triangular, and $T(\lambda^\infty, \upsilon^\infty)$ is a zero-one array with at most one nonzero entry per row and column. If λ and υ are composed of maps into and out of $\mathbf{k}$, then the entries in these arrays are maps $f : \mathbf{k} \to \mathbf{k}$. Under the standard identification $f \leftrightarrow f(1)$, the preceding identity corresponds exactly to an LU decomposition of $T(\lambda, \upsilon)$.

Remark 8.1.5. The preceding observations did not depend on our restriction to the category of linear maps on finite-dimensional vectors spaces. Rather, they yield an LU decomposition for any biproduct of finitely many simple objects in a preadditive category. The existence of such a factorization was remarked by Smith [76], though we are unaware of combinatorial treatments of the subject.

8.2 Jordan Decomposition

The reader may replace the phrase *endomorphism in an abelian category* with the phrase *linear map $W \to W$* in the statement of the following proposition, with no loss of correctness.

Theorem 8.2.1. *Let T be an endomorphism in an arbitrary abelian category. If $T^{n+1} = 0$ and $gT^n f = 1$, then there is a coproduct structure*

$$(k, T^0 f, \dots, T^n f) \tag{8.2.1}$$

with respect to which T has form

$$\begin{pmatrix} * & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0 & \cdots & 1 & 0 \end{pmatrix}. \tag{8.2.2}$$

Proof. Suppose that $T : W \to W$, set

$$\mathrm{K}_n = \mathrm{K}(gT^n),$$

and let K_n denote the inclusion $\mathrm{K}_n \to W$. By the splitting lemma, $\{f, K_n\}$ is a coproduct structure. If we identify gT^{n-1} and Tf with the unique maps $\mathrm{K}_n \to \mathrm{D}(f)$ and $\mathrm{D}(f) \to \mathrm{K}_n$, respectively, such that the following diagram commutes

$$\begin{tikzcd}
 & W \arrow[dl, "gT^{n-1}"'] & \\
\mathrm{D}(f) & & \mathrm{D}(f) \arrow[ul, "Tf"'] \arrow[dl] \\
 & \mathrm{K}_n \arrow[uu, "K_n"'] \arrow[ul] &
\end{tikzcd}$$

then $\{Tf, K_{n-1}\}$ is likewise a coproduct structure on K_n, where K_{n-1} includes

$$\mathrm{K}_{n-1} = \mathrm{K}(gT^{n+1}) \cap \mathrm{K}(gT^{n-1})$$

into W. By a simple induction, if

$$\mathrm{K}_p = \mathrm{K}(gT^n) \cap \cdots \cap \mathrm{K}(gT^p)$$

and K_p is the inclusion $\mathrm{K}_p \to W$, then $\{T^p f, K_{n-p}\}$ is a product structure on K_{n-p+1}, for $p \in \{0, \dots, n\}$. Thus $(K_0, T^0 f, \dots, T^n f)$ is an ordered coproduct structure.

If h is dual to $T^n f$ in this structure, then $hT^n f = 1$ by definition, and we may repeat the preceding argument to obtain a coproduct structure $(k, T^0 f, \ldots, T^n f)$ where

$$k = \mathrm{K}(hT^n) \cap \cdots \cap \mathrm{K}(hT^0).$$

It is elementary to check that if h is dual to $T^n f$ in this structure. Consequently, hT^{n-p} is dual to $T^p f$, for $p \in \{0, \ldots, n\}$. The desired conclusion follows. $\square$

By analogy with linear spaces, let us say that a map f *complements* g if $\{f, g\}$ is a coproduct structure. We refer to any map k that realizes the conclusion to Theorem 8.2.1 as a *Jordan complement* of f.

This proposition points to a general method for computing Jordan complements, which may be described as follows. Suppose that ω is a coproduct structure of rank-one maps into W. If T^{n+1} vanishes and T^n does not, then there exists an $f \in \omega$ such that $T^n f \neq 0$. For such f there exists at least one

$$g \in \omega - \{f\}$$

so that $g^\sharp T f$ is an isomorphism – otherwise either $T^n f$ would vanish or T would fail to be nilpotent. The pair (g, f) is then an exchange pair, and we may write

$$\omega^1 = \omega \cup \{Tf\} - \{g\}$$

for the coproduct obtained by exchanging Tf for g. The map Tf indexes a column of $T(\omega^1, \omega^1)$, so provided that $n > 1$ we may find a

$$g \in \omega^1 - \{f, Tf\}$$

so that $g^\sharp T^2 f$ is invertible, and set

$$\omega^2 = \omega^1 \cup \{T^2 f\} - \{g\}.$$

This process repeats. To see that on each iteration there exists a

$$g \in \omega^p - \{T^0 f, \ldots, T^p f\}$$

such that $g^\sharp T^p f$ is an isomorphism, we it helpful to visualize the sparsity pattern of $T(\omega^p, \omega^p)$. Once one recognizes that the columns indexed by $T^0, \ldots, T^{p-1} f$ are zero-one arrays concentrated on distinct elements of ω^p, the desired conclusion becomes clear.

For convenience, let us sum the elements of ω^n that do not take the form $T^p f$ into a single complement g, thus forming a coproduct

$$(g, T^0 f, \ldots, T^n f).$$

Let h denote the dual to $T^n f$ in this structure. If hT^{n-p} is dual to $T^n f$ for all p, then we are done, but such will not always be the case. It will be true, however, that the composition

$$(hT \times \cdots \times hT^n) \circ (T^{n-1} f \oplus \cdots \oplus T^0 f)$$

is identity, since the matrix representation of T with respect to ω^n agrees with (8.2.2) in all but the first column. Thus we may exchange $\{hT, \dots, hT^n\}$ for the duals to $\{T^0 f, \dots, T^{n-1} f\}$ in $(\omega^n)^\sharp$.

By the splitting lemma, the resulting coproduct structure will have form

$$(eg, T^0 f, \dots, T^{n-1} f, eT^n f),$$

where e is projection onto

$$\mathrm{K}_1 = \mathrm{K}(hT^1 \times \cdots \times hT^n)$$

along $T^0 f \oplus \cdots \oplus T^{n-1} f$. Since $T^n f$ factors through $\mathrm{K}(T) \subseteq \mathrm{K}_1$, this structure agrees with

$$(eg, T^0 f, \dots, T^n f). \tag{8.2.3}$$

The matrix representation of T with respect to (8.2.3) will agree with (8.2.2) in all but the second row of the first column. This too must agree, however, since otherwise T^{n+1} would fail to vanish. Thus we have established the following.

Corollary 8.2.2. *If e is the idempotent operator on W that realizes projection onto K_1 along $T^0 f \oplus \cdots \oplus T^{n-1} f$, then eg is a Jordan complement to f.*

Corollary 8.2.3. *A map k is a Jordan complement to f iff k complements $T^n f$ in the kernel of a morphism H such that*

$$H \circ (Tf \oplus \cdots \oplus T^n f) = 1.$$

Let us now describe a general algorithm to compute Jordan decompositions (not just Jordan complements) for a nilpotent morphisms. For convenience, denote the structure (8.2.3) by $\omega[[f, h]]$. We call (f, h) a *Jordan pair* in T, and refer to the operation

$$\omega \mapsto \omega[[f, h]]$$

as a *Jordan exchange.*

Informally, the decomposition algorithm works by splitting orbits off of Jordan complements. On each iteration is given a coproduct structure ω, expressed as a disjoint union of form $J_1 \cup \cdots \cup J_p \cup I$, where $J_1, \dots, J_p$ are maximal (with respect to inclusion) orbits. One selects an element $f \in \mathrm{argmax}_{i \in I} |\mathrm{Orb}(i)|$, and rotates first $\mathrm{Orb}(f)$, and then the necessary duals to $\mathrm{Orb}(f)$ into the biproduct structure. The result is is a biproduct a strictly greater number of orbits and a strictly smaller complement, I. Once this complement vanishes, the process terminates.

To aid in a formal description, let us say that (f, h) is a *proper Jordan pair* if (i) it is a Jordan pair, (ii) $\mathrm{Orb}(f)$ is not already a Jordan block in ω, and (iii) among the pairs that meet criteria (i) and (ii), $\mathrm{Orb}(f)$ has maximal cardinality.

If $T^2 = 0$, then the structure produced on iteration p of Algorithm 2 bears a simple relationship to that of step $p - 1$. Let us denote the pth Jordan pair by (f_p, h_p), and the corresponding coproduct structure ω^p. In the 2-nilpotent regime only one element rotates

Algorithm 2 Jordan Decomposition

while ω has a proper Jordan pair **do**
  fix a proper Jordan pair for ω.
  $\omega \leftarrow \omega[[f, h]]$
end while

directly into the coproduct on each iteration, namely Tf_p, and one into the product. Since neither f_p nor Tf_p are modified by subsequent iterations Algorithm 2 (a consequence of the splitting lemma) we may arrange the elements of ω^p into a tuple of form

$$(f_1, \dots, f_p, \dots, Tf_p, \dots, Tf_1). \tag{8.2.4}$$

Proposition 8.2.4. *Array* $1(\omega, \omega^\infty)$ *is upper-triangular with respect to* (8.2.4)*.*

Proof. Let n be the dimension of W. With respect to the given ordering, the off-antidiagonal support of $T(\omega^{p-1}, \omega^{p-1})$ concentrates on $[p, n-p] \times [p, n-p]$. If υ^{p-1} is the coproduct produced by rotating Tf_p into ω^{p-1}, then the same assertion holds for the support of $T(\upsilon^{p-1}, \upsilon^{p-1})$. Consequently array $1(\omega^{p-1}, \upsilon^p)$, which may be obtained by replacing the pth column of the identity array by the pth column of $T(\omega^{p-1}, \omega^{p-1})$, is upper-triangular. The same holds for $1(\upsilon^p, \omega^{p+1})$, by a similar argument, and so, too for $1(\omega^{p-1}, \omega^p) = 1(\omega^{p-1}, \upsilon^p)1(\upsilon^p, \omega^p)$. The desired conclusion follows by a simple induction. $\square$

8.3 Filtered Exchange

We define the *closure* of a collection λ of maps into W to be the class $\mathrm{cl}(\lambda)$ of all g for which some diagram of form

$$\begin{tikzcd} \bullet \arrow[dr, "\oplus\lambda"'] & & \bullet \arrow[ll] \arrow[dl, "g"] \\ & W & \end{tikzcd}$$

commutes.

Likewise, the *coclosure* of a collection υ of maps out of W is the class $\mathrm{cl}(\upsilon)$ of all g for which some diagram of form

$$\begin{tikzcd} \bullet \arrow[rr] & & \bullet \\ & W \arrow[ul, "\times\lambda"] \arrow[ur, "g"'] & \end{tikzcd}$$

commutes. Coclosures will not enter our discussion directly, however an equivalent dual story may be told in terms of this operations.

Given any family Ω of maps into W, define $\mathrm{cl}_\Omega(\omega) = \Omega \cap \mathrm{cl}(\omega)$. We say ω is *closed* if $\omega = \mathrm{cl}(\omega)$, and *closed in* Ω if $\omega = \mathrm{cl}_\Omega(\omega)$. A set is *independent* in Ω if no $f \in \omega$ lies in the closure of $\omega - \{f\}$.

In the special case where $\Omega = \mathrm{Hom}(\mathbf{k}, W)$, the closure of ω in Ω is the family of maps $\mathbf{k} \to W$ whose image lies in $\mathrm{I}(\oplus\omega)$. Consequently, ω is independent in Ω iff $\{f(1) : f \in \omega\}$ is independent as a subset of W. Thus the independent sets of Ω are the linearly independent subsets of $\mathrm{Hom}(\mathbf{k}, W)$. The benefit of this alternate characterization is that it describes independence in the language of function composition, hence the language of biproducts. The following, for example, is an elementary consequence of the splitting lemma.

Lemma 8.3.1. *If λ is a coproduct structure on W and $\omega \subseteq \lambda$, then $g \in cl(\omega)$ iff $(\lambda - \omega)^\sharp g = 0$.*

Let us apply this observation to a problem involving minimal bases. Suppose are given a linear filtration $\mathcal{F}$ on Ω, and an $\mathcal{F}$-minimal basis λ. Under what conditions will a second basis, υ, be minimal as well?

Denote the relation $\{(f, g) \in \Omega : \mathcal{F}(f) \le \mathcal{F}(g)\}$ by $\sim_\mathcal{F}$, and recall that a *partial matching* on $T(\lambda, \upsilon)$ is a partial matching d on $R = \mathrm{Supp}(T(\lambda, \upsilon))$ such that $T\left(\mathrm{d}^{(\sharp)}, \mathrm{d}^{(\flat)}\right)$ is invertible. Every exchange pair (f, g) determines a partial paring $d = \{(f, g)\}$, and while the support of $1(\lambda, \lambda[\mathrm{d}])$ is formally a relation on $\lambda \times \lambda[\mathrm{d}]$ we may naturally regard it as subset of $\lambda \times \lambda$, via the pairing functions. The relation $R_\mathrm{d} \subseteq \lambda \times \lambda$ coincides with the transitive reflexive transitive closure of

$$\mathrm{Supp}(1(\lambda, \lambda[d]))$$

under this identification. A similar interpretation holds for R^d.

Proposition 8.3.2. *If R is the support of $1(\lambda, \upsilon)$, then υ is $\mathcal{F}$-minimal iff $\sim_\mathcal{F}$ extends R_d for some perfect matching* d *on R.*

Proof. Let us assume for convenience that λ and υ are ordered tuples, and reindex as necessary so that $\mathcal{F} \circ \lambda$ and $\mathcal{F} \circ \upsilon$ are monotone-increasing functions of form

$$\{1, \ldots, m\} \to \mathbb{Z}.$$

The first condition holds iff $\mathcal{F} \circ \lambda = \mathcal{F} \circ \upsilon$. Lemma 8.3.1 therefore implies υ is minimal iff the first n_p columns of $1(\lambda, \upsilon)$ have support on the first n_p rows, where $n_p = |\lambda_{\mathcal{F} \le p}|$. This is equivalent to the condition that the array be block-upper triangular, with diagonal blocks of size $n_p - n_{p-1}$. As $1(\lambda, \upsilon)$ it is invertible, these blocks must be, also. Any perfect matching that draws its elements from the support of these diagonal blocks will extend to $\sim_\mathcal{F}$. This establishes one direction. The converse may be shown similarly. □

Suppose now that λ and υ are bases (equivalently, coproduct structures) in $\mathrm{Hom}(\mathbf{k}, W)$ and $\mathrm{Hom}(\mathbf{k}, V)$. Posit weight functions $\mathcal{F}$ and $\mathcal{G}$ on λ and υ, respectively, and extend these to the weights on $\mathrm{Hom}(\mathbf{k}, W)$ and $\mathrm{Hom}(\mathbf{k}, V)$ such that

$$\mathcal{F}_{\le p} = \mathrm{cl}_\Omega(\lambda_{\mathcal{F} \le p}) \qquad\qquad \mathcal{G}_{\le p} = \mathrm{cl}_\Omega(\upsilon_{\mathcal{G} \le p}).$$

We say that (f, g) is $\mathcal{F}$-$\mathcal{G}$ *minimal* if

$$(R_d, R^d) \subseteq (\sim_{\mathcal{F}}, \sim_{\mathcal{G}}),$$

where $\mathrm{d} = (f, g)$. Equivalently, (f, g) is $\mathcal{F}$-$\mathcal{G}$ minimal if $\mathcal{F}(f)$ is the minimum value taken by $\mathcal{F}$ on the row support of $T(\lambda, \{g\})$ and $\mathcal{G}(g)$ is the maximum value taken by $\mathcal{G}$ on the column support of $T(\{f\}, \upsilon)$. It is vacuous that the exchange pairs in any of the preceding algorithms may be chosen to be minimal, since the candidate pairs on each iteration are the nonzero elements in block N_1 of a block-diagonal array of form $\mathrm{diag}(N_0, N_1)$.

Algorithm 3 LU Decomposition (Filtered)

```
while α × β contains an exchange pair do
    fix a minimal exchange pair (a, b) ∈ α × β.
    (λ, υ) ← (λ, υ)[a, b]
    (α, β)− = (a, b)
end while
```

Proposition 8.3.3. *The bases λ^∞ and υ^∞ returned by Algorithm 3 are $\mathcal{F}$-minimal and $\mathcal{G}$-minimal, respectively.*

Proof. Let R denote the support of $1(\lambda, \lambda^\infty)$. Since $1(\lambda, \lambda^\infty)$ is triangular up to permutation, R has a unique perfect matching d. Relation $\sim_{\mathcal{F}}$ extends R_{d} by construction, so minimality for λ^∞ follows by Proposition 8.3.2. Minimality for υ^∞ may be argued similarly. □

A similar result holds for the Jordan algorithm in the special case where $T^2 = 0$. As with LU decomposition, we are guaranteed to be able to restrict our selection to minimal exchange pairs thanks to block-diagonal structure of $T(\omega, \omega)$.

Algorithm 4 Jordan Decomposition (Filtered)

```
while ω has a proper Jordan pair do
    fix an (F, F)-minimal proper Jordan pair for ω.
    ω ← ω[[f, h]]
end while
```

Proposition 8.3.4. *The structure ω^∞ returned by Algorithm 4 is $\mathcal{F}$-minimal.*

Proof. Array $1(\lambda^\infty, \lambda)$ is a product of factors of form $1(\lambda[\mathrm{d}], \lambda)$ and $1(\upsilon, \upsilon[\mathrm{d}])$. Relation $\sim_{\mathcal{F}}$ extends the induced relations of $1(\lambda, \lambda[\mathrm{d}])$ and $1(\upsilon[\mathrm{d}], \upsilon)$ by minimality, and a trivial application of Möbius inversion shows that it extends the induced relations of the factors of $1(\lambda^\infty, \lambda)$, also. The desired conclusion follows by Proposition 8.3.2. □

8.4 Block exchange

Recall that in Section 7.2 we introduced *set exchange*, the natural extension of elementary exchange where entire sets, rather than single elements, rotate into and out of a (co)product.

The analysis of preceding sections carries through with minimal modification for set operations, for instance returning arrays $1(\lambda, \lambda^{\infty})$ and $1(\upsilon^{\infty}, \upsilon)$ that are *block* lower and upper triangular, respectively, for Algorithm 1, with respect to the natural grouping of elements by iteration. Let us note some special cases.

Acyclic blocks

Suppose that block-elimination is carried out exclusively for acyclic pairings. Such pairings index triangular-up-to-permutation blocks in the associated arrays, so $1(\lambda, \lambda^{\infty})$ and $1(\upsilon^{\infty}, \upsilon)$ will be triangular up to permutation. The same holds for the block form of Jordan elimination, though here one must, as usual, invoke Proposition 7.1.3 in tandem with Möbius inversion.

Minimal blocks

The filtered versions of Algorithms 1 and 2 have natural block-generalizations as well. Here the minimality requirement for pairs is replaced by the (rather more direct) criterion that the relations induced by d on $I \times I$ and $J \times J$ extend to $\sim_{\mathcal{F}}$ and $\sim_{\mathcal{G}}$, respectively. Correctness may be argued exactly as for Algorithms 3 and 4.

Linear blocks

The two preceding cases coincide when $\mathcal{F}$ and $\mathcal{G}$ determine linear orders on λ and υ. When such is the case, minimal pairings coincide exactly with acyclic pairings *by definition*. When $\mathcal{F}$ and $\mathcal{G}$ do not induce linear orders, one may construct modified functions $\mathcal{F}'$ and $\mathcal{G}'$ for which $\sim_{\mathcal{F}'}$ and $\sim_{\mathcal{G}'}$ are linear orders contained in $\mathcal{F}$ and $\mathcal{G}$. Any bases for these modified functions will be minimal with respect to $\mathcal{F}$ and $\mathcal{G}$, by Proposition 8.3.2. In fact, one may pick a different $\mathcal{F}'$ and $\mathcal{G}'$ on each iteration, thanks to the transitivity of $\sim_{\mathcal{F}}$ and $\sim_{\mathcal{G}}$.

In this linear setting, the minimal pairings have a special structure. The proof of the following is vacuous.

Lemma 8.4.1. *If $\sim_{\mathcal{F}}$ and $\sim_{\mathcal{G}}$ are linear orders, then the set S of all $(\mathcal{F}, \mathcal{G})$-minimal pairs is itself a minimal acyclic pairing. Every minimal pairing is a subset of S, in this case.*

Remark 8.4.2. The structure of minimal pairings for linear $\sim_{\mathcal{F}}$ and $\sim_{\mathcal{G}}$ is highly natural. It plays a foundational role in work of M. Kahle on probabilistic topology [45], and has been discussed by Carlsson in reference to persistent homology calculations, in personal correspondence. More recently, the minimal pairs of a linear order have been remarked as "obvious pairs" in the work of U. Bauer on fast persistence algorithms, in reference to the computational library *Ripser* [4].

Graded blocks

In the following chapter we introduce the notion of a graded operator on a vector space $\oplus_p C_p$. To such a space one may associate a relation $u \sim v$ iff $\{u, v\} \subseteq C_p - \{0\}$ for some p. This relation is transitive, and so falls under the umbrella of Corollary 7.2.3. It may thus be shown that any variant on the Jordan or LU exchange algorithms will return a graded basis, provided that the initial (co)product structures are graded.

Part III

Applications

Chapter 9

Efficient Homology Computation

9.1 The linear complex

A $\mathbb{Z}$-grading on a $\mathbf{k}$-linear space C is a $\mathbb{Z}$-indexed family of linear subspaces $(C_p)_{p\in\mathbb{Z}}$ such that the coproduct map

$$\oplus_p C_p \to C$$

induced by the inclusions $C_p \to C$ is an isomorphism.

Remark 9.1.1. This definition agrees with that of a grading on the matroid $(C, \mathcal{I})$, where $\mathcal{I}$ is the family of $\mathbf{k}$-linearly independent subsets of C, c.f. Example 5.3.1.

Remark 9.1.2. In the special case where $C = \oplus_p C_p$, we call (C_p) the *canonical* grading on C.

If C and D are graded spaces, then a map $T : C \to D$ is *graded of degree k* if

$$TC_p \subseteq TD_{p+k}$$

for all $p \in \mathbb{Z}$. A degree -1 endomorphism on C is a *differential* if $T^2 = 0$. A *linear complex* is a pair (C, ∂), where C is a graded vector space and ∂ is a differential on C.

It is customary to write

$$Z_p = K(\partial) \cap C_p \qquad B_p = I(\partial) \cap C_p \qquad \partial_p = \partial|_{C_p}$$

when working with complexes. We call Z_p the space of *cycles*, B_p the space of *boundaries*, and ∂_p the *p-dimensional boundary operator*. Where context leaves room for doubt, the first and second of these may be expressed $Z_p(C)$ and $B_p(C)$, respectively. The *homology group in dimension p*, or simply the *pth homology group* of C is the quotient space

$$H_p(C) = Z_p/B_p.$$

By mathematical synecdoche, one generally denotes (C, ∂) by C alone, the associated differential being understood from context. For example, a *map* from one complex to another is a degree zero commutator of the respective differentials. In symbols, this means a map $T : C \to D$ for which $T\partial = \partial T$. The differential on the left is that of C, and that on the right is that of D. The situation of these maps relative to T leaves no ambiguity as to the intended meaning.

Recall that the quotient of a vector space W by a subspace U is canonically realized as the set of equivalence classes $[w] = \{w + u : u \in U\}$ equipped with an addition $[w] + [v] = [w + v]$ and a scalar multiplication $\alpha \cdot [w] = [\alpha \cdot w]$. It is an foundational fact of homological algebra that every map of chain complexes induces a map on homology groups

$$T_* : H_p(C) \to H_p(D)$$

determined by the rule

$$T_*[w] = [Tw].$$

Remark 9.1.3. If we define $H(C) = K(\partial)/I(\partial)$, then there exists a well-defined linear map $H_p(C) \to H(C)$, defined by the rule $[v] \mapsto [v]$. If we identify $H_p(C)$ with its image under this inclusion, then the $\mathbb{Z}$ indexed family of homology groups $H_p(C)$ determines a natural grading on $H(C)$.

A *combinatorial simplicial complex* $\mathcal{X}$ on ground set V is a family of subsets of V closed under inclusion. Recall that this means $I \in V$ when $J \in V$ and $I \subseteq J$. An *ordered* combinatorial simplicial complex is a pair $(\mathcal{X}, <)$, where $<$ is a linear order on the ground set V. We write $\mathcal{X}^{(p)}$ for the family of all subsets of $\mathcal{X}$ of cardinality $p + 1$.

Ordered simplicial complexes provide a useful class of complexes in algebra and topology, by the following construction. Let $C = \mathbf{k}^{\mathcal{X}}$, and for each $I \in \mathcal{X}$, define

$$\chi_I(J) = \delta_{IJ}.$$

If C_p is the linear span of

$$\{\chi_I : |I| = p + 1\},$$

then the family $(C_p)_{p \in \mathbb{Z}}$ is a grading on C. If one arranges the elements of each $I \in \mathcal{X}$ into a tuple $(i_1, \ldots, i_{|I|})$ such that $i_p < i_q$ when $p < q$, and setting

$$I_p = I - \{i_p\},$$

then it can be shown that the map $\partial : C \to C$ defined by

$$\partial(\chi_I) = \sum_{p=1}^{|I|} (-1)^p \chi_{I_p}$$

is a differential on C. The pair (C, ∂) is the $\mathbf{k}$-*linear chain complex* of $\mathcal{X}$ with respect to $<$.

Example 9.1.4. Let $G = (V, E)$ be any combinatorial graph on vertex set $V = \{1, \ldots, m\}$. The set $E \cup V \cup \{\emptyset\}$ is a simplicial complex on ground set V. The matrix representation of the associated differential with respect to basis

$$\{\chi_I : I \in E \cup V \cup \{\emptyset\}\},$$

has block form

$$
\begin{array}{c|ccc|}
 & E & V & \emptyset \\
\hline
E & & & \\
V & A & & \\
\emptyset & & & \\
\hline
\end{array}
$$

where blank entries denote zero blocks, as per convention, and A is the node-incidence matrix of G.

9.2 The linear persistence module

A $\mathbb{Z}$-graded linear persistence module is a $\mathbb{Z}$-graded map of degree one, augmented by the following data.

Let $\mathbf{k}[t]$ denote the space of polynomials in variable t with coefficients in ground field $\mathbf{k}$. An *action* of $\mathbf{k}[t]$ on a vector space W is a bilinear mapping $\mu : \mathbf{k}[t] \times W \to W$ such that

$$\mu(r, \mu(s, v)) = \mu(rs, v)$$

for all $r, s \in \mathbf{k}[t]$ and all $v \in W$. One typically writes rv for $\mu(r, v)$. Every action is uniquely determined by its restriction to $\{t\} \times W \to W$, so actions are in 1-1 correspondence with linear operators on W.

A $\mathbf{k}[t]$-*module* is a vector space equipped with an action. A *graded* module of degree k is a graded vector space on which t acts by degree k map, that is, for which

$$tW_p \subseteq W_{p+k}$$

for all $p \in \mathbb{Z}$. A $\mathbf{k}$-linear *persistence module* is a graded $\mathbf{k}[t]$-module of degree one.

A *submodule* of W is a subspace $V \subseteq W$ for which $tV \subseteq V$. A submodule is *graded* if the family of intersections $V_p = V \cap W_p$ is a grading on V. If U and V are submodules, we say that the subspace $U + V$ is an (internal) direct sum of U and V if $U \cap V = 0$. A submodule is *indecomposable* if it cannot be expressed as the internal direct sum of two nontrivial submodules. It is an elementary fact that the only indecomposable $\mathbf{k}[t]$ modules are those of form $\mathbf{k}[t]\ell = \{rv : r \in \mathbf{k}[t],\ v \in \ell\}$, where ℓ is a subspace of dimension zero or one.

The *support* of a graded module W is $\mathrm{Supp}(W) = \{p : W_p \neq 0\}$. The support of every indecomposable module is an integer interval. The following structure theorem is a mainstay of modern applications in topological data analysis.

Proposition 9.2.1. *Every finite-dimensional persistence module is an internal direct sum of indecomposable submodules. If $U_1 \oplus \cdots \oplus U_m$ and $V_1 \oplus \cdots \oplus V_m$ are any two such sums, then $|\{p : \mathrm{Supp}(U_p) = I\}| = |\{q : \mathrm{Supp}(V_q) = I\}|$ for every nontrivial interval I.*

This result has a simple proof in the language of nilpotent maps on matroids. The linear operator $T : W \to W$ induced by the action of t is nilpotent, since it is positively graded and W is finite-dimensional. Recalling that

$$\mathrm{Orb}(v) = \{T^p v : p \geq 0,\ T^p v \neq 0\}$$

is the *orbit* of v under T, observe that every orbit freely generates an indecomposable submodule. Conversely, every indecomposable submodule is freely generated by an orbit.

For precision, let us say that the *projective class* of a Jordan basis $\cup_p \mathrm{Orb}(v_p)$ is the family of all bases that may be expressed in form $\cup_p \mathrm{Orb}(\alpha_p v_p)$ for some family of nonzero scalars α. Let us further say that a decomposition $W = \oplus_p U_p$ is *proper* if each U_p is indecomposable and nontrivial.

Lemma 9.2.2. *Proper decompositions are in 1-1 correspondence with the projective classes of graded Jordan bases.*

The proof of Lemma 9.2.2 an elementary exercise in definition checking. In light of this fact, Proposition 9.2.1 follows directly from Proposition 5.3.2, which we repeat here for ease of reference.

Proposition 5.3.2. *If $I_1, \ldots, I_m$ and $J_1, \ldots, J_n$ are the orbits that compose two graded Jordan bases, then there exists a bijection $\varphi : \{1, \ldots, m\} \to \{1, \ldots, n\}$ such that*

$$\mathrm{Supp}(I_p) = \mathrm{Supp}(J_{\varphi(p)})$$

for each p in $\{1, \ldots, m\}$.

9.3 Homological Persistence

To each $k \in \mathbb{Z}$ and each sequence of $\mathbf{k}$-linear complexes and complex maps

$$\cdots \longrightarrow C^{(p-1)} \longrightarrow C^{(p)} \longrightarrow C^{(p+1)} \longrightarrow \cdots \tag{9.3.1}$$

one may associate a sequence of linear maps

$$\cdots \longrightarrow H_k(C^{(p-1)}) \longrightarrow H_k(C^{(p)}) \longrightarrow H_k(C^{(p+1)}) \longrightarrow \cdots. \tag{9.3.2}$$

Sequence (9.3.2) determines an endomorphism Q on $\oplus_p H_k(C^{(p)})$, which is graded of degree 1 with respect to the canonical grading. The associated module is the *k-dimensional homological persistence module* of (9.3.1).

Homological persistence modules play a premier role in the field of topological data analysis, and the primary mode of understanding these objects is by way of the Jordan bases of Q.

Let us focus on the special case where each map $C^{(p)} \to C^{(p+1)}$ is an inclusion of complexes, and where $H_k(C^{(p)})$ vanishes for p outside some integer interval $[0, m]$. Better still, let us assume that $C^{(p)}$ vanishes for $p < 0$, and

$$Z_k(C^{(p)}) = B_k(C^{(p)}) = B_k(C^{(m)}) \tag{9.3.3}$$

for $p > m$. The family of complexes that meet these criteria includes most of those found in modern practice, either directly or by minor technical modifications.

Let $\mathcal{Z}_p$ and $\mathcal{B}_p$ denote the space of degree-k cycles and boundaries, respectively, in $C^{(p)}$. These spaces determine filtrations

$$\mathcal{Z} = (\mathcal{Z}_0, \ldots, \mathcal{Z}_m) \qquad\qquad \mathcal{B} = (\mathcal{B}_1, \ldots, \mathcal{B}_m)$$

which we may naturally regard as integer-valued functions on

$$V = Z_k(C^{(m)}).$$

The space $H_k(C^{(p)})$ is the linear quotient

$$V_p = \mathcal{Z}_p / \mathcal{B}_p = H_k(C^{(p)})$$

and Q is the induced map on $\oplus_p V_p$. Consequently Proposition 5.3.6, which we repeat here for ease of reference, carries through without modification.

Proposition 5.3.6. *The graded Jordan bases of Q are the orbits of the $\mathcal{Z}$-$\mathcal{B}$ minimal bases of V.*

This observation entails a significant reduction in complexity, as the space $\oplus_p V_p$ is in general many orders of magnitude larger (by dimension) than V. Moreover, a simple means of computing $\mathcal{Z}$-$\mathcal{B}$ minimal bases is readily available. If we let $\mathcal{F}$ denote the integer-valued function on $C^{(m)}$ whose pth sublevel set is the subspace $C^{(p)}$, then any $\mathcal{F}$-minimal Jordan basis of the differential on $C^{(m)}$ will necessarily generate $\mathcal{Z}$ and $\mathcal{B}$. Any variation on Algorithm 4 may be used to obtain such a basis.

9.4 Optimizations

The grand challenge in a preponderance of data driven homological persistence computations is combinatorial explosion in the dimension of the input. As in the Neuroscience applications of [37] alluded to in Chapter 1, complexes with many trillions of dimensions frequently derive from modest starting data. Input reduction and the avoidance of fill in sparse matrix operations therefore bear directly on efficient computation.

9.4.1 Related work

The *standard algorithm* to compute persistent homology was introduced for coefficients in the two element field by Edelsbrunner, Letscher, and Zomorodian in [26]. The adaptation of this algorithm for arbitrary field coefficients was presented by Carlsson and Zomorodian in [85]. The standard algorithm is known to have worst case cubic complexity, a bound that was shown to be sharp by Morozov in [62]. Under certain sparsity conditions, the complexity of this algorithm is less than cubic. An algorithm by Milosavljević, Morozov, and Skraba has been shown to perform the same computation in $O(n^\omega)$ time, where ω is the matrix-multiplication coefficient [59].

A number of closely related algorithms share the cubic worst-case bound while demonstrating dramatic improvements in performance empirically. These include the *twist* algorithm of Chen and Kerber [16], and the *dual* algorithm of de Silva, Morzov, and Vejdemo-Johansson [19, 20]. Some parallel algorithms for shared memory systems include the *spectral sequence* algorithm [25], and the *chunk* algorithm [6]. Algorithms for distributed computation include the *distributed algorithm* [7] and the spectral sequence algorithm of Lipsky, Skraba, and Vejdemo-Johansson [54].

The *multifield* algorithm is a sequential algorithm that computes persistent homology over multiple fields simultaneously [12]. Efficient algorithms for multidimensional persistence have been described by Lesnick and Wright in [52]. The methods of nested dissection and simplicial collapse yielded highly efficient algorithms for spaces generated from finite metric data, e.g. [47, 21]. Discrete Morse theory has yielded remarkable advances in speed and memory efficiency, for instance in algorithms of Mrozek and Batko [63], of Dłotko and Wagner [23], and of Mischaikow and Nanda [40, 60, 65].

Efficient streaming algorithms have recently been introduced by Kerber and Schriber [46]. U. Bauer has recently introduced an algorithm that takes special advantage of the natural compression of Vietoris-Rips complexes vis-a-vis the associated distance matrix, which has yielded tremendous improvements in both time and memory performance [4].

While this list is by no means comprehensive, it reflects the breadth and variety of work in this field. An number of efficient implementations are currently available, e.g. [4, 5, 8, 14, 21, 22, 30, 31, 53, 55, 61, 64, 66, 67, 71, 79].

9.4.2 Basis selection

Suppose that T is a graded map of degree one on W, and fix any graded basis (not necessarily Jordan). If the elements of this basis are arranged in ascending order by grade, then the associated matrix representation will have block form

$$\begin{pmatrix} 0 & 0 & 0 & \cdots & 0 & 0 \\ * & 0 & 0 & \cdots & 0 & 0 \\ 0 & * & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0 & \cdots & * & 0 \end{pmatrix}$$

with respect to the partition of rows and columns by grade.

Let us denote this array by A. If $T^2 = 0$, then the change effected on A by a row-first Jordan pivot on element (i, j) of block $(p+1, p)$ admits a simple description: block $(p+1, p)$ will be replaced by the array formed by a row and column clearing operation on (i, j), and in A column i and row j will be cleared.

This observation suggests that Jordan pivots may be understood, at least in part, in terms of standard row and column operations. The influence of such operations on sparsity, in turn, relates closely to that of Gauss-Jordan elimination. If M is an $m \times n$ array on which sequential row-first pivot operations are performed on elements $(1, 1), \ldots, (p, p)$, then block $(p+1, \ldots, m, p+1, \ldots, n)$ of the resulting array will be identical to that of the array produced by row-only pivot operations on the same sequence of elements.

Let us briefly recall the elements of §3.2 regarding matrix representations and matroid circuits. If $M \in \mathbf{k}^{I \times J}$ has form $[\ 1 \mid * \ ]$, where the lefthand block is an $I \times I$ identity matrix, then the *fundamental circuit* of column j with respect to basis I in the column space of M is $r \cup \{j\}$, there r is the row support of column j. In particular, the sparsity pattern of M is entirely determined by the fundamental circuits of I.

It is in general difficult to model and control the fill pattern of a sparse matrices during factorization. It has been shown, for example, that the problem of minimizing fill in Cholesky factorization of a positive definite array is NP hard [83]. Optimizations that seek to reduce fill must therefore rely on specialized structures particular to a restricted class of inputs. The class of inputs on which we focus are the boundary operators of cellular chain complexes, and the structure that we seek to leverage is the topology of the underlying topological space.

Even in this restricted setting, it is difficult to formulate provably efficient models for optimization. Input to cellular homology computations is of an highly complex nature, and including highly regular, structured, and symmetric objects from idealized mathematical models and highly irregular, noisy, low-grade data sets from all branches sciences and engineering . No sound corpus of benchmarks exists in either case, nor do there exist broadly effective random models. Efforts in each of these directions have yielded a number of important results over the past decade, e.g. [45, 68], however none may be considered broadly representative of either mathematical or scientific data. A principled approach to optimization, therefore, must rely on one or more simplifying assumptions.

The assumption we propose in the current work is a close in spirit to that which informs spectral graph theory. Recall that spectral graph embedding formalizes the notion of "smooth variation" for a real-valued function on the vertices of a graph by means of the spectral decomposition of the graph Laplacian. A function is considered to be "smooth" if it may be expressed as a linear combination of eigenfunctions with eigenvalues close to zero. This formal notion of smoothness has been observed to agree with subjective evaluations of smooth variation, and this correspondence, while imperfect, has proved sufficiently robust to support tremendous applications algebraic graph theory, both pure and applied. A number of variations on this approach have been employed in practice, using weighted or unweighted, normalized or unnormalized Laplacians, and while significant differences are observed in the performance of these variations across application domains, in the main, a consistent relationship between input and output may be seen throughout.

The model we propose seeks to maximize another subjective quantity, "normality" relative to a putative surface or submanifold. Formally this notion is no more meaningful than smoothness on a graph, since neither graphs nor a cell complexes have smooth structures, and, like smoothness, it may be formalized in a variety of different ways.

Why is normality to be desired? Suppose G_1 and G_2 are two disjoint copies of the cycle graph on vertex set $\{1, \ldots, 12\}$, whose edges are the unordered pairs

$$\{\{p, p+1\} : 1 \leq p \leq 11\} \cup \{1, 12\}.$$

Let G be the graph obtained from the union of G_1 and G_2 by adding an edge from vertex p in G_1 to vertex p in G_2, for all p. The result may be visualized as a pair of concentric circles in the plane, with 12 spokes connecting the inner to the outer. Heuristically, we regard the spoke edges to be normal to some putative underlying manifold, and the remaining edges to be tangential.

Let T_1 be a spanning tree in G composed of all the spoke edges plus every edge from the inner circle, save one. Let T_2 be a tree composed of the spoke connecting 12 to 12, plus every tangential edge except the two copies of $\{1, 12\}$. In our heuristic formulation, T_1 is relatively normal (having a large number of normal edges) and T_2 relatively tangential (having a larger number of tangential edges).

How do the fundamental circuits of these two bases compare? Those of the normal basis are relatively small. With two exceptions, each fundamental circuit has cardinality four. Among the two exceptional circuits, the larger has size 14. The circuits of the tangential basis are much larger, in general. That determined by the spoke connecting 1 to 1, for example, will have size 24, and in total the cardinalities of the fundamental circuits of T_2 sum to 178, while those of T_1 sum to just 70.

The phenomenon described by this example is observable in practice. In Figure 9.2 are displayed two spanning trees for a graph $G = (V, E)$, where V is a sample of 200 points drawn with uniform noise from the unit circle in $\mathbf{R}^2$, and $E = \{\{u, v\} \subseteq V : \|u - v\| \leq 0.5\}$. The basis on the left is obtained by Gauss-Jordan elimination on the rows of a permuted node-incidence matrix of G whose columns have been arranged in increasing order with respect to w, the function that assigns to each simplex (in this case, edge) the volume of its convex hull in the normalized spectral embedding of the underlying graph. The basis on the right was likewise obtained by Gauss-Jordan elimination on a permuted node-incidence matrix, this one with columns ordered by $-w$. It may be shown that these orders favor edges that lie normal and tangent, respectively, to the unit circle in $\mathbf{R}^2$.

Two observations are immediate upon inspection. First, each basis appears qualitatively similar. This fact reflects a loss of information incurred by the removal of all metric data relating pairs points in V, save for the adjacency information encoded by E, when one passes from the point cloud to G. Second, the sparsity structures of the associated fundamental circuits are dramatically different, with a net difference approaching a full order of magnitude in total number of edges. This phenomenon is observably robust;we find that a similar relationship between sparsity and normality holds consistently across a wide range of examples and a wide variety of formal metrics for normality.

9.4.3 Input reduction

In a majority of the regimes where persistence is currently computed, only modules of relatively small dimension (10 and below) find application. It is therefore standard to delete (or never to construct) the rows and columns associated to cells of dimension greater than $1 + p$, where p is the greatest dimension of interest. Moreover, as the basis vectors for C_{p+1} bare only indirectly on the structure of $H_p(C)$, it is generally advantageous to avoid storing them in memory.

One means to do so is to discard cells of dimension $p + 1$ after they have been matched with cells of dimension $p + 2$ in an acyclic pairing. In the special case of Vietoris-Rips complexes, such pairings may be determined before construction of the matrix representation for the differential operator has begun, thus avoiding construction of much

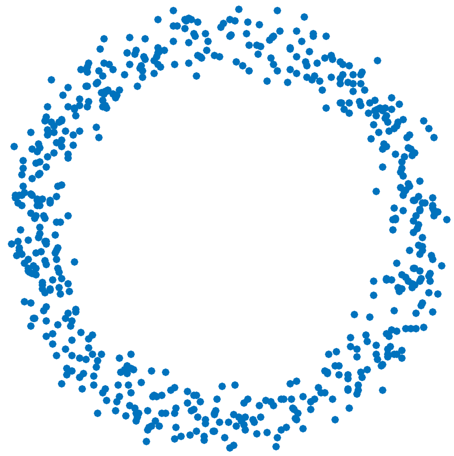


Figure 9.1: A population of 600 points sampled with noise from the unit circle in the Euclidean plane. Noise vectors were drawn from the uniform distribution on the disk of radius 0.2, centered at the origin.

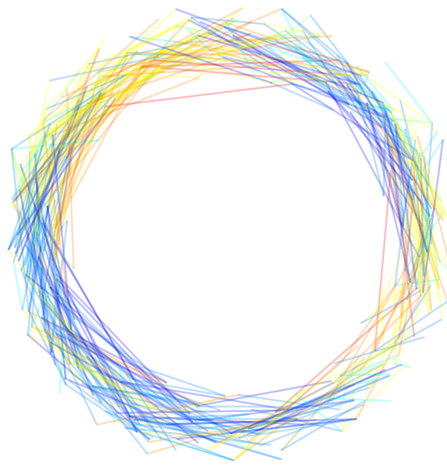
of C_{p+1} altogether. This strategy is employed in the *Eirene* library by means of linear ordering of simplices, as described in §8.4.

9.5 Benchmarks

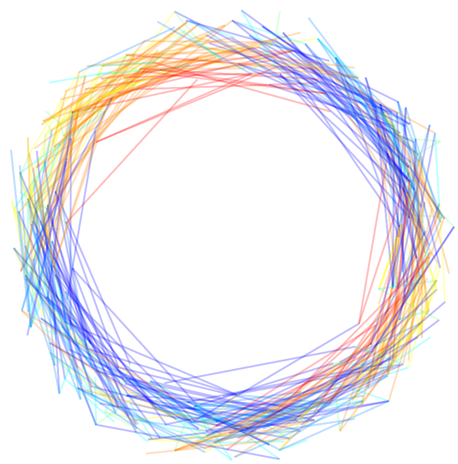
An instance of Algorithm 4 incorporating the optimizations described above has been implemented in the *Eirene* library for homological persistence [42]. Experiments were conducted on a collection of sample spaces published recently in [68].

The libraries in this publication were tested on both a cluster and a shared-memory system. The cluster is a Dell Sandybridge cluster, with 1728 (i.e. 108×16) cores of 2.0GHz Xeon SandyBridge, RAM of 64 GiB in 80 nodes and RAM of 128 GiB in 4 nodes, and a scratch disk of 20 TB. It runs the operating system (OS) Red Hat Enterprise Linux 6. The shared-memory system is an IBM System x3550 M4 server with 16 (i.e., 2×8) cores of 3.3GHz, RAM of 768 GB, and storage of 3 TB. It runs the OS Ubuntu 14.04.01.11. Results for eleg, Klein, HIV, drag 2, and random are reported for the cluster. Results for fract r are reported for the shared memory system. Detailed results for both systems may be found in the original publication.

Experiments for *Eirene* were conducted on a personal computer with Intel Core i7 processor at 2.3GHz, with 4 cores, 6MB of L3 Cache, and 16GB of RAM. Each core has 256 KB of L2 Cache. Results for time and memory performance are reported in Tables 9.1 and 9.2.

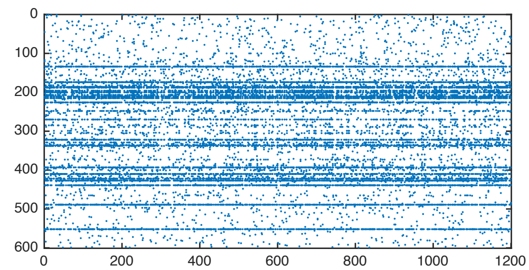


(a) A w minimal spanning tree.

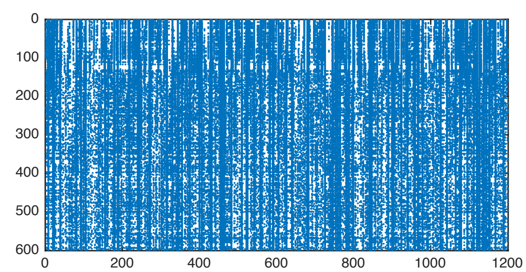


(b) A $-w$ minimal spanning tree.

Figure 9.2: Spanning trees in the one-skeleton of the Vieotris-Rips complex with scale parameter 0.5 generated by the point cloud in Figure 9.1. The cardinalities of the fundamental circuits determined by the upper basis sum to 7.3×10^5, with a median circuit length of 11 edges. The cardinalities of the fundamental circuits determined by the lower basis sum to 4.9×10^6, with a median length of 52 edges.



(a) A w minimal spanning tree.



(b) A $-w$ minimal spanning tree.

Figure 9.3: (a) Sparsity pattern of a subsample of 1200 columns selected at random from the row-reduced node incidence array determined by spanning tree (a) in Figure 9.2. The full array has a total of 7.3×10^5 nonzero entries. (b) Sparsity pattern of a subsample of 1200 columns selected at random from the row-reduced node incidence array determined by the spanning tree (b) in Figure 9.2. The full array has a total of 4.9×10^6 nonzero entries.

Data set	**eleg**	**Klein**	**HIV**	**drag 2**	**random**	**fract r**
Size of complex	4.4×10^6	1.1×10^7	2.1×10^8	1.3×10^9	3.1×10^9	2.8×10^9
Max. dim.	2	2	2	2	8	3
javaPlex (st)	84	747	-	-	-	-
Dionysus (st)	474	1830	-	-	-	-
DIPHA (st)	6	90	1631	142559	-	-
Perseus (st)	543	1978	-	-	-	-
Dionysus (d)	513	145	-	-	-	572764
DIPHA (d)	4	6	81	2358	5096	3923
Gudhi	36	89	1798	14368	-	4590
ripser	1	1	2	6	349	1517
Eirene	2	10	193	138	16	63

Table 9.1: Wall-time in seconds. Comparison results for eleg, Klein, HIV, drag 2, and random are reported for computation on the cluster. Results for fract r are reported for computation on the shared memory system.

Data set	**eleg**	**Klein**	**HIV**	**drag 2**	**random**	**fract r**†
Size of complex	4.4×10^6	1.1×10^7	2.1×10^8	1.3×10^9	3.1×10^9	2.8×10^9
Max. dim.	2	2	2	2	8	3
javaPlex (st)	< 5	< 15	> 64	> 64	> 64	> 700
Dionysus (st)	1.3	11.6	-	-	-	-
DIPHA (st)	0.1	0.2	2.7	4.9	-	-
Perseus (st)	5.1	12.7	-	-	-	-
Dionysus (d)	0.5	1.1	-	-	-	268.5
DIPHA (d)	0.1	0.2	1.8	13.8	9.6	276.1
Gudhi	0.2	0.5	8.5	62.8	-	134.8
ripser	0.007	0.02	0.06	0.2	24.7	155
Eirene	0.36	0.19	0.24	2.61	0.63	3.7

Table 9.2: Max Heap in GB. Comparison results for eleg, Klein, HIV, drag 2, and random are reported for computation on the cluster. Results for fract r are reported for computation on the shared memory system.

Chapter 10

Morse Theory

Morse theory refers to a number of theories that relate the topology of a space to the critical points of a function on that space. The seminal work in this field is attributed to Marston Morse, with applications in the study of geodesics on a Riemannian manifold: see [57, 58].

Each generation refashions Morse theory anew. Classical Morse theory specializes to smooth manifolds and smooth functions with nondegenerate Hessians: so strong a hold did this *Morse condition* exert that relaxation came but slowly. *Morse-Bott theory* permits smooth functions whose critical sets are closed submanifolds, and whose Hessians are non-degenerate in the normal direction. R. Bott used this theory in his original proof of the *Bott periodicity theorem* [13]. *Stratified Morse theory*, initiated by Goresky and MacPherson, extends this theory to the more general domain of stratified spaces [38], with phenomenal applications in algebraic geometry. Basic ties to dynamical systems came through Thom, Smale, and others, reaching their zenith in the ultimately general approach of C. Conley in his eponymous index theory [17]. Since then, Morse theory has been a key ingredient in fields as far flung as knot theory and symplectic topology [56]. E. Witten rediscovered and repackaged elements of this theory in the language of quantum field theory via deformations in the function-space attached to a manifold by Hodge theory [82].

In the late 1990s, Forman [32], following earlier work of Banchoff [2], related the ideas of smooth Morse theory to cellular complexes. The resulting Discrete Morse Theory [DMT] has been successfully applied to a wide variety of setting. Batzies and Welker have used discrete Morse theory in an algebraic setting to construct minimal resolutions of generic and shellable monomial ideals [3, 43]. Farley and Sabalka used DMT to characterize the configuration spaces of robots constrained to graphs [29]. Sköldberg has developed this theory in an abstract algebraic setting [75]. Recently, Curry, Ghrist, and Nanda have described the close relationship between discrete and algebraic Morse theories and applications on cellular sheaves [18].

This list constitutes only a fraction of the principle branches of modern Morse theory, and we will not attempt a comprehensive description of ties between the observations made in the following sections with the field as a hold. Rather we limit the scope to two branches initiated by Forman and Witten, discrete Morse theory for algebraic complexes and discrete Morse-Witten theory for cell complexes. The former has proved immensely productive in combinatorial topology and applications. It is fundamental to modern methods of homology computation, as we will see, and has been beautifully treated in a number of publications and books. The latter models after a program initiated by Smale and Witten which remains highly active in modern math and physics, for example in the study of supersymmetry.

We briefly sketch some ideas from these works before discussing the main results.

10.1 Smooth Morse Theory

This summary outlines the exposition of Bott [13] in recalling some observations of Thom, Witten, and Smale on a closed manifold M with Riemann structure g. Assume a smooth function $f : M \to \mathbf{R}$. A *critical point* of f is a point $p \in M$ for which the gradient

$$\nabla f_p = 0,$$

where ∇f is the gradient of f with respect to g. A critical point is *nondegenerate* if the Hessian

$$\mathrm{H}_p f = \left(\frac{\partial^2 f}{\partial x_i \partial x_j} \right)_{ij}$$

is nonsingular for some (equivalently) every local coordinate system (x_i). To every nondegenerate critical point we associate an *index* λ_p, the number of negative eigenvalues of $\mathrm{H}_p f$, counting multiplicity. We say f is *Morse* if it has finitely many critical points, none degenerate.

Every point q that is not a critical point of f lies on a unique 1-dimensional integral manifold X_q of ∇f which will start (limit in backwards time) at some critical point p and end (limit in forwards time) at another. This integral manifold is a *heteroclinic* or *connecting* orbit. (The use of *instanton* among physicists is an unfortunate obfuscation.)

The union of all connecting orbits having p as an initial point, which we denote W_p, is a cell of dimension λ_p. We call this the "descending cell" through p. The descending cells form a pairwise-disjoint partition M, however this partition may be quite complicated in general. It was an idea of Smale to introduce the following transversality condition to simplify their behavior. Let us denote the descending cell of $-f$ for critical point p by W'_p. We call this the *ascendening* cell of f at p, and say ∇f *transversal* if the ascending and descending cells of f meet in the most generic way possible, that is, if at any $q \in W_p \cap W'_r$ the tangent spaces $T_q W_p$ and $T_q W'_r$ span the full tangent space $T_q M$. In this case f is *Morse-Smale*.

Remark 10.1.1. By *modularity*, the transversality condition equates to the criterion that

$$\dim(T_q W_p \cap T_q W_r) = \lambda_p - \lambda_r.$$

Thus the number of connecting orbits that join points p and q is finite when $\lambda_p = \lambda_r + 1$. As remarked in 4.1.1, this is a concrete expression of modularity at the foundations of geometric topology.

We may associate a natural chain complex to a Morse-Smale function f as follows. Orient each $T_q X_q$ by $-\nabla f_x$, and fix arbitrary orientations for descending cells W_p. If we assume, for simplicity, that M is oriented, then this assignment determines orientations for the ascending cells as well. When $\lambda_p = \lambda_r + 1$, we may assign to each connecting orbit $\gamma \subseteq W_p \cap W_r'$ a value of ± 1 according to whether the exact sequence

$$0 \longrightarrow T_q X_q \longrightarrow T_q W_p \oplus T_q W_r' \longrightarrow T_q M \longrightarrow 0$$

preserves orientation. Let us denote this value $e(\gamma)$. Let

$$C^f(M) = \mathbb{Z}^{\{[W_p]\}_p}$$

be the free group over $\mathbb{Z}$ generated by the descending cells, and grade $C^f(M)$ by the dimension of W_p. We may then define a degree -1 operator ∂ by counting connecting orbits with sign:

$$\partial[W_p] = \sum_{\dim(W_q)=\dim(W_p)-1} \quad \sum_{\gamma \subseteq W_p \cap W_q'} e(\gamma)[W_q].$$

The following is a consequence of work by Smale. The complex $C^f(M)$ has been called by many names (various combinations of Thom, Smale, and Witten). This is Morse Theory; it is the Morse complex.

Theorem 10.1.2. *The operator ∂ is a differential on $C^f(M)$. With respect to this differential, $H(M, \mathbb{Z}) \cong H(C^f(M))$.*

A related construction, obtained by very different means, has been described by Witten. One considers the de Rham complex Ω^*, given by the sequence

$$\Omega^0 \longrightarrow \Omega^1 \longrightarrow \cdots \longrightarrow \Omega^n.$$

with codifferential d. The Riemannian metric determines an adjoint to d, and the spectral decomposition of the corresponding Laplacian

$$\Delta = dd^* + d^*d$$

separates Ω^q into a direct sum of finite-dimensional eigenspaces

$$\Omega^q_\lambda = \{\varphi \in \Omega^q : \Delta\varphi = \lambda\varphi\}.$$

If we denote the restriction of d to $\Omega^*_\lambda = \oplus_q \Omega^q_\lambda$, then the Hodge theory provides that

$$0 \longrightarrow \Omega^q_\lambda \xrightarrow{d_\lambda} \Omega^{q+1}_\lambda \longrightarrow \cdots \longrightarrow \Omega^n_\lambda \longrightarrow 0$$

is exact for $\lambda > 0$, and that

$$H^q(M) \cong \Omega^q_0$$

for all q. As an immediate consequence, for each $a > 0$ the finite-dimensional complex

$$\Omega^*_a = \oplus_{\lambda \leq a} \Omega^*_\lambda$$

has $H^*(M)$ as its cohomology. To this construct Witten associated a family of operators

$$d_t = e^{-tf} \circ d \circ e^{tf},$$

parametrized by $t \in \mathbf{R}$. One has a cohomology $H_t(M) = K(d_t)/I(d_t)$ since d_t squares to the zero, and it is quickly verified that $H_t^*(M) \cong H^*(M)$, since d_t is obtained by conjugation. As above, the Laplacians

$$\Delta_t = d_t d_t^* + d_t^* d_t$$

yield spectral decompositions $\Omega^*(t) = \oplus_\lambda \Omega_\lambda^*(t)$. Likewise each $a > 0$ yields a finite-dimensional complex of differential forms $\Omega_a^*(t)$ spanned by eigenforms of Δ_t with eigenvalues $\lambda \leq a$.

Witten argues that as t approaches $+\infty$ the spectrum of Δ_t separates into a finite set of values close to zero, and a complementary set of values much greater than zero. More precisely, if N_q is the set of critical values of f of index q, then there exists a t-parametrized family of maps $\psi_t : \mathrm{N}_q \to \Omega_a^q(t)$ so that for large values of t,

1. $\psi_t(\mathrm{N}_q)$ is a basis of eigenfunctions for $\Omega_a^q(t)$ and
2. $\psi_t(p)$ concentrates on a neighborhood of p, for each $p \in \mathrm{N}_q$.

As Witten claimed and Helder and Sjöstrand later confirmed [41], when f is transversal the induced codifferential on $\Omega_a^q(t)$ may be calculated to vanishingly small error as $t \to +\infty$ by counting connecting orbits, in much the same spirit as that of the Morse complex.

10.2 Discrete and Algebraic Morse Theory

The ingredients of the complex introduced by Forman [32] are conceptually close to those of the traditional Morse complex. One begins with a topological CW complex, on which is defined some suitable function f. From f is derived a collection of critical cells (critical points) and integral curves (connecting orbits). New cells are formed by combining critical cells with the integral curves that descend from them, and the resulting family is organized into the structure of a complex via incidence relations.

Let us make this description precise. Posit a regular topological CW complex X, and let $X^{(p)}$ denote the associated family of p-dimensional cells. We write $\sigma^{(p)}$ when σ has dimension p and $\tau > \sigma$ when σ lies in the boundary of τ. A function

$$f : X \to \mathbf{R}$$

is a *discrete Morse function* if

$$\#\{\tau^{(p+1)} > \sigma : f(\tau) \leq f(\sigma)\} \leq 1 \tag{10.2.1}$$

$$\#\{\upsilon^{(p-1)} > \sigma : f(\upsilon) \geq f(\sigma)\} \leq 1. \tag{10.2.2}$$

A cell $\sigma^{(p)}$ is *critical* with index p if equality fails in both cases, that is if

$$\#\{\tau^{(p+1)} > \sigma : f(\tau) \leq f(\sigma)\} = 0$$
$$\#\{\upsilon^{(p-1)} > \sigma : f(\upsilon) \geq f(\sigma)\} = 0.$$

The *gradient vector field* of f is the family

$$\nabla f = \left\{ (\sigma^{(p)}, \tau^{(p+1)}) : \sigma < \tau, \;\; f(\tau) \leq f(\sigma) \right\}.$$

It can be shown that

$$\#\{(\sigma, \tau) \in \nabla f : \sigma = \upsilon \;\text{ or }\; \tau = \upsilon\} \leq 1 \tag{10.2.3}$$

for any $\upsilon \in X$, so ∇f is a partial matching on the incidence relation $<$.

Forman defines a *discrete vector field* to be any partial matching R on $<$ for which (10.2.3) holds with R in place of ∇f. A *gradient path* on R is a sequence of form

$$\alpha_0^{(p)}, \beta_0^{(p+1)}, \alpha_1^{(p)}, \beta_1^{(p+1)}, \ldots, \beta_r^{(p+1)}, \alpha_{r+1}^{(p)} \tag{10.2.4}$$

where

$$(\alpha_i, \beta_i) \in V \qquad\qquad \alpha_{i+1} < \beta_i \qquad\qquad \textit{and} \qquad\qquad \alpha_i \neq \alpha_{i+1}$$

for all $i \in \{0, \ldots, r\}$.

Much like the Morse complex, the discrete Morse complex has a basis indexed by critical cells and a differential expressible as a sum of values determined by integral curves. One assigns an orientation to each cell of X, and defines the *multiplicity* of a gradient path γ by

$$m(\gamma) = \prod_{p=0}^{k-1} -\langle \partial \beta_p, \alpha_p \rangle \langle \partial \beta_p, \alpha_{p+1} \rangle.$$

Let C^f denote the free abelian group generated by the critical cells of X, graded by dimension, and let $\Gamma(\beta, \alpha)$ denote the set of gradient paths that run from a maximal face of β to α. The *discrete Morse boundary operator* is the degree -1 map on $\tilde{\partial} : C^f \to C^f$ defined by

$$\tilde{\partial} \beta = \sum_{\textit{critical } \alpha^{(p)}} c_{\alpha, \beta} \alpha, \tag{10.2.5}$$

where $c_{\alpha,\beta} = \sum_{\Gamma(\beta,\alpha)} m(\gamma)$.

Theorem 10.2.1 (Forman [32])**.** *The pair $(C^f, \tilde{\partial})$ is a complex, and $H(C^f) \cong H(X, \mathbb{Z})$.*

The discrete Morse complex is a highly versatile algebraic tool, with natural applications in topology, combinatorics, and algebraic geometry. The relation between smooth and discrete Morse theory has been refined by a number of results over the past two decades, e.g. [34], where it was shown that every smooth Morse complex may be realized as a discrete one, via triangulation. Given the breadth and depth of the correspondence between smooth and discrete regimes, it is reasonable to consider wether an analog to the Hodge-theoretic approach of Witten might exist for cellular spaces, as well. This is indeed the case, as shown by Forman [33].

In the spectral setting, one begins with a suitably well-behaved CW complex X, equipped with a discrete Morse function f. For each real number t one defines an operator e^{tf} on $C_*(X, \mathbf{R})$ by

$$e^{tf}\sigma = e^{tf(\sigma)}\sigma$$

for $\sigma \in X$. To the associated boundary operator one assigns a family of differentials

$$\partial_t = e^{tf}\partial e^{-tf}.$$

to which correspond a t-parametrized family of chain complexes

$$(C, \partial_t): \quad 0 \longrightarrow C_n \xrightarrow{\partial_t} C_{n-1} \xrightarrow{\partial_t} \cdots \xrightarrow{\partial_t} C_0 \longrightarrow 0$$

with associated Laplacians

$$\Delta(t) = \partial_t\partial_t^* + \partial_t^*\partial_t.$$

Here ∂_t^* denotes the linear adjoint to ∂_t with respect to the inner product on $C_*(X, \mathbf{R})$ for which basis X is orthonormal. The operator $\Delta(t)$ is symmetric, hence diagonalizable, and for each $\lambda \in \mathbf{R}$ we denote the λ-eigenspace of $\Delta(t)$ by

$$E_p^\lambda(t) = \{c \in C_p : \Delta(t)c = \lambda c\}.$$

Since $\partial_t\Delta(t) = \Delta(t)\partial_t$, operator ∂_t preserves eigenspaces. For each $\lambda \in \mathbf{R}$ one therefore has a differential complex

$$E^\lambda(t): \quad 0 \longrightarrow E_n^\lambda(t) \xrightarrow{\partial_t} E_{n-1}^\lambda(t) \xrightarrow{\partial_t} \cdots \xrightarrow{\partial_t} E_0^\lambda(t) \longrightarrow 0.$$

Forman defines a well-behaved class of Morse functions (*flat* Morse functions) and shows that for flat f

$$\Delta(t) \to \begin{pmatrix} 0 & 0 \\ 0 & D \end{pmatrix}$$

as $t \to +\infty$, where D is a block submatrix indexed by the set $\{\sigma_1, \ldots, \sigma_k\}$ of critical cells of f, and D has form

$$\operatorname{diag}(a_{\sigma_1}, \ldots, a_{\sigma_k}),$$

for some $a \in (\mathbb{Z}_{>0} \cup \{+\infty\})^{\{\sigma_1,\ldots,\sigma_k\}}$.

As $t \to \infty$, therefore, the spectrum of $\Delta(t)$ separates into a portion close to zero and a portion close to one. Fixing arbitrary $0 < \varepsilon < 1$, we thus define the *Witten complex* of f to be

$$\mathcal{W}(t) = \oplus_{\lambda<\varepsilon} E^\lambda(t)$$

writing

$$\mathcal{W}(t): \quad 0 \longrightarrow W_n(t) \xrightarrow{\partial_t} W_{n-1}(t) \xrightarrow{\partial_t} \cdots \xrightarrow{\partial_t} W_0(t) \longrightarrow 0.$$

for the natural grading induced by dimension.

Let $\pi(t)$ denote orthogonal projection from $C(X, \mathbf{R})$ to $W_p(t)$, and for each critical p-cell σ put

$$g_\sigma(t) = \pi(t)\sigma.$$

One has $g_\sigma(t) = \sigma + O(e^{-tc})$ for some positive constant c. The g_σ form a basis of $W_p(t)$ when t is sufficiently large, but not an orthonormal one in general. If we define G to be the square matrix indexed by critical p-cells such that

$$G_{\sigma_1,\sigma_2} = \langle g_{\sigma_1}, g_{\sigma_2} \rangle,$$

however, and

$$h_\sigma = G^{-1/2} g_\sigma,$$

then the h_σ do form an orthonormal basis. Forman's observation is that the matrix representation of ∂_t with respect to this basis tends to that of the discrete Morse complex.

Theorem 10.2.2 (Forman [33])**.** *Suppose that f is a flat Morse function. Then for any critical $\sigma^{(p)}$ and $\tau^{(p)}$ there exists a constant $c > 0$ such that*

$$\langle \partial_t h_\tau, h_\sigma \rangle = e^{t(f(\sigma)-f(\tau))} \left[\langle \tilde{\partial} \tau, \sigma \rangle + O(e^{-tc}) \right],$$

where $\tilde{\partial}$ is as in (10.2.5).

10.3 Results

Forman's Theorem 10.2.1 has been extended in several directions. Kozlov [48] gives a beautiful exposition of this result, illustrating that the discrete Morse complex may be realized as summand of the initial complex obtained by a sequence of splitting operations removing "atomic" subcomplexes with trivial homology. Sköldberg develops the most abstract version of which we are aware, which relies only on a direct sum decomposition of a complex of modules, together with some nondegeneracy requirements [75]. In this exposition the Morse complex is realized as the image of $\pi = \mathrm{id} - (\phi d + d\phi)$ where ϕ is a certain splitting homotopy.

These expositions share two properties in common: first, that the objects treated are differential complexes, and second, that the proofs are of an elementary but rather technical nature. We believe that these issues are related. In fact, we submit that the technical formalities found in these treatments owe not to the nature of the Morse construction, but to the restriction of a general fact about nilpotent operators to the (in this case over-structured) special case of complexes.

In evidence let us argue that Theorem 10.2.1 is not special to differentials, but holds for arbitrary 2-nilpotent operators on an object in an abelian category – even one with no biproduct structure *a priori.* In fact this was already shown in Theorem 8.2.1.

Theorem 10.3.1. *Let (C, ∂) be any chain complex and ω any graded coproduct structure on C. If*

$$f : \omega \to \boldsymbol{R}$$

is a discrete Morse function on ω, then ∇f is an acyclic Jordan pairing on $\partial(E, E)$, and the Morse complex of f is Jordan complement of ∇f in ∂.

Proof. One need only compare the discrete Morse complex (10.2.5) with the formula for the matrix representation of ∂ with respect to $\omega[[\nabla f]]$, as given by Proposition 7.2.10. $\square$

We will show in the following section that Forman's Theorem 10.2.2, moreover, requires neither that ∂ be graded nor that $\partial^2 = 0$. It is rather a property of arbitrary operators on a finite-dimensional real vector space.

Algebraic Morse theory has formed the foundation of several intensely active areas of mathematical research over the past two decades, and it is reasonable to ask why, given the high level of mathematical activity that surrounds it, the basic connection between the algebraic Morse complex and classical matrix algebra, as outlined in Theorem 10.3.1 and its proof, have gone so long unremarked. The elements of this proof, which is neither technical nor complicated, are Möbius inversion and Theorem 8.2.1. Möbius inversion began to enter maturity with a foundational paper by Rota in 1964 [74], and the essential ingredients of Theorem 8.2.1, at least in the case of complex vector spaces, entered circulation no later than 1956 [72].

The strength of the ties that link these results to combinatorial topology, moreover, has been recognized for some time. The notion of an acyclic relation is fundamental both to the general theory of Möbius inversion and to the classical formulation of algebraic Morse theory. The text of Kozlov [49] devotes an entire chapter to the Möbius transformation immediately ahead of the chapter on algebraic Morse theory – though, so far as we are aware, no direct line is drawn between the two. Even before the birth of algebraic Morse theory in its modern form, basic connections between the inversion formula and combinatorial topology were well understood, for example, vis-à-vis matroids and shellability [10].

Connections with elementary exchange have been similarly understood. This understanding is implicit in the various works that characterize the Morse complex as a direct summand. In the special case of gradient fields with cardinality one, the fundamental result of algebraic Morse theory coincides exactly with a property of clearing operations described by Chen and Kerber in [16]. Precursors to this observation can be found even among the seminal papers on persistent homology computation, e.g. [85]. Exchange, moreover, plays a pivotal role in nearly every work on computational homology, especially so in the pioneering work of Mischaikow and Nanda [60]. Adding to the body of interconnections, A. Patel has recently built on the work of Rota, Bender, Goldman, and Leinster [9, 51, 74] to extend the notion of constructible persistence to constructible maps into abelian groups, via Möbius inversion [70].

How, then, has the essential simplicity of the Morse complex (and its relation to Möbius inversion, elementary exchange) gone for so long unremarked? This question is ill posed for a several of reasons. In the first place, while we have made every effort to find a reference on this subject, there necessarily remains the possibility that one or more have been published (we would be grateful to hear of them). In the second, while several authors have offered extremely helpful historical insights, the collective editorial decisions that shaped the body of published literature as it exists today are largely unknowable.

Nevertheless, we regard it plausible that the genesis of Morse theory as a tool in homological algebra may have influenced its circulation. Morse theory assumed its first definite form in the setting smooth geometry, and entered algebra by way of combinatorial topology. The construction of the algebraic Morse complex closely mirrors that of an associated *topological* complex, which is in general more complicated and carries much more information. It would be unsurprising, given the extensive relations between the algebraic construction and its geometric and topological counterparts, that researchers in this area might infer a dependence relationship between the core results of algebraic Morse theory and the underlying geometric, topological, or homological structure.

The idea that this construction might be more general in nature came to us by way of matroid theory, and specifically matroid duality. The theory of linear matroids relies extensively on the study of Schur complements and their relation to finite families of vectors, and it was through this lens that the Möbius transformation came into view. This clarification did much to simplify the story, however a direct understanding of the Morse complex as a *Jordan complement* required the idea of a *dual exchange*, for which we find traditional language of vectors and vector space duality ill suited. It was the matroid theoretic notion of complementarity, that is *matroid duality*, that illustrated the symmetry between primal and dual exchange operations that coalesce into a single Jordan pivot. In fact, in this matter we consistently find matroid duality a more ready than the traditional venues for linear duality, e.g. tensors.

These observations tell one part of a larger story, wherein combinatorial ideas unearth new truths about algebraic, analytic, topological, and geometric objects by providing recognizable indicators of structure. We have seen this in the algebraic constructions of discrete Morse theory and will see it in the spectral/analytic constructions of Morse-Forman-Witten theory. One piece of this story which deserves independent recognition is the Schur complement. This is an extremely general *categorical* object that has been viewed, historically, as an almost exclusively as a linear construction. So much so, in fact, that so far as we are aware it has not been remarked anywhere in the literature that the Schur complement *is* in fact a complement, in the categorical sense. This despite the fact of the fundamental role played by the Schur complement in such diverse fields as probability, geometry, combinatorics, operator algebras, optimization, and matrix analysis. That entire books have been written on this subject without reference to its categorical (if not historic) foundations speaks to the breadth and depth of opportunities to expand our understanding of these application areas from a categorical perspective.

Let us now examine the algebraic foundations of Forman's Theorem 10.2.2. Posit an operator T on a finite-dimensional real vector space V with coproduct structure

$$\omega \subseteq \mathrm{Hom}(\mathbf{R}, V).$$

We write T^* for the linear adjoint to T with respect to this structure, that is, the unique map for which

$$\sigma^\sharp T^* \tau^\flat = \tau^\sharp T \sigma^\flat$$

for all $\sigma, \tau \in E$. For economy we will sometimes drop the sharp and flat operators from the elements of E, writing, for example

$$\sigma T^* T \tau$$

for $\sigma^\sharp TT^*\tau^\flat$. Context will make the intended meaning clear where this convention is invoked. A function

$$f : E \to \mathbf{R}$$

is *monotone* if $R = \mathrm{Supp}(T(E, E))$ includes into $\sim_f$, *acyclic* if

$$\mathrm{d}f = \{(\sigma, \tau) \in R : f(\sigma) = f(\tau)\}$$

is an acyclic pairing on R, and *canted* if $(\mathrm{d}f)^{(\sharp)} \cap (\mathrm{d}f)^{(\flat)} = \emptyset$. We will assume a monotone, acyclic, canted f, throughout. The elements of

$$\Gamma = \omega - \left(\mathrm{d}f^{(\flat)} \cup \mathrm{d}f^{(\sharp)}\right)$$

we call *critical cells*.

For convenience we will sometimes substitute σ and τ for $f(\sigma)$ and $f(\tau)$, for instance writing

$$\sigma \leq \tau \qquad\qquad \textit{and} \qquad\qquad e^{-|\sigma-\tau|}$$

for $f(\sigma) \leq f(\tau)$ and $e^{-|f(\sigma)-f(\tau)|}$, respectively. Given $t \in \mathbf{R}$, we write e^{tf} for the operator on V such that

$$e^{tf}\sigma = e^{tf(\sigma)}\sigma$$

for $\sigma \in E$, and define

$$T_t = e^{-tf}Te^{tf}.$$

The associated *Laplacian* is

$$\Delta_t = T_tT_t^* + T_t^*T_t.$$

As $\mathrm{Supp}\left(\lim_{t\to+\infty} T_t\right) = \mathrm{d}f$, the matrix representation of Δ_t with respect to E tends to an an array in $\mathbf{R}^{E\times E}$ supported on the diagonal of

$$\left(\mathrm{d}f^{(\flat)} \cup \mathrm{d}f^{(\sharp)}\right) \times \left(\mathrm{d}f^{(\flat)} \cup \mathrm{d}f^{(\sharp)}\right).$$

The spectrum of Δ_t therefore separates into a low sector of $|\omega| - 2|\mathrm{d}f|$ eigenvalues which, for t sufficiently large, lie in any open ball around zero, and a high sector bounded away from zero. Following the convention of Forman, we write $W(t)$ for the subspace spanned by the eigenvectors in the low sector, π_t for orthogonal projection onto $W(t)$ and

$$g_\sigma = \pi_t\sigma.$$

for each $\sigma \in \Gamma$. For t sufficiently large, $\{g_\sigma : \sigma \in \Gamma\}$ forms a basis for $W(t)$, and if for $\sigma_1, \sigma_2 \in \Gamma$

$$G_{\sigma_1,\sigma_2} = \langle g_{\sigma_1}, g_{\sigma_2}\rangle \qquad\qquad h_{\sigma_1} = G^{-1/2}g_{\sigma_1},$$

then $\{h_\sigma : \sigma \in \Gamma\}$ is an *orthonormal* basis for $W(t)$.

If g is a function $\mathbf{R} \to \mathbf{R}$ and h is a function $\mathbf{R} \to \mathbf{R}^m$, then we write

$$h \in o_e(g)$$

when there exist real constants $a < b$ such that $\|h\| \in O(e^{at})$ and $e^{bt} \in O(g)$. Similarly we write

$$h \in O_e(g)$$

when there exists a constant c so that $\|h\| \in O(e^{tc})$ and $e^{tc} \in O(g)$. If g is a vector-valued function $\mathbf{R} \to \mathbf{R}^m$, then we write $h \in o_e(g)$ if

$$h_i \in o_e(g_i) \qquad\qquad i = 1, \ldots, m,$$

where π_i is the standard projection onto the ith axis and $h_i = \pi_i h$ and $g_i = \pi_i g$.

Finally, we write k_t for the idempotent operator that projects onto the null space of

$$\times \mathrm{d} f^{(\sharp)} T_t$$

along the image of $\oplus \mathrm{d} f^{(\flat)}$, setting $k = k_0$ for economy. It is simple to check that

$$k_t = e^{-tf} k e^{tf}$$

for all t.

Theorem 10.2.2 of Forman is a special case of the following.

Theorem 10.3.2. *Under the stated conventions,*

$$\langle T_t h_\tau, h_\sigma \rangle \in e^{\tau - \sigma}(\tau T k \sigma) + o_e(e^{\tau - \sigma}).$$

for critical σ and τ.

The remainder of this section is devoted to the proof of Theorem 10.3.2. Our argument will begin and end in a fashion almost identical to that laid out by Forman in [33]. The simplifying observations are first, that the result holds for general T, not only the boundary operators of real chain complexes, and second, that the formula for the limiting array may be expressed in terms of the idempotent operator for projection onto kernel of $\mathrm{d} f^{(\sharp)} T$.

It follows from our starting hypotheses that there exists a positive constant c such that

$$\Delta_t^N = \begin{cases} O(e^{-cNt}) & \text{on} \quad W(t) \\ 1 + O(e^{-ct}) & \text{on} \quad W(t)^\perp \end{cases}$$

hence

$$1 - \Delta_t^N = \begin{cases} 1 + O(e^{-cNt}) & \text{on} \quad W(t) \\ O(e^{-ct}) & \text{on} \quad W(t)^\perp \end{cases}$$

and therefore

$$(1-\Delta_t^N)^N = \begin{cases} 1+O(e^{-cNt}) & \text{on} \quad W(t) \\ O(e^{-cNt}) & \text{on} \quad W(t)^\perp \end{cases}$$

for positive N.

Lemma 10.3.3. *One has*

$$\tau T_t^* T_t \sigma = \begin{cases} [\tau,\sigma]e^{-|\tau-\sigma|} + o_e(e^{-|\tau-\sigma|}) & \sigma \notin \mathrm{d}f^{(\flat)}, \tau \in \mathrm{d}f^{(\flat)}, \tau < \sigma, \quad \text{or} \\ & \sigma \in \mathrm{d}f^{(\flat)}, \tau \in \mathrm{d}f^{(\flat)} \\ o_e(e^{-|\tau-\sigma|}) & \sigma \notin \mathrm{d}f^{(\flat)}, \tau \in \mathrm{d}f^{(\flat)}, \tau \geq \sigma, \quad \text{or} \\ & \sigma \notin \mathrm{d}f^{(\flat)}, \tau \notin \mathrm{d}f^{(\flat)} \end{cases} \tag{10.3.1}$$

and

$$\tau T_t T_t^* \sigma = \begin{cases} [\sigma,\tau]e^{-|\tau-\sigma|} + o_e(e^{-|\tau-\sigma|}) & \sigma \notin \mathrm{d}f^{(\sharp)}, \tau \in \mathrm{d}f^{(\sharp)}, \tau > \sigma, \quad \text{or} \\ & \sigma \in \mathrm{d}f^{(\sharp)}, \tau \in \mathrm{d}f^{(\sharp)} \\ o_e(e^{-|\tau-\sigma|}) & \sigma \notin \mathrm{d}f^{(\sharp)}, \tau \in \mathrm{d}f^{(\sharp)}, \tau \leq \sigma, \quad \text{or} \\ & \sigma \notin \mathrm{d}f^{(\sharp)}, \tau \notin \mathrm{d}f^{(\sharp)}. \end{cases} \tag{10.3.2}$$

Proof of (10.3.1). In each case $\tau T_t^* T_t \sigma$ is a sum of terms of form $[\gamma,\tau][\gamma,\sigma]e^{\gamma-\tau+\gamma-\sigma}$. In the upper two cases nonzero terms satisfy $\gamma \leq \sigma$ and $\gamma \leq \tau$. Let us swap σ and τ as necessary in the second case, so that $\tau \leq \sigma$ (we may do so since $\tau T_t^* T_t \sigma = \sigma T_t^* T \tau$). Among the terms in our sum the greatest exponent can then be found on $[\tau,\sigma]e^{\tau-\sigma} = [\tau,\sigma]e^{|\tau-\sigma|}$, corresponding to $\gamma = \tau$. Thus the first two cases are justified. In the lower two cases our sum runs over γ strictly less than σ *and* τ. Since $\gamma - \tau - \gamma - \sigma < -|\tau-\sigma|$ for all such γ, the lower expression is justified. □

Proof of (10.3.2). As before, $\tau T T^* \sigma$ is a sum of terms $[\tau,\gamma][\sigma,\gamma]e^{\tau-\gamma+\sigma-\gamma}$. In each of the upper two cases, nonzero terms satisfy $\gamma \geq \sigma$ and $\gamma \geq \tau$. Let us swap σ and τ as necessary in the second case, so that $\tau \geq \sigma$ (we may do so since $\tau T_t T_t^* \sigma = \sigma T_t T^* \tau$). Among the terms in our sum $[\sigma,\tau]e^{\sigma-\tau} = [\sigma,\tau]e^{|\sigma-\tau|}$ has the greatest exponent, corresponding to $\gamma = \tau$, hence the first two cases. All nonzero terms in the lower two cases satisfy $\gamma < \sigma$ and $\gamma < \tau$, so the exponent $\tau - \gamma + \sigma - \gamma$ is strictly lower than $-|\tau - \sigma|$. □

Lemma 10.3.4. *Suppose that $\sigma \in \mathrm{d}f^{(\flat)}$ and $\sigma < \sigma_0$, and fix any τ. Then*

$$\tau T_t^* T_t \sigma \cdot e^{-|\sigma-\sigma_0|} = \begin{cases} [\tau,\sigma]e^{-|\tau-\sigma_0|} + o_e(e^{-|\tau-\sigma_0|}) & \tau \leq \sigma, \tau \in \mathrm{d}f^{(\flat)} \\ o_e(e^{-|\tau-\sigma_0|}) & \textit{else}. \end{cases} \tag{10.3.3}$$

Proof. In either case one has

$$\tau T_t^* T_t \sigma = [\tau,\sigma]e^{-|\tau-\sigma|} + o_e(e^{-|\tau-\sigma|})$$

by (10.3.1). If $\tau \leq \sigma$ then $-|\tau-\sigma| - |\sigma-\sigma_0| = -|\tau-\sigma_0|$, and if $\sigma < \tau$ then $-|\tau-\sigma| - |\sigma-\sigma_0| < -|\tau-\sigma_0|$. □

Lemma 10.3.5. *Suppose $\sigma \in \mathrm{d}f^{(\sharp)}$ and $\sigma_0 < \sigma$, and fix any τ. Then*

$$\tau T_t T_t^* \sigma \cdot e^{-|\sigma-\sigma_0|} = \begin{cases} [\tau, \sigma] e^{-|\tau-\sigma_0|} + o_e(e^{-|\tau-\sigma_0|}) & \sigma \leq \tau, \tau \in \mathrm{d}f^{(\sharp)} \\ o_e(e^{-|\tau-\sigma_0|}) & \textit{else.} \end{cases} \tag{10.3.4}$$

The proof of Lemma 10.3.5 is entirely analogous to that of Lemma 10.3.4. These observations bound scalars of form $\tau T_t^* T_t \sigma$ for $\sigma \in \mathrm{d}f^{(\flat)}$, and those of form $\tau T_t T_t^* \sigma$ for $\sigma \in \mathrm{d}f^{(\sharp)}$. The following control "cross-terms," scalars of form $\tau T_t^* T_t \sigma$ for $\sigma \in \mathrm{d}f^{(\sharp)}$, and those of form $\tau T_t T_t^* \sigma$ for $\sigma \in \mathrm{d}f^{(\flat)}$.

Lemma 10.3.6. *If $\sigma \in \mathrm{d}f^{(\flat)}$ and $\sigma < \sigma_0$, then for any τ,*

$$\tau T_t T_t^* \sigma \cdot e^{-|\sigma-\sigma_0|} = o_e(e^{-|\tau-\sigma_0|}). \tag{10.3.5}$$

Likewise, if $\sigma \in \mathrm{d}f^{(\sharp)}$ and $\sigma > \sigma_0$, then for any τ,

$$\tau T_t^* T_t \sigma \cdot e^{-|\sigma-\sigma_0|} = o_e(e^{-|\tau-\sigma_0|}). \tag{10.3.6}$$

Proof. Suppose $\sigma \in \mathrm{d}f^{(\flat)}$ and $\sigma < \sigma_0$. If $\sigma < \tau$ then $-|\sigma - \tau| - |\sigma - \sigma_0| < -|\tau - \sigma_0|$, and the desired conclusion follows. All other cases relevant to (10.3.5) are addressed by (10.3.2) directly. Now suppose $\sigma \in \mathrm{d}f^{(\sharp)}$ and $\sigma_0 < \sigma$. If $\tau < \sigma$, then $-|\sigma-\tau|-|\sigma-\sigma_0| < -|\tau-\sigma_0|$, and the desired conclusion follows. All other cases relevant to (10.3.6) are addressed by (10.3.1) directly. □

The following is a convenient repackaging of the preceding remarks. By abuse of notation, we will write ε both for the diagonal array such that

$$\varepsilon(\sigma, \sigma) = e^{-|\sigma-\sigma_0|}.$$

and for the tuple

$$\varepsilon(\sigma) = e^{-|\sigma-\sigma_0|}.$$

By a second overload, we will write $\flat$ for the linear operator on the base space of T defined

$$\flat(\sigma) = \begin{cases} \sigma^{(\flat)} & \sigma \in \mathrm{d}f^{(\sharp)} \\ 0 & \sigma \in E - \mathrm{d}f^{(\sharp)} \end{cases}$$

and for the restriction of this map to the isomorphism

$$\mathrm{span}\left(\mathrm{d}f^{(\sharp)}\right) \longrightarrow \mathrm{span}\left(\mathrm{d}f^{(\flat)}\right).$$

We define $\sharp$ similarly. Evidently $\sharp^* = \flat$, both as isomorphisms and as operators on the base space.

Lemma 10.3.7. *If α, ω are constant vectors supported on $\mathrm{d}f^{(\flat)}_{<\sigma_0}$ and $\mathrm{d}f^{(\sharp)}_{>\sigma_0}$, respectively, then for any $N \geq 0$ one has*

$$\Delta^N(\varepsilon\alpha + \varepsilon\omega) = \varepsilon(\flat T)^N \alpha + \varepsilon(\flat^* T^*)^N \omega + o_e(\varepsilon),$$

the first term having support on $\mathrm{d}f^{(\flat)}_{<\sigma_0}$*, the second on* $\mathrm{d}f^{(\sharp)}_{>\sigma_0}$.

Proof. The three preceding lemmas imply that $\flat T\varepsilon\alpha$ and $\flat^* T^* \varepsilon\omega$ have support on $\mathrm{d}f^{(\flat)}_{<\sigma_0}$ and $\mathrm{d}f^{(\sharp)}_{>\sigma_0}$, respectively, and

$$\begin{aligned} T_t^* T_t(\varepsilon\alpha) &= \varepsilon(\flat T\alpha) + o_e(\varepsilon) & \qquad T_t T_t^*(\varepsilon\omega) &= \varepsilon(\flat^* T\omega) + o_e(\varepsilon) \\ T_t^* T_t(\varepsilon\omega) &= o_e(\varepsilon) & T_t T_t^*(\varepsilon\alpha) &= o_e(\varepsilon). \end{aligned}$$

The desired conclusion follows by direct calculation. □

The α and ω that will occupy our attention are those that derive from expressions of form $\flat^* T\sigma_0$ and $\flat^* T^* \sigma_0$. Like the preceding observation, Lemma 10.3.8 is a convenient repackaging.

Lemma 10.3.8. *If* σ_0 *is critical then* $\flat T\sigma_0$ *and* $\flat^* T^* \sigma_0$ *have support on* $\mathrm{d}f^{(\flat)}_{<\sigma_0}$ *and* $\mathrm{d}f^{(\sharp)}_{>\sigma_0}$*, respectively, and*

$$T_t^* T_t \sigma_0 = \varepsilon \flat T\sigma_0 + o_e(\varepsilon) \qquad\qquad T_t T_t^* \sigma_0 = \varepsilon \flat^* T^* \sigma_0 + O(\varepsilon). \tag{10.3.7}$$

Proof. That $\flat T\sigma_0$ and $\flat^* T^* \sigma_0$ have support on $\mathrm{d}f^{(\flat)}_{<\sigma_0}$ and $\mathrm{d}f^{(\sharp)}_{>\sigma_0}$, respectively, follows directly from our hypothesis on f. The left and righthand identities in (10.3.7) are simply collations of some cases in (10.3.1) and (10.3.1), respectively. □

Lemma 10.3.9 is a direct synthesis of the two preceding remarks.

Lemma 10.3.9. *If* $\sigma_0 \in y$ *and* n *is a nonnegative integer, then*

$$\Delta_t^{N+1} \sigma_0 = \varepsilon(\flat T)^N (\flat T\sigma_0) + \varepsilon(\flat^* T^*)^N (\flat^* T^* \sigma_0) + o_e(\varepsilon), \tag{10.3.8}$$

the first term having support on $x_{<\sigma_0}$*, the second on* $x^*_{>\sigma_0}$.

We are now ready to state the main observation. In preparation, let $\lambda_\flat$ and $\lambda_\sharp$ denote the inclusion map and projection maps

$$\mathrm{span}\left(\mathrm{d}f^{(\flat)}\right) \longrightarrow V \qquad\qquad \textit{and} \qquad\qquad V \longrightarrow \mathrm{span}\left(\mathrm{d}f^{(\sharp)}\right).$$

Each $v \in V$ may be uniquely expressed as the sum of an element in the image of

$$\times(E - \mathrm{d}f^{(\flat)}). \tag{10.3.9}$$

and one from the image of $\times \mathrm{d}f^{(\flat)}$. We refer to the latter as the *flat component* of v.

Proposition 10.3.10. *For critical* σ *and* τ*,*

$$\langle T_t \pi_t \tau, \pi_t \sigma \rangle \in e^{\tau - \sigma} \tau T k \sigma + o_e(e^{\tau-\sigma}).$$

Proof. Let $\varphi_\flat$ and $\varphi_\sharp$ denote the isomorphisms

$$\varphi_\flat = \flat \lambda_\sharp T_t \lambda_\flat \qquad\qquad \textit{and} \qquad\qquad \varphi_\sharp = \flat^* \lambda_\flat^* T_t^* \lambda_\sharp^*$$

respectively. Under this convention (10.3.8) may be expressed

$$\Delta_t^N(\varepsilon\alpha + \varepsilon\omega) = \varepsilon\varphi_\flat^N\alpha + \varepsilon\varphi_\sharp^N\omega + o_e(\varepsilon).$$

If

$$u = (1 - \Delta_t^N)^N\sigma_0.$$

then $\lambda_\sharp T_t u$ does not in general vanish, however we claim that there is a "small" time-varying vector v such that $\lambda_\sharp T_t(u - v)$ does vanish, specifically,

$$v = \varphi_\flat^{-1}(\flat T_t)u.$$

Claim: $v \in o_e(\varepsilon)$.
Proof: There exist constant vectors α, ω supported on $\mathrm{d}f^{(\flat)}_{<\sigma_0}$ and $\mathrm{d}f^{(\sharp)}_{>\sigma_0}$, respectively, such that $\Delta_t u = \varepsilon\alpha + \varepsilon\omega + o_e(\varepsilon)$. Precomposition with Δ^{N-1} yields $O(e^{-cNt})$ on the lefthand side, and $\varepsilon\varphi_\flat^{N-1}\alpha + \varepsilon\varphi_\sharp^{N-1}\omega + o_e(\varepsilon)$ on the right. As $\varphi_\flat$ and $\varphi_\sharp$ are isomorphisms, it follows that α and ω vanish for sufficiently large N. When they do, $\Delta_t u \in o_e(\varepsilon)$. The flat components of $\Delta_t u$ and $(\flat T_t)u$ agree up to an error of $o_e(\varepsilon)$, so the latter is $o_e(\varepsilon)$, also. Moreover, it is simple to check that $\varphi_\flat^{-1}$ sends $O_e(\varepsilon)$ to $O_e(\varepsilon)$ and, consequently, $o_e(\varepsilon)$ to $o_e(\varepsilon)$. (The operant observation is that $\varphi_\flat^{-1}$ is triangular, and its (τ, σ) entry is proportional to $e^{-|\tau-\sigma|}$). Consequently $\varphi_\flat^{-1}(\flat T_t)u \in o_e(\varepsilon)$, which was to be shown.

Now, fix $\tau \in y$, and set $u_\tau = (1 - \Delta_t^N)^N\tau$. Since

$$\langle u_\tau, T_t u\rangle = \langle \pi\tau, T_t\pi\sigma_0\rangle + O(e^{-cNt}) \qquad\qquad \langle u_\tau, T_t v\rangle = O(e^{\tau-\sigma}),$$

our ultimate objective may be realized by establishing

$$\langle u_\tau, T_t(u + v)\rangle \in e^{\tau-\sigma_0}Tk_0\sigma_0 + o_e(e^{\tau-\sigma}). \tag{10.3.10}$$

For this we require one further observation.
Claim: $\langle u_\tau, T_t k_t\eta\rangle = e^{\tau-\eta}\tau Tk_0\eta + o_e(e^{\tau-\eta})$, for any $\eta \in E - \mathrm{d}f^{(\flat)}$.
Proof: The inner product is a sum of terms that are either proportional to or dominated by $\mu \cdot e^{\mu-\eta-|\mu-\tau|}$, with μ running over all cells. We consider the individual contribution of each term in three exhaustive cases:
Case 1: $\mu = \tau$. This term contributes $e^{\tau-\gamma}\tau Tk\eta$.
Case 2: $\mu \geq \eta$. Since $k_t = e^{tf}ke^{-tf}$, one has

$$e^{tf}Tke^{-tf} = T_t k_t.$$

As T_t tends to zero on $E - \mathrm{d}f^{(\flat)}$ and k_t tends to orthogonal projection onto the span of $E - \mathrm{d}f^{(\flat)}$, the product $T_t k_t$ tends to zero. Consequently $\mu T_t k_t\eta = e^{\mu-\eta}\mu Tk\eta$ tends to zero. When $\mu \geq \eta$, this is only possible if $\mu Tk\eta$ vanishes. The contributions of all such μ are therefore vacuous.
Case 3: $\mu < \eta$. If $\mu < \tau$ then the exponent $\mu - \eta - |\mu - \tau|$ is strictly lower than $-|\eta - \tau|$. If $\tau < \mu$ then $\mu - \eta - |\mu - \tau| = |\eta - \tau|$, and we may consider three subcases: (a) $\mu \in \mathrm{d}f^{(\sharp)}$. By definition of k, $\mu Tk\eta$ vanishes. (b) $\mu \notin \mathrm{d}f^{(\flat)} \cup \mathrm{d}f^{(\sharp)}$. The μ-component of u_τ is strictly dominated by $e^{-|\mu-\tau|}$, so the contribution is strictly dominated by $e^{-|\eta-\tau|}$. (c) $\mu \in \mathrm{d}f^{(\flat)}$.

Since by assumption $\tau < \mu$, the μ-component of u_τ is strictly dominated by $e^{-|\mu-\tau|}$, hence the contribution is strictly dominated by $e^{-|\mu-\tau|}$.

The stated claim follows.

Since $u + v = k_t u$, one has $k_t(u + v) = u + v$, and therefore

$$T_t(u + v) = T_t k_t(u + v).$$

Recalling that k_t annihilates $\mathrm{d}f^{(\flat)}$, we may express the righthand side as a sum of terms $T_t k_t \eta(u+v)$ with η running over $E - \mathrm{d}f^{(\flat)}$. The second claim asserts that, up to negligible error, the inner product of u_τ with any such term is

$$e^{\tau-\eta}[\eta(u + v)](\tau T k \eta).$$

If $\eta = \sigma_0$ then $\eta(u + v) = 1 + o_e(1)$. This term contributes $e^{\tau-\sigma_0}\tau T k\sigma_0 + o_e(e^{\tau-\sigma_0})$ to (10.3.10). If $\eta \neq \sigma$ and either $\eta \notin (\mathrm{d}f^{(\flat)} \cup \mathrm{d}f^{(\sharp)})$ or $\eta \in \mathrm{d}f^{(\sharp)}_{\leq\sigma_0}$, then $\eta(u+v) \in o_e(e^{-|\eta-\sigma_0|})$, so its contribution is $o_e(e^{\tau-\sigma_0})$. If $\eta \in \mathrm{d}f^{(\sharp)}_{>\sigma_0}$ then $e^{\tau-\sigma_0}$ strictly dominates $e^{\tau-\eta}$. As $\eta(u + v)$ is bounded, $e^{\tau-\sigma_0}$ strictly dominates this contribution as well. In summary, we have established (10.3.10), which was to be shown. □

One may rephrase Proposition 10.3.10 as the statement that

$$\langle T_t g_\tau, g_\sigma\rangle \in e^{\tau-\sigma}\tau T k\sigma + o_e(e^{\tau-\sigma})$$

for critical σ and τ. We would like the same to hold for $\langle T_t h_\tau, h_\sigma\rangle$. To check that it does, let us first bound G.

Proposition 10.3.11. *For critical σ_0 and σ_1,*

$$G_{\sigma_0,\sigma_1} = \langle g_{\sigma_0}, g_{\sigma_1}\rangle = \begin{cases} o_e(e^{-|\sigma_0-\sigma_1|}) & \sigma_0 \neq \sigma_1 \\ 1 + o_e(1) & \sigma_0 = \sigma_1. \end{cases}$$

Proof. For N sufficiently large, the projection $\pi\sigma_0 = (1 - \Delta_t^N)^N \sigma_0 + O(e^{-cNt})$ may be expressed $\sigma_0 + \varepsilon\alpha + \varepsilon\omega + o_e(\varepsilon)$ for some α, ω supported on $\mathrm{d}f^{(\flat)}_{<\sigma_0}$ and $\mathrm{d}f^{(\sharp)}_{>\sigma_0}$, respectively. A similar expression may be derived for π_{σ_1}. The desired conclusion follows. □

As an immediate consequence, one has

$$(G^{-1/2})_{\sigma_0,\sigma_1} = \begin{cases} o_e(e^{-|\sigma_0-\sigma_1|}) & \sigma_0 \neq \sigma_1 \\ 1 + o_e(1) & \sigma_0 = \sigma_1. \end{cases}$$

for critical σ_0, σ_1, whence

$$\begin{aligned} \langle T_t h_\tau, h_\sigma\rangle &= \sum_{(\sigma_1,\tau_1)\in\Gamma} (G^{-1/2})_{\tau\tau_1}\langle T_t g_{\tau_1}, g_{\sigma_1}\rangle (G^{-1/2})_{\sigma_1\sigma} \\ &= (G^{-1/2})_{\tau\tau}\langle T_t g_\tau, g_\sigma\rangle (G^{-1/2})_{\sigma\sigma} + \sum_{\Gamma-\{(\sigma,\tau)\}} (G^{-1/2})_{\tau\tau_1}\langle T_t g_{\tau_1}, g_{\sigma_1}\rangle (G^{-1/2})_{\sigma_1\sigma}. \end{aligned}$$

The first term is

$$e^{\tau-\sigma}\tau T k\sigma + o_e(e^{\tau-\sigma})$$

while the second is a sum over $(\sigma_1, \tau_1) \in \Gamma - \{(\sigma, \tau)\}$ of terms in

$$O_e(e^{-|\tau-\tau_1|})O_e(e^{\tau-\sigma})O_e(e^{-|\sigma-\sigma_1|}).$$

Each of these is $o_e(e^{\tau-\sigma})$, so

$$\langle T_t h_\tau, h_\sigma\rangle \in e^{\tau-\sigma}\tau T k\sigma + o_e(e^{\tau-\sigma})$$

as desired.

Chapter 11

Abelian Matroids

This chapter proposes a novel, category-theoretic treatment of representation theory for finite-rank matroids. The main contribution is the definition of a *(semisimple) abelian* matroid representation, which lifts the notion of a linear representation to the regime of abelian categories. We show that the main drivers of matroid representation theory – deletion, contraction, and dualization – have natural analogs for abelian representations. These abstractions are simpler more general than their linear counterparts; the relation between primal and dual representations, for example, is little more than a restatement of the (categorical) exchange lemma. Duality is broadly recognized as a primary source of depth and structure in the study of matroids, and the light cast on this construction by the categorical approach speaks to tremendous potential for interaction between the fields.

For the reader unfamiliar with the language of category theory, we invoke the conventions laid out at the opening to Chapter 6. The terms *object* and *morphism* may be replaced by $\mathbf{k}$-*linear vector space* and $\mathbf{k}$-*linear map*, respectively. The term *map* is occasionally used in place of morphism. A *monomorphism* is an injection and an *epimorphism* is a surjection. An *endomorphism* on W is a morphism $W \to W$, and an *automorphism* is an invertible endomorphism. With these substitutions in place, the phrases *in a preadditive category* and *in an abelian category* may be stricken altogether.

11.1 Linear Matroids

Let us recall the elements of linear representation theory outlined in §3.5.

Linear representations

A $\mathbf{k}$-*linear representation* of a matroid $\mathcal{M} = (E, \mathcal{I})$ is a function

$$r : E \to W$$

such that

1. W is a $\mathbf{k}$-linear vector space, and
2. $S \subseteq E$ is independent in $\mathcal{M}$ iff $r|_S$ is a linearly independent indexed family in W.

Note that independence for $r|_S$ implies independence for $r(S)$, but not vice-versa.

To every linear representation r and every $S \subseteq E$ correspond a canonical *restriction* and a non-canonical *contraction* operation (Lemma 11.1.1). We refer to $r|_S$ as the restriction of r to S, and to $q \circ (r|_{E-S})$ as the *contraction* of r by S. We do not define an operation to produce a matrix representation of $\mathcal{M}^*$, given r, either canonical or otherwise.

Lemma 11.1.1. *If $r : E \to W$ is a linear representation of $\mathcal{M}$, then*

$$r|_S \qquad \textit{and} \qquad q \circ (r|_{E-S})$$

represent $\mathcal{M}|_S$ and $\mathcal{M}/S$, respectively, where q is any morphism such that

$$\mathrm{K}(q) = \mathrm{span}(S).$$

Matrix representations

A $\mathbf{k}$-linear *matrix representation* of $\mathcal{M}$ is an array $M \in \mathbf{k}^{I \times E}$ such that

$$r_M : E \to \mathbf{k}^I$$

is a linear representation, where

$$r_M(e)(i) = M(i, e)$$

for all $i \in I$ and $e \in E$.

For every subset $S \subseteq E$ the restriction $M|_{I \times S}$ yields a canonical representation of $\mathcal{M}|_E$. There is no analogous operation to produce a canonical representation of $\mathcal{M}/S$, however there are many non-canonical operations. To illustrate, fix subsets $\alpha \subseteq I$ and $\beta \subseteq S$ such that $M(\alpha, \beta)$ is invertible and $|\alpha| = |\beta| = \rho(S)$. For convenience identify each $i \in I$ with the standard unit vector $\chi_i \in \mathbf{k}^I$, and let T be any endomorphism on $\mathbf{k}^I$ such that

$$Tr_M(\beta) = \alpha \qquad \textit{and} \qquad T(I - \alpha) = I - \alpha.$$

If $U = qT$ is the postcomposition of T with the deletion operator $q : \mathbf{k}^I \to \mathbf{k}^{I-\alpha}$, then the kerenel of U is the linear span of $r_M(S)$, hence

$$s = (U \circ r_M)|_{E-S} \tag{11.1.1}$$

is a linear representation of $\mathcal{M}/S$. The array $N \in \mathbf{k}^{(I-\alpha)\times(E-S)}$ defined

$$N(i, e) = s(e)(i)$$

is then a matrix representation $\mathcal{M}/S$. We do not define an operation to produce a matrix representation of $\mathcal{M}^*$, given M, either canonical or otherwise.

Based Representations

A *based* representation is a pair (r, B), where $B \in \mathcal{B}(M)$ and

$$r : E \to \mathbf{k}^B$$

satisfies

$$r(b) = \chi_b$$

for all $b \in B$. Based representations inherit the canonical restriction operation $r \mapsto r|_S$ and the non-canonical contraction operation $r \mapsto s$, where s is the representation defined by (11.1.1) in the special case where M satisfies $r = r_M$.

Based representations have, in addition, a canonical dual operation $(r, B) \mapsto (r^*, E-B)$, where

$$r : E \to \mathbf{k}^{E-B}$$

satisfies

$$r^*(e)(b) = r(b)(e)$$

for all $e \in (E - B)$ and all $b \in B$. That r^* is a bona fide representation for $\mathcal{M}^*$ is a nontrivial fact of representation theory.

Standard Representations

An array $M \in \mathbf{k}^{B \times (E-B)}$ is a B-*standard* matrix representation of $\mathcal{M}$ if the pair (r, B) is a based representation, where $r : E \to \mathbf{k}^B$ is the function defined by

$$r(b) = \chi_b \qquad\qquad r(e)(b) = M(b, e)$$

for all $b \in B$ and all $e \in (E - B)$.

It is worth noting that standard representations are not matrix representations, since their column indices do not run over all of E. Every standard representation uniquely determines a matrix representation, however, which may be characterized either as the unique array N such that $r = r_N$, or more concretely by

$$N = \left[\begin{array}{c|c} \delta^B & M \end{array}\right],$$

where δ^B is the Dirac delta on $B \times B$.

Lemma 11.1.2. *If* $M \in \mathbf{k}^{B \times (E-B)}$ *is a standard representation for* $\mathcal{M}$*, then*

$$M^* \in \mathbf{k}^{(E-B) \times B}$$

is a standard representation for $\mathcal{M}^*$*.*

11.2 Covariant Matroids

Let us fix an abelian category $\mathcal{C}$. Recall that an object in $\mathcal{C}$ is *simple* if it has exactly one proper subobject – namely 0. An object W is *semisimple* if it isomorphic to a coproduct

$\oplus V$, where V is an indexed family of simple objects. As a special case, the Jordan-Hölder theorem for abelian categories states that such decompositions are essentially unique up to permutation.

Theorem 11.2.1 (Jordan-Hölder)**.** *If $V = (V_i)_{i=1}^m$ and $W = (W_j)_{j \in J}$ are indexed families of simple objects in $\mathcal{C}$ and*

$$\oplus V \cong \oplus W,$$

then there exists a bijection $\varphi : \{1, \ldots, m\} \to J$ such that

$$V_p \cong W_{\varphi(p)}$$

for all $p \in \{1, \ldots, m\}$.

In light of the Jordan-Hölder theorem, we say that a an object an object W isomorphic to the coproduct of m simple objects has *finite length.* Specifically, it has *length m.*

Let Y denote the class of all simple objects in $\mathcal{C}$, and for each W define

$$\mathrm{Sim}(*, W) = \bigcup_{y \in Y} \mathrm{Hom}(y, W) \qquad \mathrm{Sim}(W, *) = \bigcup_{y \in Y} \mathrm{Hom}(W, y).$$

If W has finite length, then we may define $\mathcal{I}^\flat$ to be the class of all subfamilies $\lambda \subseteq \mathrm{Sim}(*, W)$ such that $\oplus\lambda$ is an monomorphism. Dually, $\mathcal{I}^\sharp$ is the class of all subfamilies $\upsilon \subseteq \mathrm{Sim}(W, *)$ such that $\times\upsilon$ is an epimorphism.

Proposition 11.2.2. *The pairs*

$$\Big(\mathrm{Sim}(*, W),\ \mathcal{I}^\flat\Big) \qquad \textit{and} \qquad \Big(\mathrm{Sim}(W, *),\ \mathcal{I}^\sharp\Big)$$

are matroid independence systems.

Proof. It is evident that $\mathcal{I}^\flat$ and $\mathcal{I}^\sharp$ are closed under inclusion. The Steinitz Exchange property follows from the Jordan-Hölder theorem, and from the fact that subobjects of semisimple objects are semisimple. $\square$

We call submatroids of $\big(\mathrm{Sim}(*, W),\ \mathcal{I}^\flat\big)$ *covariant*, and submatroids of matroids of $\big(\mathrm{Sim}(W, *),\ \mathcal{I}^\sharp\big)$ *contravariant.*

Abelian representations

A *covariant representation* of a matroid $\mathcal{M} = (E, \mathcal{I})$ is a function

$$r : E \to \mathrm{Sim}(*, W)$$

such that

$$\mathcal{I} = \{I \subseteq E : r|_I \in \mathcal{I}^\flat\}.$$

A *contravariant representation* is a function

$$r : E \to \mathrm{Sim}(W, *)$$

such that

$$\mathcal{I} = \{I \subseteq E : r|_I \in \mathcal{I}^\sharp\}.$$

By a slight abuse of terms, we will say that r is *semisimple* if W is a semisimple object of finite length.

To every abelian representation r and every $S \subseteq E$ corresponds a canonical *restriction* operation $r \mapsto r|_S$. While there exists no canonical *contraction*, there do exist many non-canonical contractions when r is semisimple (Lemma 11.2.3). We do not define an operation to produce a matrix representation of $\mathcal{M}^*$ from a give representation r, either canonical or otherwise.

Lemma 11.2.3. *If r is a semisimple covariant representation of $\mathcal{M}$ and e is an idempotent such that*

$$\mathrm{K}(e) = \mathrm{I}(\oplus r|_S),$$

then

$$e \circ (r|_{E-S})$$

is a semisimple covariant representation of $\mathcal{M}/S$. Dually, if r is a semisimple contravariant representation and e is an idempotent such that

$$\mathrm{I}(e) = \mathrm{K}(\times r|_S),$$

then

$$(r|_{E-S}) \circ e$$

is a semisimple contravariant representation of $\mathcal{M}/S$.

Morphic Representations

Posit a morphism

$$T : W \to W^{op}.$$

We say that a product structure λ^{op} on W^{op} is semisimple if $\lambda^{op} \subseteq \mathrm{Sim}(W, *)$, and a coproduct structure λ on W is semisimple if $\lambda \subseteq \mathrm{Sim}(*, W)$.

A *covariant morphic representation* is a pair (T, λ), where $\lambda : E \to \mathrm{Sim}(*, W)$ is a semisimple coproduct structure and

$$T\lambda = (T\lambda_e)_{e \in E}$$

is a covariant representation. Dually, a *contravariant morphic representation* is a pair

(T, λ^{op}), where λ^{op} is a semisimple product structure on W^{op} and

$$\lambda^{op} T = (\lambda^{op}_e T)_{e \in E}$$

is a contravariant representation.

The restriction and contraction operations described for linear matrix representations have natural analogs for abelian representations. Restriction for a covariant representation may be achieved by replacing λ with $\lambda|_S$ and W with $\mathrm{I}(\lambda|_S)$, for example. Contraction may be realized by first identifying a pair (α, β) such that $\beta \subseteq \lambda$ and $T(\alpha, \beta)$ is invertible and $|\alpha| = |\beta| = \rho(S)$, then substituting T with the restriction to $\mathrm{I}(\lambda|_{E-S})$ of Te, where e is projection onto the kernel of $(\times\alpha)T$ along $\mathrm{I}(\oplus\beta)$.

Standard representations

Let B be a basis in $\mathcal{M}$. A *covariant B-standard representation* of $\mathcal{M}$ is a triple (T, λ, υ), where

$$T : W \to W^{op},$$

$T\lambda : B \to \mathrm{Sim}(\,*\,, W^{op})$ and $\upsilon : (E - B) \to \mathrm{Sim}(\,*\,, W)$ are coproduct structures, and Tr is a covariant representation of $\mathcal{M}$, where

$$r|_B = \lambda \qquad\qquad r|_{E-B} = \upsilon.$$

Dually, a *contravariant B-standard representation* of $\mathcal{M}$ is a triple $(T, \lambda^{op}, \upsilon^{op})$, where $\lambda^{op}T : B \to \mathrm{Sim}(W^{op}, \,*\,)$ and $\upsilon^{op} : (E - B) \to \mathrm{Sim}(W, \,*\,)$ are product structures and $r^{op}T$ is a covariant representation of $\mathcal{M}$, where

$$r|_B = \lambda^{op} \qquad\qquad r|_{E-B} = \upsilon^{op}.$$

(Note that by a slight abuse of notation, we write $\lambda^{op}T$ and $r^{op}T$ for the functions $e \mapsto \lambda^{op}_e T$ and $e \mapsto r^{op}_e T$).

Remark 11.2.4. In the special case where $T\lambda$ and υ are the canonical coproduct structures on $\mathbf{k}^B$ and $\mathbf{k}^{E-B}$, respectively, then the rule

$$(T, \lambda, \upsilon) \mapsto T(\lambda, \upsilon)$$

determines a 1-1 correspondence between covariant B-standard representations and $\mathbf{k}$-linear B-standard matrix representations. A similar statement holds for the corresponding dual structures. Therefore Theorem 11.2.5 implies Lemma 11.1.2.

The simplicity of the proof of Theorem 11.2.5 should be contrasted with the technical arguments found in most introductory treatments.

Theorem 11.2.5. *If*

$$(T, \lambda, \upsilon)$$

is a B-standard covariant representation of $\mathcal{M}$, then

$$(T, \lambda^\sharp, \upsilon^\sharp)$$

is an $(E-B)$*-standard contravariant representation of* $\mathcal{M}^*$*. Dually, if* (T, λ, υ) *is a* B*-standard contravariant representation of* $\mathcal{M}$*, then*

$$(T, \lambda^\flat, \upsilon^\flat)$$

is an $(E-B)$ *standard covariant representation of* $\mathcal{M}^*$*.*

Proof. Let C be any basis of $\mathcal{M}$, and let

$$\alpha^{op} = B - C \qquad\qquad \alpha = C - B.$$

Submatrix $T(\alpha^{op}, \alpha)$ is invertible, so the exchange lemma provides that

$$(B-C)^\sharp T \cup ((E-B)-C)^\sharp$$

is a product structure. Thus $E-C$ is independent in the matroid represented by $(T, \lambda^\sharp, \upsilon^\sharp)$. This argument is easily reversed, and the desired conclusion follows. $\square$

Bibliography

[1] Baker, M. The Jordan Canonical Form, Matthew Baker's Math Blog, July 2015.

[2] Banchoff, T. F. Critical points and curvature for embedded polyhedral surfaces. *The American Mathematical Monthly 77*, 5 (1970), 475–485.

[3] Batzies, E., and Welker, V. Discrete Morse theory for cellular resolutions. *J. Reine Angew. Math 543* (2000), 147–168.

[4] Bauer, U. Ripser, available at: https://github.com/ripser/ripser.

[5] Bauer, U., Kerber, M., and Reininghaus, J. Dipha (a distributed persistent homology algorithm). available at https://code.google.com/p/dipha/.

[6] Bauer, U., Kerber, M., and Reininghaus, J. Clear and compress: computing persistent homology in chunks. In *Topological Methods in Data Analysis and Visualization III*, Mathematics and Visualization. 2014, pp. 103–117.

[7] Bauer, U., Kerber, M., and Reininghaus, J. Distributed computation of persistent homology. In *Proceedings of the Meeting on Algorithm Engineering & Expermiments* (Philadelphia, PA, USA, 2014), Society for Industrial and Applied Mathematics, pp. 31–38.

[8] Bauer, U., Kerber, M., Reininghaus, J., and Wagner, H. *PHAT – Persistent Homology Algorithms Toolbox.* Springer Berlin Heidelberg, Berlin, Heidelberg, 2014, pp. 137–143.

[9] Bender, E., and Goldman, J. On the applications of Möbius inversion in combinatorial analysis. *The American Mathematical Monthly 82*, 8 (1975), 789–803.

[10] Björner, A. *Matroid Applications.* Cambridge University Press, 1992, ch. 7, Homology and Shellability of Matroids and Geometric Lattices, pp. 226–283.

[11] Björner, A., Vergnas, M. L., Sturmfels, B., White, N., and Ziegler, G. M. *Oriented Matroids.* Cambridge University Press, 1993.

[12] Boissonnat, J.-D., and Maria, C. Computing persistent homology with various coefficient fields in a single pass. In *European Symposium on Algorithms* (2014), Springer, pp. 185–196.

[13] Bott, R. Lectures on Morse theory, old and new. *Bull. Amer. Math. Soc. (N.S.) 7*, 2 (09 1982), 331–358.

[14] Bubenik, P., and Dlotko, P. A persistence landscapes toolbox for topological statistics. *CoRR abs/1501.00179* (2015).

[15] Carlsson, G. Topology and data. *Bull. Amer. Math. Soc. (N.S.) 46*, 2 (2009), 255–308.

[16] Chen, C., and Kerber, M. Persistent homology computation with a twist. In *27th European Workshop on Computational Geometry (EuroCG 2011)* (2011).

[17] Conley, C. *Isolated Invariant Sets and the Morse Index*. Regional conference series in mathematics. American Mathematical Society, 1978.

[18] Curry, J., Ghrist, R., and Nanda, V. Discrete morse theory for computing cellular sheaf cohomology. *Found. Comput. Math. 16*, 4 (Aug. 2016), 875–897.

[19] De Silva, V., Morozov, D., and Vejdemo-Johansson, M. Dualities in persistent (co)homology. *Inverse Problems 27*, 12 (2011), 124003, 17.

[20] De Silva, V., Morozov, D., and Vejdemo-Johansson, M. Persistent cohomology and circular coordinates. *Discrete Comput. Geom. 45*, 4 (2011), 737–759.

[21] Dey, T. K., Shi, D., and Wang, Y. SimBa: An Efficient Tool for Approximating Rips-filtration Persistence via Simplicial Batch-collapse. *ArXiv e-prints* (Sept. 2016).

[22] Dlotko, P. Persistence landscape toolbox. available at: https://www.math.upenn.edu/dlotko/persistencelandscape.html.

[23] Dłotko, P., and Wagner, H. Computing homology and persistent homology using iterated Morse decomposition. *arXiv:1210.1429v2 [math.AT]* (2012).

[24] Edelsbrunner, H., and Harer, J. Persistent homology — a survey. In *Surveys on Discrete and Computational Geometry: Twenty Years Later.*, J. E. Goodman, J. Pach, and R. Pollack, Eds., vol. 453 of *Contemporary Mathematics*. American Mathematical Society, 2008, pp. 257–282.

[25] Edelsbrunner, H., and Harer, J. *Computational Topology: an Introduction*. American Mathematical Society, Providence, RI, 2010.

[26] Edelsbrunner, H., Letscher, D., and Zomorodian, A. Topological persistence and simplification. *Discrete and Computational Geometry 28* (2002), 511–533.

[27] Egevary, E. Ü eine konstruktive methode zer reduktion einer matrix auf jordanische normalform. *Acta Math. Acad. Sci. Hung. 10*, 31-54 (1959).

[28] Farahat, H. K. A note on the classical canonical form. *J. Lond. Math. Soc. 32* (1957), 178–180.

[29] Farley, D., and Sabalka, L. Discrete Morse theory and Graph Braid Groups. *Algebraic & Geometric Topology 5*, 3 (2005), 1075–1109.

[30] Fasy, B., Kim, J., Lecci, F., Maria, C., and Rouvreau, V. Tda: Statistical tools for topological data analysis. available at https://cran.r-project.org/web/packages/tda/index.html.

[31] Fasy, B. T., Kim, J., Lecci, F., and Maria, C. Introduction to the R package TDA. *ArXiv e-prints* (Nov. 2014).

[32] Forman, R. Morse theory for cell complexes. *Adv. Math. 134*, 1 (1998), 90–145.

[33] Forman, R. Witten-Morse theory for cell complexes. *Topology 37*, 5 (1998), 945–979.

[34] Gallais, E. Combinatorial realization of the Thom-Smale complex via discrete Morse theory. *Annali della Scuola Normale Superiore di Pisa, Classe di Scienze 9*, 2 (2010), 229–252. 20 pages.

[35] Gelfand, S. I., and Manin, Y. I. *Methods of Homological Algebra*, second ed. Springer Monographs in Mathematics. Springer-Verlag, Berlin, 2003.

[36] Ghrist, R. Barcodes: the persistent topology of data. *Bull. Amer. Math. Soc. (N.S.) 45*, 1 (2008), 61–75.

[37] Giusti, C., Pastalkova, E., Curto, C., and Itskov, V. Clique topology reveal intrinsic structure in neural connections. *Proc. Nat. Acad. Sci. 112*, 44 (2015), 13455–13460.

[38] Goresky, M., and MacPherson, R. *Stratified Morse Theory*, vol. 14 of *Ergebnisse Der Mathematik Und Ihrer Grenzgebiete.* Springer-Verlag, 1988.

[39] Guillemin, V., and Pollack, A. *Differential Topology.* AMS Chelsea Publishing, Providence, RI, 2010. Reprint of the 1974 original.

[40] Harker, S., Mischaikow, K., Mrozek, M., Nanda, V., Wagner, H., Juda, M., and Dlotko, P. The efficiency of a homology algorithm based on discrete Morse theory and coreductions. In *Proceedings of the 3rd International Workshop on Computational Topology in Image Context*, vol. 1 of *Image A*. 2010, pp. 41–47.

[41] Helffer, B., and Sjöstrand, J. Puits multiples en mécanique semi-classique vi.(cas des puits sous-variétés). In *Annales de l'IHP Physique théorique* (1987), vol. 46, pp. 353–372.

[42] Henselman, G. Eirene: a platform for computational homological algebra. `http://gregoryhenselman.org/eirene.html`, May 2016.

[43] Jöllenbeck, M., and Welker, V. *Minimal resolutions via algebraic discrete Morse theory.* American Mathematical Soc., 2009.

[44] Kaczynski, T., Mischaikow, K., and Mrozek, M. *Computational Homology*, vol. 157 of *Applied Mathematical Sciences.* Springer-Verlag, New York, 2004.

[45] Kahle, M. Topology of random clique complexes. *Discrete & Computational Geometry 45*, 3 (2011), 553–573.

[46] Kerber, M., and Schreiber, H. Barcodes of towers and a streaming algorithm for persistent homology. *CoRR abs/1701.02208* (2017).

[47] Kerber, M., Sheehy, D. R., and Skraba, P. Persistent homology and nested dissection. In *Proceedings of the Twenty-seventh Annual ACM-SIAM Symposium on Discrete Algorithms* (Philadelphia, PA, USA, 2016), SODA '16, Society for Industrial and Applied Mathematics, pp. 1234–1245.

[48] Kozlov, D. Discrete Morse theory for free chain complexes. *Comptes Rendus Mathematique 340* (2005), 867–872.

[49] Kozlov, D. *Combinatorial Algebraic Topology*, vol. 21 of *Algorithms and Computation in Mathematics*. Springer, 2008.

[50] Kurepa, S., and Veselic, K. Transformation of a matrix to normal forms. *Glasnik Mat. 2*, 22 (1967), 39–51.

[51] Leinster, T. Notions of Möbius inversion. *Bull. Belg. Math. Soc. Simon Stevin 19*, 5 (2012), 909–933.

[52] Lesnick, M., and Wright, M. Interactive Visualization of 2-D Persistence Modules. *ArXiv e-prints* (Dec. 2015).

[53] Lesnick, M., and Write, M. Rivet: The rank invariant visualization and exploration tool, 2016. available at http://rivet.online/.

[54] Lipsky, D., Skraba, P., and Vejdemo-Johansson, M. A spectral sequence for parallelized persistence. *arXiv preprint arXiv:1112.1245* (2011).

[55] Maria, C., Boissonnat, J.-D., Glisse, M., and Yvinec, M. The gudhi library: Simplicial complexes and persistent homology. In *International Congress on Mathematical Software* (2014), Springer, pp. 167–174.

[56] McDuff, D., and Salamon, D. *Introduction to Symplectic Topology*, second ed. Oxford Mathematical Monographs. Oxford University Press, 1998.

[57] Milnor, J. *Morse Theory*. Princeton University Press, 1963.

[58] Milnor, J. Differential topology forty-six years later. *Notices Amer. Math. Soc. 58*, 6 (2011), 804–809.

[59] Milosavljevic, N., Morozov, D., and Skraba, P. Zigzag persistent homology in matrix multiplication time. In *Proceedings of the 27th Annual Symposium on Computational Geometry (SCG'11)* (Paris, France, June 2011), ACM, pp. 216–225.

[60] Mischaikow, K., and Nanda, V. Morse theory for filtrations and efficient computation of persistent homology. *Discrete Comput. Geom. 50*, 2 (2013), 330–353.

[61] Morozov, D. Dionysus. available at http://www.mrzv.org/software/ dionysus/.

[62] Morozov, D. Persistence algorithm takes cubic time in the worst case. *BioGeometry News, Department of Computer Science, Duke University* (2005).

[63] Mrozek, M., and Batko, B. The coreduction homology algorithm. *Discrete and Computational Geometry 41*, 1 (2009), 96–118.

[64] Nanda, V. Perseus, the persistent homology software. available at http://www.sas.upenn.edu/ vnanda/perseus.

[65] Nanda, V. *Discrete morse theory for filtrations.* PhD thesis, Rutgers, The State University of New Jersey, 2012.

[66] o. C. S. Jyamiti research group (Prof. Tamal K. Dey), D., and Engi-neering, O. S. U. Simppers, 2014. available at http://web.cse.ohio-state.edu/ tamaldey/simppers/simppers-software/.

[67] o. C. S. Jyamiti Research Group (Prof. Tamal K. Dey), D., and Engineering, O. S. U. Gicomplex, 2013. available at http://web.cse.ohio-state.edu/ tamaldey/gic/gicsoftware/.

[68] Otter, N., Porter, M., Tillmann, U., Grindrod, P., and Harrington, H. A roadmap for the computation of persistent homology. ArXiv:1506.08903v5, Jan 2017.

[69] Oxley, J. *Matroid Theory*, 2nd ed. ed. Oxford University Press, 2011.

[70] Patel, A. Generalized persistence diagrams. *ArXiv e-prints*, 2016arXiv160103107P (2016).

[71] Perry, P., and de Silva, V. Plex, 2000–2006. available at http://mii.stanford.edu/research/comptop/programs/.

[72] Pták, V. Eine bemerkung zur jordanschen normalform von matrizen. *Acta Sci. Math. Szeged 17* (1956), 190–194.

[73] Pták, V. Norms and the spectral radius of matrices. *Czech. Math. J. 87* (1962), 555–557.

[74] Rota, G.-C. On the foundations of combinatorial theory i: Theory of möbius functions. *Z. Wahrsh. Verw. Gebiete 2* (1964), 340–368.

[75] Sköldberg, E. Morse theory from an algebraic viewpoint. *Transactions of the American Mathematical Society 358*, 1 (2006), 115–129.

[76] Smith, J. H. Schur complements and ldu decomposition in an additive category. *Linear and Multilinear Algebra 14*, 1 (1983), 11–19.

[77] Szigeti, J. Linear algebra in lattices and nilpotent endomorphisms of semisimple modules. *Journal of Algebra 319*, 1 (2008), 296 – 308.

[78] Tao, T. The Jordan Normal Form and the Euclidean Algorithm, What's New, October 2007.

[79] Tausz, A., Vejdemo-Johansson, M., and Adams, H. javaPlex: a research platform for persistent homology. In *Book of Abstracts Minisymposium on Publicly Available Geometric/Topological Software* (2012), p. 7.

[80] Truemper, K. *Matroid Decomposition.* Academic Press, San Diego, 1992.

[81] Wildon, M. A Short Proof of the Existence of Jordan Normal Form. 2007.

[82] Witten, E. Supersymmetry and morse theory. *J. Differential Geom. 17*, 4 (1982), 661–692.

[83] Yannakakis, M. Computing the minimum fill-in is np-complete. *SIAM Journal on Algebraic and Discrete Methods 2*, 1 (1981), 77–79.

[84] Zhang, F. *The Schur Complement and Its Applications*, vol. 4. Springer Science & Business Media, 2006.

[85] Zomorodian, A., and Carlsson, G. Computing persistent homology. *Discrete Comput. Geom. 33*, 2 (2005), 249–274.