Tensors: Geometry and Applications

J.M. Landsberg

Contents

Preface

Tensors are ubiquitous in the sciences. One reason for their ubiquity is that they provide a useful way to organize data. Geometry is a powerful tool for extracting information from data sets, and a beautiful subject in its own right. This book has three intended uses: as a classroom textbook, a reference work for researchers, and a research manuscript.

0.1. Usage

Classroom uses. Here are several possible courses one could give from this text:

(1) The first part of this text is suitable for an advanced course in multilinear algebra - it provides a solid foundation for the study of tensors and contains numerous applications, exercises, and examples. Such a course would cover Chapters 1,2,3 and parts of Chapters 4,5,6.

(2) For a graduate course on the geometry of tensors not assuming algebraic geometry, one can cover Chapters 1,2,4,5,6,7 and 8 skipping 2.9-12, 4.6, 5.7, 6.7 (except Pieri), 7.6 and 8.6-8.

(3) For a graduate course on the geometry of tensors assuming algebraic geometry and with more emphasis on theory, one can follow the above outline only skimming Chapters 2 and 4 (but perhaps add 2.12) and add selected later topics.

(4) I have also given a one semester class on the complexity of matrix multiplication using selected material from earlier chapters and then focusing on Chapter 11.

(5) Similarly I have used chapter 13 as a basis of a semester long class on **P** v. **NP**, assuming some algebraic geometry. Here Chapter 8 is important.

(6) I have also given several intensive short classes on various topics. A short class for statisticians, focusing on cumulants and tensor decomposition is scheduled for the near future.

Reference uses. I have compiled information on tensors in table format (e.g. regarding border rank, maximal rank, typical rank etc.) for easy reference. In particular, Chapter 3 contains most what is known on rank and border rank, stated in elementary terms. Up until now there had been no reference for even the classical results regarding tensors. (Caveat: I do not include results relying on a metric or Hermitian metric.)

Research uses. I have tried to state all the results and definitions from geometry and representation theory needed to study tensors. When proofs are not included, references for them are given. The text includes the state of the art regarding ranks and border ranks of tensors, and explains for the first time many results and problems coming from outside mathematics in geometric language. For example, a very short proof of the well-known Kruskal theorem is presented, illustrating that it hinges upon a basic geometric fact about point sets in projective space. Many other natural subvarieties of spaces of tensors are discussed in detail. Numerous open problems are presented throughout the text.

Many of the topics covered in this book are currently very active areas of research. However, there is no reasonable reference for all the wonderful and useful mathematics that is *already* known. My goal has been to fill this gap in the literature.

0.2. Overview

The book is divided into four parts: I. First applications, multilinear algebra, and overview of results, II. Geometry and Representation theory, III. More applications and IV. Advanced topics.

Chapter 1: Motivating problems. I begin with a discussion of the *complexity of matrix multiplication*, which naturally leads to a discussion of basic notions regarding tensors (bilinear maps, rank, border rank) and the central question of determining equations that describe the set of tensors of border rank at most r. The ubiquitous problem of *tensor decomposition* is illustrated with two examples: *fluorescence spectroscopy* in chemistry and *cumulants* in statistics. A brief discussion of **P** v. **NP** and its variants is presented as a prelude to Chapter 13, where the study of symmetric tensors plays an especially important role. *Tensor product states* arising in quantum information theory and *algebraic statistics* are then introduced as they are typical of applications where one studies subvarieties of spaces of tensors. I conclude by briefly mentioning how the geometry and representation theory that occupies much of the first part of the book will be useful for future research on the motivating problems.

This chapter should be accessible to anyone who is scientifically literate.

Chapter 2: Multilinear algebra. The purpose of this chapter is to introduce the language of tensors. While many researchers using tensors often think of tensors as n-dimensional $\mathbf{a}_1 \times \cdots \times \mathbf{a}_n$-tables, I emphasize coordinate-free definitions. The coordinate-free descriptions make it easier for one to take advantage of symmetries and to apply theorems. Chapter 2 includes: numerous exercises where familiar notions from linear algebra are presented in an invariant context, a discussion of rank and border rank, and first steps towards explaining how to decompose spaces of tensors. Three appendices are included. The first includes basic definitions from algebra for reference, the second reviews Jordon and rational canonical forms. The third describes *wiring diagrams*, a pictorial tool for understanding the invariance properties of tensors and as a tool for aiding calculations.

This chapter should be accessible to anyone who has had a first course in linear algebra. It may be used as the basis of a course in multi-linear algebra.

Chapter 3: Elementary results on rank and border rank. Rank and border rank are the most important properties of tensors for applications. In this chapter I report on the state of the art. When the proofs are elementary and instructional, they are included as well, otherwise they are proven later in the text. The purpose of this chapter is to provide a reference for researchers.

Chapter 4: Algebraic geometry for spaces of tensors. A central task to be accomplished in many of the motivating problems is to test if a tensor has membership in a given set (e.g., if a tensor has *rank* r). Some of these sets are defined as the zero sets of collections of polynomials, i.e., as *algebraic varieties*, while others can be expanded to be varieties by taking their *Zariski closure* (e.g., the set of tensors of border rank at most r is the Zariski closure of the set of tensors of rank at most r). I present only the essentials of projective geometry here, in order to quickly arrive at the study of groups and their modules essential to this book. Other topics in algebraic geometry are introduced as needed.

This chapter may be difficult for those unfamiliar with algebraic geometry - it is terse as numerous excellent references are available (e.g. [**154, 284**]). Its purpose is primarily to establish language. Its prerequisite is Chapter 2.

Chapter 5: Secant varieties. The notion of border rank for tensors has a vast and beautiful generalization in the context of algebraic geometry, to that of *secant varieties of projective varieties*. Many results on border rank are more easily proved in this larger geometric context, and it is easier to develop intuition regarding the border ranks of tensors when one examines properties of secant varieties in general.

The prerequisite for this chapter is Chapter 4.

Chapter 6: Exploiting symmetry: Representation theory for spaces of tensors. Representation theory provides a language for taking advantage of symmetries. Consider the space $Mat_{n\times m}$ of $n \times m$ matrices: one is usually interested in the properties of a matrix up to changes of bases (that is, the underlying properties of the linear map it encodes). This is an example of a vector space with a group acting on it. Consider polynomials on the space of matrices. The minors are the most important polynomials. Now consider the space of $n_1 \times n_2 \times \cdots \times n_k$-way arrays (i.e., a space of tensors) with $k > 2$. What are the spaces of important polynomials? Representation theory helps to give an answer.

Chapter 6 discusses representations of the group of permutations on d elements, denoted $\mathfrak{S}_d$, and of the group of invertible $n \times n$ matrices, denoted $GL_n\mathbb{C}$, and applies it to the study of homogeneous varieties. The material presented in this chapter is standard and excellent texts already exist (e.g., [**264, 132, 140**]). I focus on the aspects of representation theory useful for applications and its implementation.

The prerequisite for this chapter is Chapter 2.

Chapter 7: Tests for border rank: Equations for secant varieties. This chapter discusses the equations secant varieties in general and gives a detailed study of the equations of secant varieties of the varieties of rank one tensors and symmetric tensors, i.e., the varieties of tensors, and symmetric tensors of border rank at most r. These are the most important objects for tensor decomposition so an effort is made to present the state of the art and to give as many different perspectives as possible.

The prerequisite to Chapter 7 is Chapter 6.

Chapter 8: Additional varieties useful for spaces of tensors. In addition to secant varieties, there are general classes of varieties, such as *tangential varieties*, *dual varieties*, and the *Fano varieties of lines* that generalize certain attributes of tensors to a more general geometric situation. In the special cases of tensors, these varieties play a role in classifying normal forms and the study of rank. For example, dual varieties play a role in distinguishing the different *typical ranks* that can occur for tensors over the real numbers. They should also be useful for future applications. Chapter 8 discusses these as well as the *Chow variety* of polynomials that decompose to a product of linear factors. I also present differential-geometric tools for studying these varieties.

Chapter 8 can mostly be read immediately after Chapter 4.

Chapter 9: Rank. It is more natural in algebraic geometry to discuss border rank than rank because it relates to projective varieties. Yet, for applications sometimes one needs to determine the ranks of tensors. I first regard rank in a more general geometric context, and then specialize to the cases of interest for applications. Very little is known about the possible ranks of tensors, and what little is known is mostly in cases where there are normal forms, which is presented in Chapter 10. The main discussion in this chapter regards the ranks of symmetric tensors, because more is known about them. Included are the Comas-Seguir theorem classifying ranks of symmetric tensors in two variables as well as results on maximum possible rank. Results presented in this chapter indicate there is beautiful geometry associated to rank that is only beginning to be discovered.

Chapter 9 can be read immediately after Chapter 5.

Chapter 10: Normal forms for small tensors. The chapter describes the spaces of tensors admitting normal forms, and the normal forms of tensors in those spaces, as well as normal forms for points in small secant varieties.

The chapter can be read on a basic level after reading Chapter 2, but the proofs and geometric descriptions of the various orbit closures require material from other chapters.

The next four chapters deal with applications.

Chapter 11: The complexity of matrix multiplication. This chapter brings the reader up to date on what is known regarding the complexity of matrix multiplication, including new proofs of many standard results.

Much of the chapter needs only Chapter 2, but parts require results from Chapters 5 and 6.

Chapter 12: Tensor decomposition. In many applications one would like to express a given tensor as a sum of rank one tensors, or some class of simple tensors. In this chapter I focus on examples coming from signal processing and discuss two such: blind source separation and deconvolution of DS-CMDA signals. The blind source separation problem is similar to many questions arising in statistics, so I explain the larger context of the study of cumulants.

Often in applications one would like unique expressions for tensors as a sum of rank one tensors. I bring the reader up to date on what is known regarding when a unique expression is possible. A geometric proof of the often cited Kruskal uniqueness condition for tensors is given. The proof is short and isolates the basic geometric statement that underlies the result.

The chapter can be read after reading Chapters 2 and 3.

Chapter 13: P versus NP. This chapter includes an introduction to several algebraic versions of **P** and **NP**, as well as a discussion of Valiant's *holographic algorithms*. It concludes with a discussion of the GCT program of Mulmuley and Sohoni, which requires a knowledge of algebraic geometry and representation theory, although the rest of the chapter is elementary and only requires Chapter 2.

Chapter 14: Varieties of tensors in phylogenetics and quantum mechanics. This chapter discusses two different applications with very similar underlying mathematics. In both cases one is interested in isolating subsets (subvarieties) of spaces of tensors with certain attributes coming from physics or statistics. It turns out the resulting varieties for phylogenetics and tensor network states are strikingly similar, both in their geometry, and in the methods they were derived (via auxiliary graphs).

Much of this chapter only needs Chapter 2 as a prerequisite.

The final three chapters deal with more advanced topics.

Chapter 15: Outline of the proof of the Alexander-Hirschowitz theorem. The dimensions of the varieties of symmetric tensors of border rank at most r were determined by Alexander and Hirschowitz. A brief outline of a streamlined proof appearing in [**255**] is given here.

This chapter is intended for someone who has already had a basic course in algebraic geometry.

Chapter 16: Representation theory. This chapter includes a brief description of the rudiments of the representation theory of complex simple Lie groups and algebras. There are many excellent references for this subject so I present just enough of the theory for our purposes: the proof of Kostant's theorem that the ideals of homogeneous varieties are generated in degree two, the statement of the Bott-Borel-Weil theorem, and the presentation of the *inheritance* principle of Chapter 6 in a more general context.

This chapter is intended for someone who has already had a first course in representation theory.

Chapter 17: Weyman's method. The study of secant varieties of triple Segre products naturally leads to the Kempf-Weyman method for determining ideals and singularities of G-varieties. This chapter contains an exposition of the rudiments of the method, intended primarily to serve as an introduction to the book [**328**].

The prerequisites for this chapter include Chapter 16 and a first course in algebraic geometry.

0.3. Clash of cultures

In the course of preparing this book I have been fortunate to have had many discussions with computer scientists, applied mathematicians, engineers, physicists, and chemists. Often the beginnings of these conversations were very stressful to all involved. I have kept these difficulties in mind, attempting to write both to geometers and researchers in these various areas.

Tensor practitioners want practical results. To quote Rasmus Bro (personal communication): "*Practical* means that a user of a given chemical instrument in a hospital lab can push a button and right after get a result."

My goal is to initiate enough communication between geometers and scientists so that such practical results will be realized. While both groups are interested in such communication, there are language and even philosophical barriers to be overcome. The purpose of this paragraph is to alert geometers and scientists to some of the potential difficulties in communication.

To quote G. Folland [**123**] "For them [scientists], mathematics is the discipline of manipulating symbols according to certain sophisticated rules, and the external reality to which those symbols refer lies not in an abstract universe of sets but in the real-world phenomena that they are studying."

But mathematicians, as Folland observed, are Platonists, we think the things we are manipulating on paper have a higher existence. To quote Plato [**262**]: "*Let us take any common instance; there are beds and tables in the world –plenty of them, are there not?*

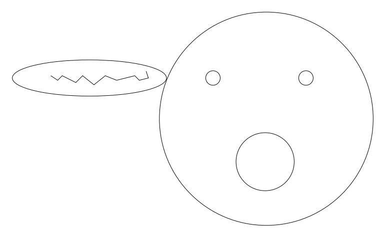
Yes. But there are only two ideas or forms of them –one the idea of a bed, the other of a table.

True. And the maker of either of them makes a bed or he makes a table for our use, in accordance with the idea –that is our way of speaking in *this and similar instances –but no artificer makes the ideas themselves: how could he?*

And what of the maker of the bed? Were you not saying that he too makes, not the idea which, according to our view, is the essence of the bed, but only a particular bed?

Yes, I did. Then if he does not make that which exists he cannot make true existence, but only some semblance of existence; and if any one were to say that the work of the maker of the bed, or of any other workman, has real existence, he could hardly be supposed to be speaking the truth. ”

This difference of cultures is particularly pronounced when discussing tensors: for some practitioners these are just multi-way arrays that one is allowed to perform certain manipulations on. For geometers these are spaces equipped with certain group actions. To emphasize the geometric aspects of tensors, geometers prefer to work invariantly: to paraphrase W. Fulton: “Don’t use coordinates unless someone holds a pickle to your head”[1]



0.4. Further reading

For gaining a basic grasp of representation theory as used in this book, one could consult [**140, 132, 264, 167**]. The styles of these books vary significantly and the reader’s taste will determine which she or he prefers. To go further with representation theory [**184**] is useful, especially for the presentation of the Weyl character formula. An excellent (and pictorial!) presentation of the implementation of the Bott-Borel-Weil theorem is in [**20**].

For basic algebraic geometry as in Chapter 4, [**154, 284**] are useful. For the more advanced commutative algebra needed in the later chapters [**116**] is written with geometry in mind. The standard and only reference for the Kempf-Weyman method is [**328**].

[1] This modification of the actual quote in tribute to my first geometry teacher, Vincent Gracchi. A problem in our 9-th grade geometry textbook asked us to determine if a 3-foot long rifle could be packed in a box of certain dimensions and Mr. Gracchi asked us all to cross out the word ‘rifle’ and substitute the word ‘pickle’ because he “did not like guns”. A big $5q + 5q$ to Mr. Gracchi for introducing his students to geometry!

The standard reference for what was known in algebraic complexity theory up to 1997 is [**53**].

0.5. Conventions, acknowledgments

0.5.1. Notations. This subsection is included for quick reference. All notations are defined properly the first time they are used in the text.

Vector spaces are usually denoted A, B, C, V, W, and A_j and the dimensions are usually the corresponding bold letters $\mathbf{a}, \mathbf{b}, \mathbf{c}$ etc... If $v_1, \dots, v_p \in V$, $\langle v_1, \dots, v_p\rangle$ denotes the span of $v_1, \dots, v_p$. If $e_1, \dots, e_{\mathbf{v}}$ is a basis of V, $e^1, \dots, e^{\mathbf{v}}$ denotes the dual basis of V^*. $GL(V)$ denotes the general linear group of invertible linear maps $V \to V$ and $\mathfrak{gl}(V)$ its Lie algebra. If G denotes a Lie or algebraic group, $\mathfrak{g}$ denotes its associated Lie algebra.

If $X \subset \mathbb{P}V$ is an algebraic set, then $\hat{X} \subset V$ is the cone over it, its inverse image plus 0 under $\pi : V\backslash 0 \to \mathbb{P}V$. If $v \in V$, $[v] \in \mathbb{P}V$ denotes $\pi(v)$. The linear span of a set $X \subset \mathbb{P}V$ is denoted $\langle X\rangle \subseteq V$.

For a variety X, X_{smooth} denotes its smooth points and X_{sing} denotes its singular points. $X_{general}$ denotes the set of general points of X. Sometimes I abuse language and refer to a point as a general point with respect to other data. For example, if $L \in G(k, V)$ and one is studying the pair (X, L) where $X \subset \mathbb{P}V$ is a subvariety, I will call L a general point if L is in general position with respect to X.

$\Lambda^k V$ denotes the k-th exterior power of the vector space V, the symbols $\wedge$ and $\bigwedge$ denote exterior product. $S^k V$ is the k-th symmetric power. The tensor product of $v, w \in V$ is denoted $v \otimes w \in V^{\otimes 2}$, and symmetric product has no marking, e.g., $vw = \frac{1}{2}(v\otimes w + w\otimes v)$. If $p \in S^d V$ is a homogeneous polynomial of degree d, write $p_{k,d-k} \in S^k V \otimes S^{d-k} V$ for its partial polarization and $\overline{p}$ for p considered as a d-multilinear form $V^* \times \cdots \times V^* \to \mathbb{C}$. When needed, $\circ$ is used for the symmetric product of spaces, e.g., given a subspace $W \subset S^q V$, $W \circ S^p V \subset S^{q+p} V$.

$\mathfrak{S}_d$ denotes the group of permutations on d elements. To a partition $\pi = (p_1, \dots, p_r)$ of d, i.e., a set of integers $p_1 \geq p_2 \geq \cdots \geq p_r$, $p_i \in \mathbb{Z}_+$, such that $p_1 + \cdots + p_r = d$, $[\pi]$ denotes the associated irreducible $\mathfrak{S}_d$ module and $S_\pi V$ denotes the associated irreducible $GL(V)$-module. I write $|\pi| = d$, and $\ell(\pi) = r$.

0.5.2. Layout. All theorems, propositions, remarks, examples, etc., are numbered together within each section; for example, Theorem 1.3.2 is the second numbered item in Section 1.3. Equations are numbered sequentially within each Chapter. I have included hints for selected exercises, those marked with the symbol $\circledcirc$ at the end, which is meant to be suggestive of a life preserver.

0.5.3. Acknowledgments. This project started as a collaboration with Jason Morton, who contributed significantly to the writing and editing of chapters 2,4,5 and 6. The book has greatly benefitted from his input. The first draft of this book arose out of a series of lectures B. Sturmfels invited me to give for his working group at UC Berkeley in spring 2006. I then gave a graduate class at Texas A&M University in fall 2007 on the complexity of matrix multiplication and a class on complexity theory in spring 2009. J. Morton and I gave lectures from the notes at a summer graduate workshop at MSRI in July 2008, as well as several lectures at a follow-up research workshop at AIM in July 2008. I also gave a GNSAGA lecture series in Florence, Italy in June 2009 on secant varieties, a short course on the geometry of tensors June 2010 in Nordfjordeid, Norway, and a one semester graduate class on the same subject at Texas A&M, Fall 2010. It is a pleasure to thank the students in these classes as well as my hosts B. Sturmfels, MSRI, AIM, G. Ottaviani and K. Ranestad. Much of the material in this book comes from joint work with L. Manivel, G. Ottaviani, and J. Weyman. It is a pleasure to thank these three collaborators for significant help at each step along the way. Other material comes from joint work with the (at the time) post-docs J. Buczyński, J. Morton, S. Norine, and Z. Teitler, who have also significantly helped with the book. I also thank my students and post-docs A. Boralevi, L. Nguyen, L. Oeding, Y. Qi, D. The, M. Yang, and K. Ye for their useful comments and questions, as well as R. Bardeli, M. Bläser, P. Bürgisser, E. Briand, J-Y Cai, P. Comon, L. De Lathauwer, M. Dillon, M. Eastwood, J. von zur Gathen, J. Grochow, C. Hammond, L-H Lim, and G. Paouris for useful comments and patiently answering my questions. It would not have been possible to write this book without such wonderful help. I was very fortunate to have careful and thoughtful anonymous referees who made numerous outstanding suggestions for improving the manuscript for which I am truly grateful. Finally I thank my family for their help, support and understanding.

Part 1

Motivation from applications, multilinear algebra and elementary results

Chapter 1

Introduction

The workhorse of scientific computation is matrix multiplication. In many applications one would like to multiply large matrices, ten thousand by ten thousand or even larger. The standard algorithm for multiplying $n \times n$ matrices uses on the order of n^3 arithmetic operations, whereas addition of matrices only uses n^2. For a $10,000 \times 10,000$ matrix this means 10^{12} arithmetic operations for multiplication compared with 10^8 for addition. Wouldn't it be great if all matrix operations were as easy as addition? As "pie in the sky" as this wish sounds, it might not be far from reality. After reviewing the standard algorithm for comparison, §1.1 begins with Strassen's algorithm for multiplying 2×2 matrices using seven multiplications. As shocking as this algorithm may be already, it has an even more stunning consequence: $n \times n$ matrices can by multiplied by performing on the order of $n^{2.81}$ arithmetic operations. This algorithm is implemented in practice for multiplication of large matrices. More recent advances have brought the number of operations needed even closer to the n^2 of addition.

To better understand Strassen's algorithm, and to investigate if it can be improved, it helps to introduce the language of *tensors*, which is done in §1.2. In particular, the *rank* and *border rank* of a tensor are the standard measures of its complexity.

The problem of determining the complexity of matrix multiplication can be rephrased as the problem of decomposing a particular tensor (the matrix multiplication operator) according to its rank. *Tensor decomposition* arises in numerous application areas: locating the area causing epileptic seizures in a brain, determining the compounds in a solution using flourescence spectroscopy, and data mining, to name a few. In each case, researchers compile data into a "multi-way array" and isolate essential features of the data by decomposing the corresponding tensor into a sum of rank one tensors. Chapter 12 discusses several examples of tensor decomposition arising in wireless communication. In §1.3, I provide two examples of tensor decomposition: *flourescence spectroscopy* in chemistry, and *blind source separation*. Blind source separation (BSS) was proposed in 1982 as a way to study how, in vertebrates, the brain detects motion from electrical signals sent by tendons (see [**95**, p. 3]). Since then numerous researchers have applied BSS in many fields, in particular engineers in signal processing. A key ingredient of BSS comes from statistics, the *cumulants* defined in §12.1 and also discussed briefly in §1.3. P. Comon utilized cumulants in [**93**], initiating *independent component analysis* (ICA), which has led to an explosion of research in signal processing.

In a letter addressed to von Neumann from 1956, see [**289**, Appendix], Gödel attempted to describe the apparent difference between intuition and systematic problem solving. His ideas, and those of others at the time, evolved to become the complexity classes **NP** (modeling intuition, or theorem proving) and **P** (modeling problems that could be solved systematically in a reasonable amount of time, like proof checking). See Chapter 14 for a brief history and explanation of these classes. Determining if these two classes are indeed distinct has been considered a central question in theoretical computer science since the 1970's. It is now also considered a central question in mathematics. I discuss several mathematical aspects of this problem in Chapter 13. In §1.4, I briefly discuss an algebraic variant due to L. Valiant: determine whether or not the *permanent* of a matrix of size m can be computed as the determinant of a matrix of size m^c for some constant c. (It is known that a determinant of size on the order of $m^2 2^m$ will work, see Chapter 13.)

A new approach to statistics via algebraic geometry has been proposed in the past 10 years. This *algebraic statistics* [**261, 257**] associates geometric objects (algebraic varieties) to sets of tables having specified statistical attributes. In a similar vein, since the 1980's, in physics, especially quantum information theory, geometric subsets of the Hilbert space describing all possible quantum states of a crystal lattice have been studied. These subsets are meant to model the states that are physically reasonable. These so-called *tensor network states* go under many names in the literature (see §1.5.2) and have surprising connections to both algebraic statistics and complexity theory. Chapter 14 discusses these two topics in detail. They are introduced briefly in §1.5.

The purpose of this chapter is to introduce the reader to the problems mentioned above and to motivate the study of the geometry of tensors that underlies them. I conclude the introduction in §1.6 by mentioning several key ideas and techniques from geometry and representation theory that are essential for the study of tensors.

1.1. The complexity of matrix multiplication

1.1.1. The standard algorithm. Let A, B be 2×2 matrices

$$A = \begin{pmatrix} a_1^1 & a_2^1 \\ a_1^2 & a_2^2 \end{pmatrix}, \quad B = \begin{pmatrix} b_1^1 & b_2^1 \\ b_1^2 & b_2^2 \end{pmatrix}.$$

The usual algorithm to calculate the matrix product $C = AB$ is

$$\begin{aligned} c_1^1 &= a_1^1 b_1^1 + a_2^1 b_1^2, \\ c_2^1 &= a_1^1 b_2^1 + a_2^1 b_2^2, \\ c_1^2 &= a_1^2 b_1^1 + a_2^2 b_1^2, \\ c_2^2 &= a_1^2 b_2^1 + a_2^2 b_2^2. \end{aligned}$$

It requires 8 multiplications and 4 additions to execute, and applied to $n \times n$ matrices, it uses n^3 multiplications and $n^3 - n^2$ additions.

1.1.2. Strassen's algorithm for multiplying 2×2 matrices using only seven scalar multiplications [296]. Set

$$\begin{aligned} I &= (a_1^1 + a_2^2)(b_1^1 + b_2^2), \\ II &= (a_1^2 + a_2^2) b_1^1, \\ III &= a_1^1 (b_2^1 - b_2^2) \\ IV &= a_2^2(-b_1^1 + b_1^2) \\ V &= (a_1^1 + a_2^1) b_2^2 \\ VI &= (-a_1^1 + a_1^2)(b_1^1 + b_2^1), \\ VII &= (a_2^1 - a_2^2)(b_1^2 + b_2^2), \end{aligned} \tag{1.1.1}$$

Exercise 1.1.2.1: Show that if $C = AB$, then

$$\begin{aligned} c_1^1 &= I + IV - V + VII, \\ c_1^2 &= II + IV, \\ c_2^1 &= III + V, \\ c_2^2 &= I + III - II + VI. \end{aligned}$$

Remark 1.1.2.2. According to P. Bürgisser (personal communication), Strassen was actually attempting to prove such an algorithm did not exist when he arrived at it by process of elimination. In fact he initially was working over $\mathbb{Z}_2$ (where a systematic study was feasible), and then realized that by carefully choosing signs the algorithm works over an arbitrary field. We will see in §5.2.2 why the algorithm could have been anticipated using elementary algebraic geometry.

Remark 1.1.2.3. In fact there is a nine parameter family of algorithms for multiplying 2×2 matrices using just seven scalar multiplications. See (2.4.5).

1.1.3. Fast multiplication of $n \times n$ matrices. In Strassen's algorithm, the entries of the matrices need not be scalars - they could themselves be matrices. Let A, B be 4×4 matrices, and write

$$A = \begin{pmatrix} a_1^1 & a_2^1 \\ a_1^2 & a_2^2 \end{pmatrix}, \quad B = \begin{pmatrix} b_1^1 & b_2^1 \\ b_1^2 & b_2^2 \end{pmatrix}.$$

where a^i_j, b^i_j are 2×2 matrices. One may apply Strassen's algorithm to get the blocks of $C = AB$ in terms of the blocks of A, B performing 7 multiplications of 2×2 matrices. Since one can apply Strassen's algorithm to each block, one can multiply 4×4 matrices using $7^2 = 49$ multiplications instead of the usual $4^3 = 64$. If A, B are $2^k \times 2^k$ matrices, one may multiply them using 7^k multiplications instead of the usual 8^k.

The total number of arithmetic operations for matrix multiplication is bounded by the number of multiplications, so counting multiplications is a reasonable measure of complexity. For more on how to measure complexity see Chapter 11.

Even if n is not a power of two, one can still save multiplications by enlarging the matrices with blocks of zeros to obtain matrices whose size is a power of two. Asymptotically, one can multiply $n \times n$ matrices using approximately $n^{\log_2(7)} \simeq n^{2.81}$ arithmetic operations. To see this, let $n = 2^k$ and write $7^k = (2^k)^a$ so $k\log_2 7 = ak\log_2 2$ so $a = \log_2 7$.

Definition 1.1.3.1. The *exponent* ω of matrix multiplication is

$$\omega := \inf\{h \in \mathbb{R} \mid Mat_{n\times n} \text{ may be multiplied using } O(n^h) \text{ arithmetic operations}\}$$

Strassen's algorithm shows $\omega \leq \log_2(7) < 2.81$. Determining ω is a central open problem in complexity theory. The current "world record" is $\omega < 2.38$, see [**100**]. I present several approaches to this problem in Chapter 11.

Strassen's algorithm is actually implemented in practice when large matrices must be multiplied. See §11.1 for a brief discussion.

1.2. Definitions from multilinear algebra

1.2.1. Linear maps. In this book I generally work over the complex numbers $\mathbb{C}$. The definitions below are thus presented for complex vector spaces, however the identical definitions hold for real vector spaces (just adjusting the ground field where necessary). Let $\mathbb{C}^n$ denote the vector space of n-tuples of complex numbers, i.e., if $v \in \mathbb{C}^n$, write the vector v as $v = (v_1, \dots, v_n)$ with $v_j \in \mathbb{C}$. The vector space structure of $\mathbb{C}^n$ means that for $v, w \in \mathbb{C}^n$ and $\alpha \in \mathbb{C}$, $v + w = (v_1 + w_1, \dots, v_n + w_n) \in \mathbb{C}^n$ and $\alpha v = (\alpha v_1, \dots, \alpha v_n) \in \mathbb{C}^n$. A map $f : \mathbb{C}^n \to \mathbb{C}^m$ is *linear* if $f(v + \alpha w) = f(v) + \alpha f(w)$ for all $v, w \in \mathbb{C}^n$ and $\alpha \in \mathbb{C}$. In this book vector spaces will generally be denoted by capital letters A, B, C, V, W, with the convention $\dim A = \mathbf{a}$, $\dim B = \mathbf{b}$ etc.. I will generally reserve the notation $\mathbb{C}^n$ for an n-dimensional vector space equipped with a basis as above. The reason for making this distinction is that the geometry of many of the phenomena we will study is more transparent if one does not make choices of bases.

If A is a vector space, let $A^* := \{f : A \to \mathbb{C} \mid f \text{ is linear}\}$ denote the *dual vector space*. If $\alpha \in A^*$ and $b \in B$, one can define a linear map $\alpha \otimes b : A \to B$ by $a \mapsto \alpha(a)b$. Such a linear map has *rank one*. The *rank* of a linear map $f : A \to B$ is the smallest r such that there exist $\alpha_1, \dots, \alpha_r \in A^*$ and $b_1, \dots, b_r \in B$ such that $f = \sum_{i=1}^r \alpha_i \otimes b_i$. (See Exercise 2.1.(4) for the equivalence of this definition with other definitions of rank.)

If $\mathbb{C}^{2*}$ and $\mathbb{C}^3$ are equipped with bases (e^1, e^2), (f_1, f_2, f_3) respectively, and $A : \mathbb{C}^2 \to \mathbb{C}^3$ is a linear map given with respect to this basis by a matrix

$$\begin{pmatrix} a^1_1 & a^1_2 \\ a^2_1 & a^2_2 \\ a^3_1 & a^3_2 \end{pmatrix},$$

then A may be written as the tensor

$$A = a^1_1 e^1 \otimes f_1 + a^1_2 e^2 \otimes f_1 + a^2_1 e^2 \otimes f_1 + a^2_2 e^2 \otimes f_2 + a^3_1 e^1 \otimes f_3 + a^3_2 e^2 \otimes f_3$$

and there exists an expression $A = \alpha^1 \otimes b_1 + \alpha^2 \otimes b_2$ because A has rank at most two.

Exercise 1.2.1.1: Find such an expression explicitly when the matrix of A is $\begin{pmatrix} 1 & 2 \\ -1 & 0 \\ 3 & 1 \end{pmatrix}$. ⊚

1.2.2. Bilinear maps. Matrix multiplication is an example of a *bilinear map*, that is, a map $f : A \times B \to C$ where A, B, C are vector spaces and for each fixed element $b \in B$, $f(\cdot, b) : A \to C$ is linear and similarly for each fixed element of A. Matrix multiplication of square matrices is a bilinear map:

$$M_{n,n,n} : \mathbb{C}^{n^2} \times \mathbb{C}^{n^2} \to \mathbb{C}^{n^2}. \tag{1.2.1}$$

If $\alpha \in A^*$, $\beta \in B^*$ and $c \in C$, the map $\alpha \otimes \beta \otimes c : A \times B \to C$ defined by $(a, b) \mapsto \alpha(a)\beta(b)c$ is a bilinear map. For any bilinear map $T : A \times B \to C$, one can represent it as a sum

$$T(a,b)=\sum_{i=1}^{r}\alpha^i(a)\beta^i(b)c_i \tag{1.2.2}$$

for some r, where $\alpha^i \in A^*$, $\beta^i \in B^*$, and $c_i \in C$.

Definition 1.2.2.1. For a bilinear map $T : A\times B \to C$, the minimal number r over all such presentations (1.2.2) is called the *rank* of T and denoted $\mathbf{R}(T)$.

This notion of rank was first defined by F. Hitchcock in 1927 [**161**].

1.2.3. Aside: Differences between linear and multilinear algebra. Basic results from linear algebra are that "rank equals row rank equals column rank", i.e., for a linear map $f : A \to B$, $\mathrm{rank}(f) = \dim f(A) = \dim f^T(B^*)$. Moreover the maximum possible rank is $\min\{\dim A, \dim B\}$ and that for "most" linear maps $A \to B$ the maximum rank occurs. Finally, if $\mathrm{rank}(f) > 1$, the expression of f as a sum of rank one linear maps is never unique, there are parameters of ways of doing so.

We will see that all these basic results fail in multi-linear algebra. Already for bilinear maps, $f : A \times B \to C$, the rank of f is generally different from $\dim f(A \times B)$, the maximum possible rank is generally greater than $\max\{\dim A, \dim B, \dim C\}$, and that "most" bilinear maps have rank *less than* the maximum possible. Finally, in many cases of interest, the expression of f as a sum of rank one linear maps *is* unique, which turns out to be crucial for applications to signal processing and medical imaging.

1.2.4. Rank and algorithms. If T has rank r it can be executed by performing r scalar multiplications (and $\mathcal{O}(r)$ additions). Thus the rank of a bilinear map gives a measure of its complexity.

In summary, $\mathbf{R}(M_{n,n,n})$ measures the number of multiplications needed to compute the product of two $n \times n$ matrices and gives a measure of its complexity. Strassen's algorithm shows $\mathbf{R}(M_{2,2,2}) \leq 7$. S. Winograd [**329**] proved $\mathbf{R}(M_{2,2,2}) = 7$. The exponent ω of matrix multiplication may be rephrased as

$$\omega = \underline{\lim}_{n\to\infty}\log_n(\mathbf{R}(M_{n,n,n}))$$

Already for 3×3 matrices, all that is known is $19 \leq \mathbf{R}(M_{3,3,3}) \leq 23$ [**29, 196**]. The best asymptotic lower bound is $\frac{5}{2}m^2 - 3m \leq \mathbf{R}(M_{m,m,m})$ [**28**]. See Chapter 11 for details.

The rank of a bilinear map and the related notions of *symmetric rank*, *border rank*, and *symmetric border rank* will be central to this book so I introduce these additional notions now.

1.2.5. Symmetric rank of a polynomial. Given a quadratic polynomial $p(x) = ax^2 + bx + c$, algorithms to write p as a sum of two squares $p = (\alpha x + \beta)^2 + (\gamma x + \delta)^2$ date back at least to ancient Babylon 5,000 years ago [**183**, Chap 1].

In this book it will be more convenient to deal with homogeneous polynomials. It is easy to convert from polynomials to homogeneous polynomials, one simply adds an extra variable, say y, and if the highest degree monomial appearing in a polynomial is d, one multiplies each monomial in the expression by the appropriate power of y so that the monomial has degree d. For example, the above polynomial becomes $p(x, y) = ax^2 + bxy + cy^2 = (\alpha x + \beta y)^2 + (\gamma x + \delta y)^2$. See §2.6.5 for more details.

A related notion to the rank of a bilinear map is the *symmetric rank* of a homogeneous polynomial. If P is homogeneous of degree d in n variables, it may always be written as a sum of d-th powers. Define the *symmetric rank* of P, $\mathbf{R}_S(P)$ to be the smallest r such that P is expressible as the sum of r d-th powers.

For example, a general homogeneous polynomial of degree 3 in two variables, $P = ax^3 + bx^2y + cxy^2 + ey^3$, where a, b, c, e are constants, will be the sum of two cubes (see Exercise 5.3.2.3).

1.2.6. Border rank and symmetric border rank. Related to the notions of rank and symmetric rank, and of equal importance for applications, are that of *border rank* and *symmetric border rank* defined below, respectively denoted $\underline{\mathbf{R}}(T)$, $\underline{\mathbf{R}}_S(P)$. Here is an informal example to illustrate symmetric border rank.

Example 1.2.6.1. While a general homogeneous polynomial of degree three in two variables is a sum of two cubes, it is *not* true that every cubic polynomial is either a cube or the sum of two cubes. For an example, consider

$$P = x^3 + 3x^2y.$$

P is not the sum of two cubes. (To see this, write $P = (sx + ty)^3 + (ux + vy)^3$ for some constants s, t, u, v, equate coefficients and show there is no solution.) However, it is the limit as $\epsilon \to 0$ of polynomials P_ϵ that are sums of two cubes, namely

$$P_\epsilon := \frac{1}{\epsilon}((\epsilon - 1)x^3 + (x + \epsilon y)^3).$$

This example dates back at least to Terracini nearly 100 years ago. Its geometry is discussed in Example 5.2.1.2.

Definition 1.2.6.2. The *symmetric border rank* of a homogeneous polynomial P, $\underline{\mathbf{R}}_S(P)$, is the smallest r such that there exists a sequence of polynomials P_ϵ, each of rank r, such that P is the limit of the P_ϵ as ϵ tends to zero. Similarly, the *border rank* $\underline{\mathbf{R}}(T)$ of a bilinear map $T : A \times B \to C$, is the smallest r such that there exists a sequence of bilinear maps T_ϵ, each of rank r, such that T is the limit of the T_ϵ as ϵ tends to zero.

Thus the polynomial P of Example 1.2.6.1 has symmetric border rank two and symmetric rank three.

The border rank bounds the rank as follows: if T can be approximated by a sequence T_ϵ where one has to divide by ϵ^q to obtain the limit (i.e., take q derivatives in the limit), then $\mathbf{R}(T) \leq q^2\underline{\mathbf{R}}(T)$, see [**53**, p. 379-380] for a more precise explanation. A similar phenomenon holds for the symmetric border rank.

By the remark above, the exponent ω of matrix multiplication may be rephrased as

$$\omega = \underline{\lim}_{n\to\infty}\log_n(\underline{\mathbf{R}}(M_{n,n,n})).$$

In [**198**] it was shown that $\underline{\mathbf{R}}(M_{2,2,2}) = 7$. For 3×3 matrices, all that is known is $14 \leq \underline{\mathbf{R}}(M_{3,3,3}) \leq 21$, and for $n \times n$ matrices, $\frac{3n^2}{2} + \frac{n}{2} - 1 \leq \underline{\mathbf{R}}(M_{n,n,n})$ [**217**]. See Chapter 11 for details.

The advantage of border rank over rank is that the set of bilinear maps of border rank at most r can be described as the zero set of a collection of polynomials, as discussed in the next paragraph.

1.2.7. Our first spaces of tensors and varieties inside them. Let $A^*{\otimes}B$ denote the vector space of linear maps $A \to B$. The set of linear maps of rank at most r will be denoted $\hat{\sigma}_r = \hat{\sigma}_{r,A^*\otimes B}$. This set is the zero set of a collection of homogeneous polynomials on the vector space $A^*{\otimes}B$. Explicitly, if we choose bases and identify $A^*{\otimes}B$ with the space of $\mathbf{a} \times \mathbf{b}$ matrices, $\hat{\sigma}_r$ is the set of matrices whose $(r+1)\times(r+1)$ minors are all zero. In particular there is a simple test to see if a linear map has rank at most r.

A subset of a vector space defined as the zero set of a collection of homogeneous polynomials is called an *algebraic variety*.

Now let $A^*{\otimes}B^*{\otimes}C$ denote the vector space of bilinear maps $A\times B \to C$. This is our first example of a space of *tensors*, defined in Chapter 2, beyond the familiar space of linear maps. Expressed with respect to bases, a bilinear map is a “three dimensional matrix” or array.

The set of bilinear maps of rank at most r is not an algebraic variety, i.e., it is not the zero set of a collection of polynomials. However the set of bilinear maps of border rank at most r is and algebraic variety. The set of bilinear maps $f : A \times B \to C$ of border rank at most r will be denoted $\hat{\sigma}_r = \hat{\sigma}_{r,A^*\otimes B^*\otimes C}$. It is the zero set of a collection of homogeneous polynomials on the vector space $A^*{\otimes}B^*{\otimes}C$.

In principle, to test for membership of a bilinear map T in $\hat{\sigma}_r$, one could simply evaluate the defining equations of $\hat{\sigma}_r$ on T and see if they are zero. However, unlike the case of linear maps, defining equations for $\hat{\sigma}_{r,A^*\otimes B^*\otimes C}$ are *not* known in general. Chapter 7 discusses what is known about them.

More generally, if $A_1, \dots, A_n$ are vector spaces, one can consider the *multi-linear maps* $A_1 \times \cdots \times A_n \to \mathbb{C}$, the set of all such forms a vector space, which is a space of tensors and is denoted $A_1^* \otimes \cdots \otimes A_n^*$. The rank of an element of $A_1^* \otimes \cdots \otimes A_n^*$ is defined similarly to the definitions above, see §2.4.1. In bases one obtains the set of $\mathbf{a}_1 \times \cdots \times \mathbf{a}_n$-way arrays. Adopting a coordinate-free perspective on these arrays will make it easier to isolate the properties of tensors of importance.

1.3. Tensor decomposition

One of the most ubiquitous applications of tensors is to *tensor decomposition*. In this book I will use the term *CP decomposition* for the decomposition of a tensor into a sum of rank one tensors. Other terminology used (and an explantion of the origin of the term "CP") is given in the introduction to Chapter 13.

We already saw an example of this problem in §1.1, where we wanted to write a bilinear map $f : A \times B \to C$ as a sum of maps of the form $(v, w) \mapsto \alpha(v)\beta(w)c$ where $\alpha \in A^*$, $\beta \in B^*$, $c \in C$.

In applications, one collects data arranged in a multi-dimensional array T. Usually the science indicates that this array, considered as a tensor, should be "close" to a tensor of small rank, say r. The problems researchers face are to i) find the correct value of r and/or ii) to find a tensor of rank r that is "close" to T. §1.3.1 is an example arising in chemistry, taken from the book [**1**].

A central problem in signal processing is *source separation*. An often used metaphor to explain the problem is the "cocktail party problem" where a collection of people in a room is speaking. Several receivers are set up, recording the conversations. What they actually record is not the individual conversations, but all of them mixed together and the goal is to recover what each individual is saying. Remarkably, this "unmixing" can often be accomplished using a CP decomposition. This problem relates to larger questions in statistics and is discussed in detail in Chapter 12, and briefly in §1.3.2.

1.3.1. Tensor decomposition and fluorescence spectroscopy.

I samples of solutions are analyzed, each contains different chemicals diluted at different concentrations. The first goal is to determine the number r of different chemicals present. In [**1**, §10.2] there are four such, *dihydroxybenzene,*

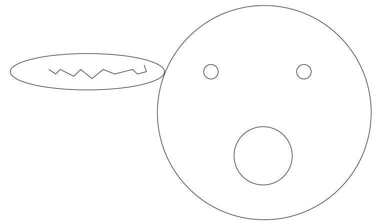


Figure 1.3.1. temporary picture until get party

tryptophan, phenylalanine and *DOPA*. Each sample, say sample number i, is successively excited by light at J different wavelengths. For every excitation wavelength one measures the emitted spectrum. Say the intensity of the fluorescent light emitted is measured at K different wavelengths. Hence for every i, one obtains a $J \times K$ table of excitation-emission matrices.

Thus the data one is handed is an $I \times J \times K$ array. In bases, if $\{e_i\}$ is a basis of $\mathbb{C}^I$, $\{h_j\}$ a basis of $\mathbb{C}^J$, and $\{g_k\}$ a basis of $\mathbb{C}^K$, then $T = \sum_{ijk} T_{ijk} e_i \otimes h_j \otimes g_k$. A first goal is to determine r such that

$$T \simeq \sum_{f=1}^{r} a_f \otimes b_f \otimes c_f$$

where each f represents a substance. Writing $a_f = \sum a_{i,f} e_i$, then $a_{i,f}$ is the concentration of the f-th substance in the i-th sample, and similarly using the given bases of $\mathbb{R}^J$ and $\mathbb{R}^K$, $c_{k,f}$ is the fraction of photons the f-th substance emits at wavelength k, and $b_{j,f}$ is the intensity of the incident light at excitation wavelength j multiplied by the absorption at wavelength j.

There will be noise in the data, so T will actually be of generic rank, but there will be a very low rank tensor $\tilde{T}$ that closely approximates it. (For all complex spaces of tensors, there is a rank that occurs with probability one which is called the *generic rank*, see Definition 3.1.4.2.) There is no metric naturally associated to the data, so the meaning of "approximation" is not clear. In [**1**], one proceeds as follows to find r. First of all, r is assumed to be very small (at most 7 in their exposition). Then for each r_0, $1 \leq r_0 \leq 7$, one assumes $r_0 = r$ and applies a numerical algorithm that attempts to find the r_0 *components* (i.e. rank one tensors) that $\tilde{T}$ would be the sum of. The values of r_0 for which the algorithm does not converge quickly are thrown out. (The authors remark that this procedure is not mathematically justified, but seems to work well in practice. In the example, these discarded values of r_0 are too large.) Then, for the remaining values of r_0, one looks at the resulting tensors to see if they are reasonable physically. This enables them to remove values of r_0 that are too small. In the example, they are left with $r_0 = 4, 5$.

Now assume r has been determined. Since the value of r is relatively small, up to trivialities, the expression of $\tilde{T}$ as the sum of r rank one elements will be unique, see §3.3. Thus, by performing the decomposition of $\tilde{T}$, one recovers the concentration of each of the r substances in each solution by determining the vectors a_f as well as the individual excitation and emission spectra by determining the vectors b_f.

1.3.2. Cumulants. In statistics one collects data in large quantities that are stored in a multi-dimensional array, and attempts to extract information from the data. An important sequence of quantities to extract from data sets are *cumulants*, the main topic of the book [**229**]. This subsection is an introduction to cumulants, which is continued in Chapter 12.

Let $\mathbb{R}^{\mathbf{v}}$ have a probability measure $d\mu$, i.e., a reasonably nice measure such that $\int_{\mathbb{R}^{\mathbf{v}}} d\mu = 1$. A measurable function $f : \mathbb{R}^{\mathbf{v}} \to \mathbb{R}$ is called a *random variable* in the statistics literature. (The reason for this name is that if we pick a "random point" $x \in \mathbb{R}^{\mathbf{v}}$, according to the probability measure μ, f determines a "random value" $f(x)$.) For a random variable f, write $E\{f\} := \int_{\mathbb{R}^{\mathbf{v}}} f(x) d\mu$. $E\{f\}$ is called the *expectation* or the *mean* of the function f.

For example, consider a distribution of mass in space with coordinates x^1, x^2, x^3 and the density given by a probability measure μ. (Each coordinate function may be considered a random variable.) Then the integrals $m^j := E\{x^j\} := \int_{\mathbb{R}^3} x^j d\mu$ give the coordinates of the center of mass (called the *mean*).

More generally, define the *moments* of random variables x^j:

$$m_x^{i_i,\dots,i_p} := E\{x^{i_1} \cdots x^{i_p}\} = \int_{\mathbb{R}^{\mathbf{v}}} x^{i_1} \cdots x^{i_p} d\mu.$$

Remark 1.3.2.1. In practice, the integrals are approximated by discrete statistical data taken from samples.

A first measure of the dependence of functions x^j is given by the quantities

$$\kappa^{ij} = \kappa_x^{ij} := m^{ij} - m^i m^j = \int_{\mathbb{R}^{\mathbf{v}}} x^i x^j d\mu - (\int_{\mathbb{R}^{\mathbf{v}}} x^j d\mu)(\int_{\mathbb{R}^{\mathbf{v}}} x^j d\mu).$$

called *second order cumulants* or the *covariance matrix*.

One says that the functions x^j are *statistically independent at order two* if $\kappa_x^{ij} = 0$ for all $i \neq j$. To study statistical independence, it is better not to consider the κ_x^{ij} individually, but to form a symmetric matrix $\kappa_{2,x} = \kappa_2(x) \in S^2\mathbb{R}^{\mathbf{v}}$. If the measurements depend on $r < \mathbf{v}$ independent events, the rank of this matrix will be r. (In practice, the matrix will be "close" to a matrix of rank r.) The matrix $\kappa_2(x)$ is called a *covariance matrix*. One can define higher order cumulants to obtain further measurements of statistical independence. For example, consider

$$\kappa^{ijk} = m^{ijk} - (m^i m^{jk} + m^j m^{ik} + m^k m^{ij}) + 2m^i m^j m^k. \tag{1.3.1}$$

We may form a third order symmetric tensor from these quantities, and similarly for higher orders.

Cumulants of a set of random variables (i.e. functions on a space with a probability measure) give an indication of their mutual statistical dependence, and higher-order cumulants of a single random variable are some measure of its non-Gaussianity.

Definition 1.3.2.2. In probability, two events A, B are independent if $Pr(A \wedge B) = Pr(A)Pr(B)$, where $Pr(A)$ denotes the probability of the event A. If x is a random variable, one can compute $Pr(x \leq a)$. Two random variables x, y are *statistically independent* if $Pr(\{x \leq a\} \wedge \{y \leq b\}) = Pr(\{x \leq a\})Pr(\{y \leq b\})$ for all $a, b \in \mathbb{R}_+$. The statistical independence of random variables $x^1, \ldots, x^m$ is defined similarly.

An important property of cumulants, explained in §12.1, is:

If the x^i are determined by r statistically independent quantities, then $\mathbf{R}_S(\kappa_p(x)) \leq r$ (with equality generally holding) for all $p \geq 2$.

We will apply this observation in the next subsection.

1.3.3. Blind source separation. A typical application of blind source separation (BSS) is as follows: Big Brother would like to determine the location of pirate radio transmissions in Happyville. To accomplish this, antennae are set up at several locations to receive the radio signals. How can one determine the location of the sources from the signals received at the antennae?

Let $y^j(t)$ denote the measurements at the antennae at time t. Assume there is a relation of the form

$$\begin{pmatrix} y^1(t) \\ \vdots \\ y^m(t) \end{pmatrix} = \begin{pmatrix} a_1^1 \\ \vdots \\ a_1^m \end{pmatrix} x^1(t) + \cdots + \begin{pmatrix} a_r^1 \\ \vdots \\ a_r^m \end{pmatrix} x^r(t) + \begin{pmatrix} v^1(t) \\ \vdots \\ v^m(t) \end{pmatrix} \tag{1.3.2}$$

which we write in vector notation as

$$y = Ax + v. \tag{1.3.3}$$

Here A is a fixed $m \times r$ matrix, $v = v(t) \in \mathbb{R}^m$ is a vector valued function representing the noise, and $x(t) = (x^1(t), \ldots, x^r(t))^T$ represents the statistically independent functions of t that correspond to the locations of the sources. ("T" denotes transpose.) The $v^i(t)$ are assumed (i.) to be independent random variables which are independent of the $x^j(t)$, (ii.) $E\{v^i\} = 0$, and (iii.) the moments $E\{v^{i_1} \cdots v^{i_p}\}$ are bounded by a small constant.

One would like to recover $x(t)$, plus the matrix A, from knowledge of the function $y(t)$ alone. Note that the x^j are like eigenvectors, in the sense that they are only well defined up to scale and permutation, so "recover" means modulo this ambiguity. And "recover" actually means recover approximate samplings of x from samplings of y.

At first this task appears impossible. However, note that if we have such an equation as (1.3.3), then we can compare the quantities κ_y^{ij} with κ_x^{st} because, by the linearity of the integral

$$\kappa_y^{ij} = A_s^i A_t^j \kappa_x^{st} + \kappa_v^{ij}$$

By the assumption of statistical independence, $\kappa_x^{st} = \delta^{st}\kappa_x^{ss}$, where δ^{st} is the Kronecker delta, and κ_v^{ij} is small, so we ignore it to obtain a system of $\binom{m}{2}$ linear equations for $mr+r$ unknowns. If we assume $m > r$ (this case is called an *overdetermined mixture*), this count is promising. However, although a rank r symmetric $m \times m$ matrix may be written as a sum of rank one symmetric matrices, that sum is never unique. (An algebraic geometer would recognize this fact as the statement that the secant varieties of quadratic Veronese varieties are degenerate.)

Thus we turn to the order three symmetric tensor (cubic polynomial) $\kappa_3(y)$, and attempt to decompose it into a sum of r cubes in order to recover the matrix A. Here the situation is much better: as long as the rank is less then generic, with probability one there will be a unique decomposition (except in $S^3\mathbb{C}^5$ where one needs two less than generic), see Theorem 12.3.4.3. Once one has the matrix A, one can recover the x^j at the sampled times. What is even more amazing is that this algorithm will work in principle even if the number of sources is greater than the number of functions sampled, i.e., if $r > m$ (this is called an *underdetermined mixture*) - see §12.1.

Example 1.3.3.1 (BSS was inspired by nature)**.** How does our central nervous system detect where a muscle is and how it is moving? The muscles send electrical signals through two types of transmitters in the tendons, called primary and secondary, as the first type sends stronger signals. There are two things to be recovered, the function $p(t)$ of angular position and $v(t) = \frac{dp}{dt}$ of angular speed. (These are to be measured at any given instant so your central nervous system can't simply "take a derivative".) One might think one type of transmitter sends information about $v(t)$ and the other about $p(t)$, but the opposite was observed, there is some kind of mixing: say the signals sent are respectively given by functions $f_1(t), f_2(t)$. Then it was observed there is a matrix A, such that

$$\begin{pmatrix} f_1(t) \\ f_2(t) \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} v(t) \\ p(t) \end{pmatrix}$$

and the central nervous system somehow decodes $p(t), v(t)$ from $f_1(t), f_2(t)$. This observation is what led to the birth of blind source separation, see [**95**, §1.1.1].

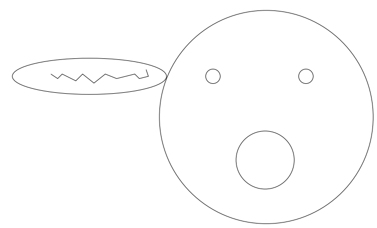


Figure 1.3.2. placeholder until muscle arrives

1.3.4. Other uses of tensor decomposition. Tensor decomposition occurs in numerous areas, here are just a few:

- An early use of tensor decomposition was in the area of *psychometrics*, which sought to use it to help evaluate intelligence and other personality characteristics. Early work in the area includes [**306, 69, 155**].
- In geophysics; for example the interpretation of magnetotelluric data for one-dimensional and two-dimensional regional structures, see, e.g., [**330**] and the references therein.
- In interpreting magnetic resonance imaging (MRI), see, e.g. [**280, 122**], and the references therein. One such use is in determining the location in the brain that is causing epileptic seizures in a patient. Another is the diagnosis and management of stroke.

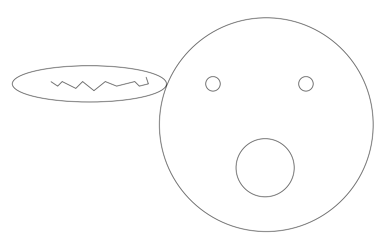


Figure 1.3.3. When locating sources of epileptic seizures in order to remove them, unique up to finite decomposition is not enough.

- In data-mining, see, e.g., [**216**].
- In numerical analysis, thanks to the convergence of Fourier series in $L^2(\mathbb{R}^n)$, one has $L^2(\mathbb{R}^n) = L^2(\mathbb{R}) \otimes \cdots \otimes L^2(\mathbb{R})$, and one often approximates a function of n variables by a finite sum of products of functions of one variable, generalizing classical separation of variables. See, e.g., [**152**] and the references therein.

1.3.5. Practical issues in tensor decomposition. Four issues to deal with are *existence*, *uniqueness*, *executing the decomposition* and *noise*. I now discuss each briefly. In this subsection I use "tensor" for tensors, symmetric tensors, and partially symmetric tensors, "rank" for rank and symmetric rank etc..

Existence. In many tensor decomposition problems, the first issue to resolve is to determine the rank of the tensor T one is handed. In cases where one has explicit equations for the tensors of border rank r, if T solves the equations, then with probability one, it is of rank at most r. (For symmetric tensors of small border rank, it is always of rank at most r, see Theorem 3.5.2.2.)

Uniqueness. In the problems coming from spectroscopy and signal processing, one is also concerned with uniqueness of the decomposition. If the rank is sufficiently small, uniqueness is assured with probability one, see §3.3.3. Moreover there are explicit tests one can perform on any given tensor to be assured of uniqueness, see §3.3.2.

Performing the decomposition. In certain situations there are algorithms that exactly decompose a tensor, see, e.g., §3.5.3 - these generally are a consequence of having equations that test for border rank. In most situations one uses numerical algorithms, which is an area of active research outside the scope of this book. See [**94, 185**] for surveys of decomposition algorithms.

Noise. In order to talk about noise in data, one must have a distance function. The properties of tensors discussed in this book are defined independent of any distance function, and there are no natural distance functions on spaces of tensors (but rather classes of such). For this reason I do not discuss noise or approximation in this book. In specific applications there are often distance functions that are natural based on the science, but often in the literature such functions are chosen by convenience. R. Bro (personal communication) remarks that assuming that the noise has a certain behavior (iid and Gaussian) can determine a distance function.

1.4. P v. NP and algebraic variants

1.4.1. Computing the determinant of a matrix. Given an $n \times n$ matrix, how do you compute its determinant? Perhaps in your first linear algebra class you learned the Laplace expansion - given a 2×2 matrix

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

its determinant is $ad - bc$, then given an $n \times n$ matrix $X = (x^i_j)$,

$$\det(x^i_j) = x^1_1 \Delta_{1,1} - x^1_2 \Delta_{1,2} + \cdots + (-1)^{n-1} x^1_n \Delta_{1,n}$$

where $\Delta_{i,j}$ is the determinant of the $(n-1) \times (n-1)$ matrix obtained by deleting the i-th row and j-th column of X. This algorithm works fine for 3×3 and 4×4 matrices, but by the time one gets to 10×10 matrices, it becomes a bit of a mess.

Aside 1.4.1.1. Let's pause for a moment to think about exactly *what* we are computing, i.e., the meaning of $\det(X)$. The first thing we learn is that X is invertible if and only if its determinant is nonzero. To go deeper, we first need to remind ourselves of the meaning of an $n \times n$ matrix X: what does it *represent*? It may represent many different things - a table of data, a bilinear form, a map between two different vector spaces of the same dimension, but one gets the most meaning of the determinant when X represents a linear map from a vector space to itself. Then, $\det(X)$ represents the product of the eigenvalues of the linear map, or equivalently, the measure of the change in oriented volume the linear map effects on n-dimensional parallelpipeds with one vertex at the origin. These interpretations will be important for what follows, see, e.g., §13.4.2 but for now we continue our computational point of view.

The standard formula for the determinant is

$$\det(X) = \sum_{\sigma \in \mathfrak{S}_n} \operatorname{sgn}(\sigma) x^1_{\sigma(1)} \cdots x^n_{\sigma(n)} \tag{1.4.1}$$

where $\mathfrak{S}_n$ is the *group* of all permutations on n elements and $\operatorname{sgn}(\sigma)$ is the sign of the permutation $\sigma : \{1, \dots, n\} \to \{1, \dots, n\}$. The formula is a sum of $n!$ terms, so for a 10×10 matrix, one would have to perform $9(10!) \simeq 10^7$ multiplications and $10! - 1$ additions to compute the determinant.

The standard method for computing the determinant of large matrices is *Gaussian elimination*. Recall that if X is an upper triangular matrix, its determinant is easy to compute: it is just the product of the diagonal entries. On the other hand, $\det(gX) = \det(g)\det(X)$, so multiplying X by a matrix with determinant one will not change its determinant. To perform Gaussian elimination one chooses a sequence of matrices with determinant one to multiply X by until one obtains an upper triangular matrix whose determinant is then easy to compute. Matrix multiplication, even of arbitrary matrices, uses on the order of n^3 scalar multiplications (actually less, as discussed in §1.1), but even executed naïvely, one uses approximately n^4

multiplications. Thus for a 10×10 matrix one has 10^4 for Gaussian elimination applied naïvely versus 10^7 for (1.4.1) applied naïvely. This difference in complexity is discussed in detail in Chapter 13.

The determinant of a matrix is unchanged by the following operations:

$$X \mapsto gXh$$
$$X \mapsto gX^T h$$

where g, h are $n \times n$ matrices with $\det(g)\det(h) = 1$. In 1897 Frobenius [**129**] proved these are exactly the symmetries of the determinant. On the other hand, a random homogeneous polynomial of degree n in n^2 variables will have hardly any symmetries, and one might attribute the facility of computing the determinant to this large group of symmetries.

We can make this more precise as follows: first note that polynomials with small formulas, such as $f(x^i_j) = x^1_1 x^2_2 \cdots x^n_n$ are easy to compute.

Let $\mathfrak{b} \subset Mat_{n\times n}$ denote the subset of upper triangular matrices, and $G(det_n)$ the symmetry group of the determinant. Observe that $\det_n|_{\mathfrak{b}}$ is just the polynomial f above, and that $G(det_n)$ can move any matrix into $\mathfrak{b}$ relatively cheaply. This gives a geometric proof that the determinant is easy to compute.

One can ask: what other polynomials are easy to compute? More precisely, what polynomials with no known small formulas are easy to compute? An example of one such is given in §13.5.3.

1.4.2. The permanent and counting perfect matchings.

The *marriage problem*: n people are attending a party where the host has n different flavors of ice cream for his guests. Not every guest likes every flavor. We can make a *bipartite graph*, a graph with one set of nodes the set of guests, and another set of nodes the set of flavors, and then draw an edge joining any two nodes that are compatible. A *perfect matching* is possible if everyone can get a suitable dessert. The host is curious as to how many different perfect matchings are possible.

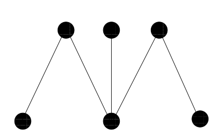


Figure 1.4.1. Andrei Zelevinsky's favorite bipartite graph: top nodes are A, B, C, bottom are a, b, c. Amy is allergic to chocolate, Bob insists on banana, Carol is happy with banana or chocolate. Only Amy likes apricot.

Given a bipartite graph on (n,n) vertices one can check if the graph has a complete matching in polynomial time [**153**]. However there is no known polynomial time algorithm to count the number of perfect matchings.

Problems such as the marriage problem appear to require a number of arithmetic operations that grows exponentially with the size of the data in order to solve them, however a proposed solution can be verified by performing a number of arithmetic operations that grows polynomialy with the size of the data. Such problems are said to be of class **NP**. (See Chapter 13 for precise definitions.)

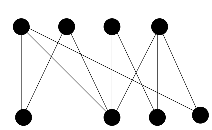
Form an *incidence matrix* $X = (x^i_j)$ for a bipartite graph by letting the upper index correspond to one set of nodes and the lower index the other. One then places a 1 in the (i,j)-th slot if there is an edge joining the corresponding nodes and a zero if there is not.

Define the *permanent* of an $n \times n$ matrix $X = (x^i_j)$ by

$$perm_n(X) := \sum_{\sigma \in \mathfrak{S}_n} x^1_{\sigma(1)} x^2_{\sigma(2)} \cdots x^n_{\sigma(n)}, \tag{1.4.2}$$

and observe the similarities with (1.4.1).

Exercise 1.4.2.1: Verify directly that the permanent of the incidence matrix for the following graph indeed equals its number of perfect matchings.



Exercise 1.4.2.2: Show that if X is an incidence matrix for an (n,n)-bipartite graph Γ, that the number of perfect matchings of Γ is given by the permanent.

1.4.3. Algebraic variants of P v. NP. Matrix multiplication, and thus computing the determinant of a matrix, can be computed by performing a number of arithmetic operations that is polynomial in the size of the data. (If the data size is of order $m = n^2$ then one needs roughly $m^{\frac{3}{2}} = n^3$ operations, or roughly n^4 if one wants an algorithm without divisions or decisions, see §13.4.2.) Roughly speaking, such problems are said to be of class **P**, or are *computable in polynomial time*.

L. Valiant [**307**] had the following idea: Let $P(x^1, \ldots, x^{\mathbf{v}})$ be a homogeneous polynomial of degree m in $\mathbf{v}$ variables. We say P is an *affine projection* of a determinant of size n if there exists an affine linear function $f : \mathbb{C}^{\mathbf{v}} \to Mat_n(\mathbb{C})$ such that $P = det \circ f$. Write $dc(P)$ for the smallest n

such that P may be realized as the affine projection of a determinant of size n.

Comparing the two formulas, the difference between the permanent and the determinant is "just" a matter of a few minus signs. For example, one can convert the permanent of a 2×2 matrix to a determinant by simply changing the sign of one of the off diagonal entries. Initially there was hope that one could do something similar in general, but these hopes were quickly dashed. The next idea was to write the permanent of an $m \times m$ matrix as the determinant of a larger matrix.

Using the sequence $(\det_n)$, Valiant defined an algebraic analog of the class $\mathbf{P}$, denoted $\mathbf{VP}_{ws}$, as follows: a sequence of polynomials (p_m) where p_m is a homogeneous polynomial of degree $d(m)$ in $\mathbf{v}(m)$ variables, with $d(m), \mathbf{v}(m)$ polynomial functions of m, is in the class $\mathbf{VP}_{ws}$ if $dc(p_m) = \mathcal{O}(m^c)$ for some constant c. (For an explanation of the cryptic notation $\mathbf{VP}_{ws}$, see §13.3.3.)

Conjecture 1.4.3.1 (Valiant)**.** *The function $dc(perm_m)$ grows faster than any polynomial in m, i.e.,* $(\mathrm{perm}_m) \notin \mathbf{VP}_{ws}$.

How can one determine $dc(P)$? I discuss geometric approaches towards this question in Chapter 13. One, that goes back to Valiant and has been used in [**324, 232, 231**] involves studying the local differential geometry of the two sequences of zero sets of the polynomials det_n and $perm_m$. Another, due to Mulmuley and Sohoni [**241**] involves studying the algebraic varieties in the vector spaces of homogenous polynomials obtained by taking the orbit closures of the polynomials $\det_n$ and perm_m in the space of homogeneous polynomials of degree n (resp. m) in n^2 (resp. m^2) variables. This approach is based in algebraic geometry and representation theory.

1.5. Algebraic Statistics and tensor networks

1.5.1. Algebraic statistical models. As mentioned above, in statistics one is handed data, often in the form of a multi-dimensional array, and is asked to extract meaningful information from the data. Recently the field of *algebraic statistics* has arisen.

Instead of asking *What is the meaningful information to be extracted from this data?* one asks *How can I partition the set of* **all** *arrays of a given size into subsets of data sets sharing similar attributes?*

Consider a weighted 6 sided die, for $1 \leq j \leq 6$, let p_j denote the probability that j is thrown, so $0 \leq p_j \leq 1$ and $\sum_j p_j = 1$. We record the information in a vector $p \in \mathbb{R}^6$. Now say we had a second die, say 20 sided, with probabilities q_s, $0 \leq q_s \leq 1$ and $q_1 + \cdots + q_{20} = 1$. Now if we throw the dice together, assuming the events are independent, the probability of throwing i for the first and s for the second is simply $p_i q_s$. We may form a 6×20 matrix $x = (x_{i,s}) = (p_i q_s)$ recording all the possible throws with their probabilities. Note $x_{i,s} \geq 0$ for all i, s and $\sum_{i,s} x_{i,s} = 1$. The matrix x has an additional property: *x has rank one.*

Were the events not independent we would not have this additional constraint. Consider the set $\{T \in \mathbb{R}^6 \otimes \mathbb{R}^{20} \mid T_{i,s} \geq 0, \sum_{i,s} T_{i,s} = 1\}$. This is the set of all discrete probability distributions on $\mathbb{R}^6 \otimes \mathbb{R}^{20}$, and the set of the previous paragraph is this set intersected with the set of rank one matrices.

Now say some gamblers were cheating with κ sets of dice, each with different probabilities. They watch to see how bets are made and then choose one of the sets accordingly. Now we have probabilities $p_{i,u}$, $q_{s,u}$, and a $6 \times 20 \times \kappa$ array $z_{i,s,u}$ with $\mathrm{rank}(z) = 1$, in the sense that if we consider z as a bilinear map, it has rank one.

Say that we cannot observe the betting. Then, to obtain the probabilities of what we can observe, we must sum over all the κ possibilities. We end up with an element of $\mathbb{R}^6 \otimes \mathbb{R}^{20}$, with entries $r_{i,s} = \sum_u p_{i,u} q_{i,u}$. That is, we obtain a 6×20 matrix of probabilities of rank (at most) κ, i.e., an element of $\hat{\sigma}_{\kappa, \mathbb{R}^6 \otimes \mathbb{R}^{20}}$. The set of all such distributions is the set of matrices of $\mathbb{R}^6 \otimes \mathbb{R}^{20}$ of rank at most κ intersected with $PD_{6,20}$.

This is an example of a *Bayesian network* . In general, one associates a graph to a collection of random variables having various conditional dependencies and then from such a graph, one defines sets (varieties) of distributions. More generally an *algebraic statistical model* is the intersection of the probability distributions with a closed subset defined by some dependence relations. Algebraic statistical models are discussed in Chapter 14.

A situation discussed in detail in Chapter 14 are algebraic statistical models arising in phylogeny: Given a collection of species, say humans, monkeys, gorillas, orangutans, and ...., all of which are assumed to have evolved from some common ancestor, ideally we might like to reconstruct the corresponding evolutionary tree from sampling DNA. Assuming we can only measure the DNA of existing species, this will not be completely possible, but it might be possible to, e.g., determine which pairs are most closely related.

One might imagine, given the numerous possibilities for evolutionary trees, that there would be a horrific amount of varieties to find equations for. A major result of E. Allmann and J. Rhodes states that this is not the case:

Theorem 1.5.1.1. [**10, 9**] *Equations for the algebraic statistical model associated to any bifurcating evolutionary tree can be determined explicitly from equations for* $\hat{\sigma}_{4, \mathbb{C}^4 \otimes \mathbb{C}^4 \otimes \mathbb{C}^4}$.

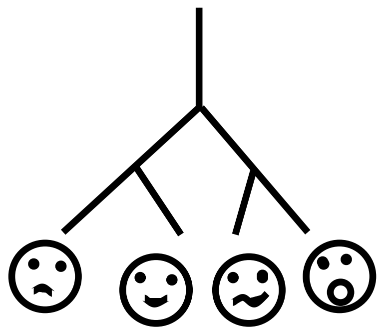


Figure 1.5.1. Evolutionary tree, extinct ancestor gave rise to 4 species

See §14.2.3 for a more precise statement. Moreover, motivated by Theorem 1.5.1.1 (and the promise of a hand smoked alaskan salmon), equations for $\hat{\sigma}_{4,\mathbb{C}^4\otimes\mathbb{C}^4\otimes\mathbb{C}^4}$ were recently found by S. Friedland, see §3.9.3 for the equations and salmon discussion.

The geometry of this field has been well developed and exposed by geometers [**257, 165**]. I include a discussion in Chapter 14.

1.5.2. Tensor network states. Tensors describe states of quantum mechanical systems. If a system has n particles, its state is an element of $V_1\otimes\cdots\otimes V_n$ where V_j is a Hilbert space associated to the j-th particle. In many-body physics, in particular solid state physics, one wants to simulate quantum states of thousands of particles, often arranged on a regular lattice (e.g., atoms in a crystal). Due to the exponential growth of the dimension of $V_1\otimes\cdots\otimes V_n$ with n, any naïve method of representing these tensors is intractable on a computer. *Tensor network states* were defined to reduce the complexity of the spaces involved by restricting to a subset of tensors that is physically reasonable, in the sense that the corresponding spaces of tensors are only "locally" entangled because interactions (entanglement) in the physical world appear to just happen locally.

Such spaces have been studied since the 1980's. These spaces are associated to graphs, and go under different names: *tensor network states*, *finitely correlated states (FCS)*, *valence-bond solids (VBS)*, *matrix product states (MPS)*, *projected entangled pairs states (PEPS)*, and *multi-scale entanglement renormalization ansatz states (MERA)*, see, e.g., [**276, 121, 170, 120, 318, 87**] and the references therein. I will use the term tensor network states.

The topology of the tensor network is often modeled to mimic the physical arrangement of the particles and their interactions.

Just as phylogenetic trees, and more generally Bayes models, use graphs to construct varieties in spaces of tensors that are useful for the problem at hand, in physics one uses graphs to construct varieties in spaces of tensors that model the "feasible" states. The precise recipe is given in §14.1, where I also discuss geometric interpretations of the tensor network states arising from chains, trees and loops. The last one is important for physics; large loops are referred to as "1-D systems with periodic boundary conditions" in the physics literature and are the prime objects people use in practical simulations today.

To entice the reader uninterested in physics, but perhaps interested in complexity, here is a sample result:

Proposition 1.5.2.1. [**213**] *Tensor networks associated to graphs that are triangles consist of matrix multiplication (up to relabeling) and its degenerations*

See Proposition 14.1.4.1 for a more precise statement. Proposition 1.5.2.1 leads to a surprising connection between the study of tensor network states and the geometric complexity theory program mentioned above and discussed in §13.6.

1.6. Geometry and representation theory

So far we have seen the following

- A key to determining the complexity of matrix multiplication will be to find explicit equations for the set of tensors in $\mathbb{C}^{n^2}\otimes\mathbb{C}^{n^2}\otimes\mathbb{C}^{n^2}$ of border rank at most r in the range $\frac{3}{2}n^2 \leq r \leq n^{2.38}$.
- For flourescence spectroscopy and other applications, one needs to determine the ranks of tensors (of small rank) in $\mathbb{R}^I\otimes\mathbb{R}^J\otimes\mathbb{R}^K$ and to decompose them. Equations will be useful both for determining what the rank of an approximating tensor should be and for developing explicit algorithms for tensor decomposition.
- To study cumulants and blind source separation, one is interested in the analogous questions for symmetric tensor rank.
- In other applications discussed in this book, such as the GCT program for the Mulmuley-Sohoni variant of **P** v. **NP**, and in algebraic statistics, a central goal is to find the equations for other algebraic varieties arising in spaces of tensors.

To find equations we will exploit the symmetries of the relevant varieties. We will also seek geometric descriptions of the varieties. For example, Example 1.2.6.1 admits the simple interpretation that the limit of the sequence of secant lines is a tangent line as illustrated in Figure 2.4.5. This interpretation is explained in detail in Chapter 5, and exploited many times in later chapters.

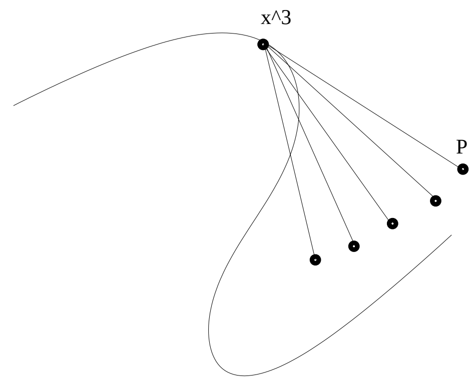


Figure 1.6.1. Unlabeled points are various P_ϵ's lying on secant lines to the curve of perfect cubes

Given a multidimensional array in $\mathbb{R}^{\mathbf{a}} \otimes \mathbb{R}^{\mathbf{b}} \otimes \mathbb{R}^{\mathbf{c}}$, its rank and border rank will be unchanged if one changes bases in $\mathbb{R}^{\mathbf{a}}$, $\mathbb{R}^{\mathbf{b}}$, $\mathbb{R}^{\mathbf{c}}$. Another way to say this is that the properties of rank and border rank are *invariant* under the action of the group of changes of bases $G := GL_{\mathbf{a}} \times GL_{\mathbf{b}} \times GL_{\mathbf{c}}$. When one looks for the defining equations of $\hat{\sigma}_r$, the space of equations will also be invariant under the action of G. *Representation theory* provides a way to organize all polynomials into sums of subspaces invariant under the action of G. It is an essential tool for the study of equations.

Chapter 2

Multilinear algebra

This chapter approaches multilinear algebra from a geometric perspective. If $X = (a^i_s)$ is a matrix, one is not so much interested in the collection of numbers that make up X, but what X *represents* and what qualitative information can be extracted from the entries of X. For this reason and others, in §2.3 an invariant definition of tensors is given and its utility is explained, especially in terms of groups acting on spaces of tensors. Before that, the chapter begins in §2.1 with a collection of exercises to review facts from linear algebra that will be important in what follows. For those readers needing a reminder of the basic definitions in linear algebra, they are given in an appendix §2.9. Basic definitions regarding group actions that will be needed throughout are stated in §2.2. In §2.4, rank and border rank of tensors are defined, Strassen's algorithm is revisited, and basic results about rank are established. A more geometric perspective on the matrix multiplication operator in terms of contractions is given in §2.5. Among subspaces of spaces of tensors, the *symmetric* and *skew-symmetric* tensors discussed in §2.6 are distinguished, not only because they are the first subspaces one generally encounters, but all other natural subspaces may be built out of symmetrizations and skew-symmetrizations. As a warm up for the detailled discussions of polynomials that appear later in the book, polynomials on the space of matrices are discussed in §2.7. In §2.8, $V^{\otimes 3}$ is decomposed as a $GL(V)$-module, which serves as an introduction to Chapter 6.

There are three appendices to this chapter. As mentioned above, in §2.9 basic definitions are recalled for the reader's convenience. §2.10 reviews Jordan and rational canonical forms. Wiring diagrams, a useful pictorial tool for studying tensors are introduced in §2.11.

I work over the field $\mathbb{C}$. Unless otherwise noted, A_j, A, B, C, V, and W are finite dimensional complex vector spaces respectively of dimensions $\mathbf{a}_j, \mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{v}$, and $\mathbf{w}$. V^* denotes The dual vector space to V (see §1.2.1) is denoted V^*.

2.1. Rust removal exercises

For $\alpha \in V^*$, let $\alpha^\perp := \{v \in V \mid \alpha(v) = 0\}$ and for $w \in V$, let $\langle w \rangle$ denote the span of w.

Definition 2.1.0.2. Let $\alpha \in V^*$, and $w \in W$. Consider the linear map $\alpha \otimes w : V \to W$ defined by $v \mapsto \alpha(v)w$. A linear map of this form is said to be of *rank one*.

Observe that

$$\ker(\alpha \otimes w) = \alpha^\perp,$$
$$\mathrm{Image}(\alpha \otimes w) = \langle w \rangle.$$

(1) Show that if one chooses bases of V and W, the matrix representing $\alpha \otimes w$ has rank one.

(2) Show that every rank one $n \times m$ matrix is the product of a column vector with a row vector. To what extent is this presentation unique?

(3) Show that a nonzero matrix has rank one if and only if all its 2×2 minors are zero. ⊚

(4) Show that the following definitions of the *rank* of a linear map $f : V \to W$ are equivalent
   (a) $\dim f(V)$
   (b) $\dim V - \dim(\ker(f))$
   (c) The smallest r such that f is the sum of r rank one linear maps.
   (d) The smallest r such that any matrix representing f has all size $r+1$ minors zero.
   (e) There exist choices of bases in V and W such that the matrix of f is $\begin{pmatrix} \mathrm{Id}_r & 0 \\ 0 & 0 \end{pmatrix}$ where the blocks in the previous expression come from writing $\dim V = r + (\dim V - r)$ and $\dim W = r + (\dim W - r)$ and Id_r is the $r \times r$ identity matrix.

(5) Given a linear subspace $U \subset V$, define $U^\perp \subset V^*$, the *annihilator of* U, by $U^\perp := \{\alpha \in V^* \mid \alpha(u) = 0\ \forall u \in U\}$. Show that $(U^\perp)^\perp = U$.

(6) Show that for a linear map $f : V \to W$, that $\ker f = (\mathrm{Image}\, f^T)^\perp$. (See 2.9.1.6 for the definition of the transpose $f^T : W^* \to V^*$.)

This is sometimes referred to as the *fundamental theorem of linear algebra*. It implies $\operatorname{rank}(f) = \operatorname{rank}(f^T)$, i.e., that for a matrix, *row rank equals column rank*, as was already seen in Exercise (4) above.

(7) Let V denote the vector space of 2×2 matrices. Take a basis

$$v_1 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \ v_2 = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \ v_3 = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \ v_4 = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$$

Fix $a, b, c, d \in \mathbb{C}$ and let

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

Write out a 4×4 matrix expressing the linear map

$$\begin{aligned} L_A : V &\to V \\ X &\mapsto AX \end{aligned}$$

that corresponds to left multiplication by A. Write the analogous matrix for right multiplication. For which matrices A are the two induced linear maps the same?

(8) Given a 2×2 matrix A, write out a 4×4 matrix expressing the linear map

$$\begin{aligned} ad(A) : V &\to V \\ X &\mapsto AX - XA. \end{aligned}$$

What is the largest possible rank of this linear map?

(9) Let A be a 3×3 matrix and write out a 9×9 matrix representing the linear map $L_A : Mat_{3\times 3} \to Mat_{3\times 3}$.

(10) Choose new bases such that the matrices of Exercises 7 and 9 become block diagonal (i.e., the only non-zero entries occur in 2×2 (resp. 3×3) blocks centered about the diagonal). What will the $n \times n$ case look like?

(11) A vector space V admits a *direct sum decomposition* $V = U \oplus W$ if $U, W \subset V$ are linear subspaces and if for all $v \in V$ there exists unique $u \in U$ and $w \in W$ such that $v = u + w$. Show that that a necessary and sufficient condition to have a direct sum decomposition $V = U \oplus W$ is that $\dim U + \dim W \geq \dim V$ and $U \cap W = (0)$. Similarly, show that another necessary and sufficient condition to have a direct sum decomposition $V = U \oplus W$ is that $\dim U + \dim W \leq \dim V$ and $\operatorname{span}\{U, W\} = V$.

(12) Let $S^2\mathbb{C}^n$ denote the vector space of symmetric $n \times n$ matrices. Calculate $\dim S^2\mathbb{C}^n$. Let $\Lambda^2\mathbb{C}^n$ denote the vector space of skew

symmetric $n \times n$ matrices. Calculate its dimension, and show that there is a direct sum decomposition

$$Mat_{n\times n} = S^2(\mathbb{C}^n) \oplus \Lambda^2(\mathbb{C}^n).$$

(13) Let $v_1, \ldots, v_{\mathbf{v}}$ be a basis of V, let $\alpha^i \in V^*$ be defined by $\alpha^i(v_j) = \delta^i_j$. (Recall that a linear map is uniquely specified by prescribing the image of a basis.) Show that $\alpha^1, \ldots, \alpha^{\mathbf{v}}$ is a basis for V^*, called the *dual basis* to $v_1, \ldots, v_n$. In particular, $\dim V^* = \mathbf{v}$.

(14) Define, in a coordinate-free way, an injective linear map $V \to (V^*)^*$. (Note that the map would not necessarily be surjective if V were an infinite-dimensional vector space.)

(15) A *filtration* of a vector space is a sequence of subspaces $0 \subset V_1 \subset V_2 \subset \cdots \subset V$. Show that a filtration of V naturally induces a filtration of V^*.

(16) Show that by fixing a basis of V, one obtains an identification of the group of invertible endomorphisms of V, denoted $GL(V)$, and the set of bases of V.

2.2. Groups and representations

A significant part of our study will be to exploit symmetry to better understand tensors. The set of symmetries of any object forms a *group*, and the realization of a group as a group of symmetries is called a *representation* of a group. The most important group in our study will be $GL(V)$, the group of invertible linear maps $V \to V$, which forms a group under the composition of mappings.

2.2.1. The group $GL(V)$. If one fixes a reference basis, $GL(V)$ is the group of changes of bases of V. If we use our reference basis to identify V with $\mathbb{C}^{\mathbf{v}}$ equipped with its standard basis, $GL(V)$ may be identified with the set of invertible $\mathbf{v} \times \mathbf{v}$ matrices. I sometimes write $GL(V) = GL_{\mathbf{v}}$ or $GL_{\mathbf{v}}\mathbb{C}$ if V is $\mathbf{v}$-dimensional and comes equipped with a basis. I emphasize $GL(V)$ as a group rather than the invertible $\mathbf{v} \times \mathbf{v}$ matrices because it not only acts on V, but on many other spaces constructed from V.

Definition 2.2.1.1. Let W be a vector space, let G be a group, and let $\rho : G \to GL(W)$ be a *group homomorphism* (see §2.9.2.4). (In particular, $\rho(G)$ is a subgroup of $GL(W)$.) A group homomorphism $\rho : G \to GL(W)$ is called a *(linear) representation* of G. One says *G acts on W*, or that W is a *G-module*.

For $g \in GL(V)$ and $v \in V$, write $g \cdot v$, or $g(v)$ for the action. Write $\circ$ for the composition of maps.

Example 2.2.1.2. Here are some actions: $g \in GL(V)$ acts on

(1) V^* by $\alpha \mapsto \alpha \circ g^{-1}$.
(2) $End(V)$ by $f \mapsto g \circ f$.
(3) A second action on $End(V)$ is by $f \mapsto g \circ f \circ g^{-1}$.
(4) The vector space of homogeneous polynomials of degree d on V for each d by $P \mapsto g \cdot P$,where $g \cdot P(v) = P(g^{-1}v)$. Note that this agrees with (1) when $d = 1$.
(5) Let $V = \mathbb{C}^2$ so the standard action of $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in GL(V)$ on $\mathbb{C}^2$ is, $\begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$. Then GL_2 also acts on $\mathbb{C}^3$ by

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} \mapsto \begin{pmatrix} a^2 & ac & c^2 \\ 2ab & ad+bc & 2cd \\ b^2 & bd & d^2 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix}.$$

The geometry of this action is explained by Exercise 2.6.23

These examples give group homomorphisms $GL(V) \to GL(V^*)$, $GL(V) \to GL(End(V))$ (two different ones) and $GL(V) \to GL(S^dV^*)$, where S^dV^* denotes the vector space of homogeneous polynomials of degree d on V.

Exercise 2.2.1.3: Verify that each of the above examples are indeed actions, e.g., show $(g_1g_2) \cdot \alpha = g_1(g_2 \cdot \alpha)$.

Exercise 2.2.1.4: Let $\dim V = 2$, choose a basis of V so that $g \in GL(V)$ may be written $g = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$. Write out the 4×4 matrices for examples 2.2.1.2.(2) and 2.2.1.2.(3).

2.2.2. Modules and submodules. If W is a G-module and there is a linear subspace $U \subset W$ such that $g \cdot u \in U$ for all $g \in G$ and $u \in U$, then one says U is a *G-submodule* of W.

Exercise 2.2.2.1: Using Exercise 2.2.1.4 show that both actions on $End(V)$ have nontrivial submodules, in the first case (when $\dim V = 2$) one can find two-dimensional subspaces preserved by $GL(V)$ and in the second there is a unique one dimensional subspace and an unique three dimensional subspace preserved by $GL(V)$.⊚

A module is *irreducible* if it contains no nonzero proper submodules. For example, the action 2.2.1.2.(3) restricted to the trace free linear maps is irreducible.

If $Z \subset W$ is a subset and a group G acts on W, one says Z is *invariant* under the action of G if $g \cdot z \in Z$ for all $z \in Z$ and $g \in G$.

2.2.3. Exercises.

(1) Let $\mathfrak{S}_n$ denote the group of permutations of $\{1, \ldots, n\}$ (see Definition 2.9.2.2). Endow $\mathbb{C}^n$ with a basis. Show that the action of $\mathfrak{S}_n$ on $\mathbb{C}^n$ defined by permuting basis elements, i.e., given $\sigma \in \mathfrak{S}_n$ and a basis $e_1, \ldots, e_n$, $\sigma \cdot e_j = e_{\sigma(j)}$, is not irreducible.⊚

(2) Show that the action of GL_n on $\mathbb{C}^n$ is irreducible.

(3) Show that the map $GL_p \times GL_q \to GL_{pq}$ given by (A, B) acting on a $p \times q$ matrix X by $X \mapsto AXB^{-1}$ is a linear representation.

(4) Show that the action of the group of invertible upper triangular matrices on $\mathbb{C}^n$ is not irreducible. ⊚

(5) Let Z denote the set of rank one $p \times q$ matrices inside the vector space of $p \times q$ matrices. Show Z is invariant under the action of $GL_p \times GL_q \subset GL_{pq}$.

2.3. Tensor products

In physics, engineering and other areas, tensors are often defined to be multidimensional arrays. Even a linear map is often defined in terms of a matrix that represents it in a given choice of basis. In what follows I give more invariant definitions, and with a good reason.

Consider for example the space of $\mathbf{v} \times \mathbf{v}$ matrices, first as representing the space of linear maps $V \to W$, where $\mathbf{w} = \mathbf{v}$. Then given $f : V \to W$, one can always make changes of bases in V and W such that the matrix of f is of the form

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

where the blocking is $(k, \mathbf{v} - k) \times (k, \mathbf{v} - k)$, so there are only $\mathbf{v}$ different such maps up to equivalence. On the other hand, if the space of $\mathbf{v} \times \mathbf{v}$ matrices represents the linear maps $V \to V$, then one can only have Jordan (or rational) canonical form, i.e., there are parameters worth of distinct matrices up to equivalence.

Because it will be essential to keep track of group actions in our study, I give basis-free definitions of linear maps and tensors.

2.3.1. Definitions.

Notation 2.3.1: Let $V^* \otimes W$ denote the vector space of linear maps $V \to W$. With this notation, $V \otimes W$ denotes the linear maps $V^* \to W$.

The space $V^* \otimes W$ may be thought of in four different ways: as the space of linear maps $V \to W$, as the space of linear maps $W^* \to V$ (using the isomorphism determined by taking transpose), as the dual vector space to $V\otimes W^*$, by Exercise 2.3.2.(3) below, and as the space of bilinear maps $V\times W^*\to\mathbb{C}$. If one chooses bases and represents $f\in V^*\otimes W$ by a $\mathbf{v}\times\mathbf{w}$ matrix $X=(f^i_s)$, the first action is multiplication by a column vector $v\mapsto Xv$, the second by right multiplication by a row vector $\beta\mapsto\beta X$, the third by, given an $\mathbf{w}\times\mathbf{v}$ matrix $Y=(g^s_i)$, taking $\sum_{i,s}f^i_sg^s_i$, and the fourth by $(v,\beta)\mapsto f^i_sv_i\beta^s$.

Exercise 2.3.1.1: Show that the rank one elements in $V\otimes W$ span $V\otimes W$. More precisely, given bases (v_i) of V and (w_s) of W, show that the $\mathbf{vw}$ vectors $v_i\otimes w_s$ provide a basis of $V\otimes W$.

Exercise 2.3.1.2: Let $v_1,\dots,v_n$ be a basis of V with dual basis $\alpha^1,\dots,\alpha^n$. Write down an expression for a linear map as a sum of rank one maps $f:V\to V$ such that each v_i is an eigenvector with eigenvalue λ_i, that is $f(v_i)=\lambda_iv_i$ for some $\lambda_i\in\mathbb{C}$. In particular, write down an expression for the identity map (case all $\lambda_i=1$).

Definition 2.3.1.3. Let $V_1,\dots,V_k$ be vector spaces. A function

$$f:V_1\times\ldots\times V_k\to\mathbb{C} \tag{2.3.1}$$

is *multilinear* if it is linear in each factor V_ℓ. The space of such multilinear functions is denoted $V_1^*\otimes V_2^*\otimes\cdots\otimes V_k^*$, and called the *tensor product* of the vector spaces $V_1^*,\dots,V_k^*$. Elements $T\in V_1^*\otimes\cdots\otimes V_k^*$ are called *tensors*. The integer k is sometimes called the *order* of T. The sequence of natural numbers $(\mathbf{v}_1,\dots,\mathbf{v}_k)$ is sometimes called the *dimensions* of T.

More generally, a function

$$f:V_1\times\ldots\times V_k\to W \tag{2.3.2}$$

is *multilinear* if it is linear in each factor V_ℓ. The space of such multilinear functions is denoted $V_1^*\otimes V_2^*\otimes\ldots\otimes V_k^*\otimes W$, and called the *tensor product* of $V_1^*,\dots,V_k^*,W$.

If $f:V_1\times V_2\to W$ is bilinear, define the left kernel

$$\mathrm{Lker}(f)=\{v\in V_1\mid f(v_1,v_2)=0\ \forall v_2\in V_2\}$$

and similarly for the right kernel $\mathrm{Rker}(f)$. For multi-linear maps one analogously defines the *i-th kernel*.

When studying tensors in $V_1\otimes\cdots\otimes V_n$, introduce the notation $V_{\hat{j}}:=V_1\otimes\cdots\otimes V_{j-1}\otimes V_{j+1}\otimes\cdots\otimes V_n$. Given $T\in V_1\otimes\cdots\otimes V_n$, write $T(V_j^*)\subset V_{\hat{j}}$ for the image of the linear map $V_j^*\to V_{\hat{j}}$.

Definition 2.3.1.4. Define the *multilinear rank* (sometimes called the *duplex rank* or *Tucker rank*) of $T\in V_1\otimes\cdots\otimes V_n$ to be the n-tuple of natural numbers $\mathrm{R}_{\mathrm{multlin}}(T):=(\dim T(V_1^*),\dots,\dim T(V_n^*))$.

The number $\dim(T(V_j^*))$ is sometimes called the *mode j rank* of T.

Write $V^{\otimes k} := V\otimes\cdots\otimes V$ where there are k copies of V in the tensor product.

Remark 2.3.1.5. Some researchers like to picture tensors given in bases in terms of *slices*. Let A have basis $a_1,\dots,a_{\mathbf{a}}$ and similarly for B, C, let $T\in A\otimes B\otimes C$, so in bases $T = T^{i,s,u}a_i\otimes b_s\otimes c_u$. Then one forms an $\mathbf{a}\times\mathbf{b}\times\mathbf{c}$ rectangular solid table whose entries are the T^{isu}. This solid is then decomposed into *modes* or *slices*, e.g., consider T as a collection of $\mathbf{a}$ matrices of size $\mathbf{b}\times\mathbf{c}$: $(T^{1,s,u}),\dots,(T^{\mathbf{a},s,u})$, which might be referred to as *horizontal slices* (e.g. [**185**, p. 458]), or a collection of $\mathbf{b}$ matrices $(T^{i,1,u}),\dots,(T^{i,\mathbf{b},u})$ called *lateral slices*, or a collection of $\mathbf{c}$ matrices called *frontal slices*. When two indices are fixed, the resulting vector in the third space is called a *fiber*.

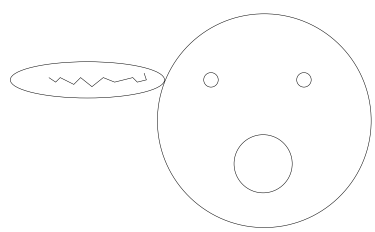


Figure 2.3.1. Placeholder for pictures to be furnished by Kolda

2.3.2. Exercises.

(1) Write out the slices of the 2×2 matrix multiplication operator $M\in A\otimes B\otimes C = (U^*\otimes V)\otimes(V^*\otimes W)\otimes(W^*\otimes U)$ with respect to the basis $a_1 = u^1\otimes v_1, a_2 = u^1\otimes v_2, a_3 = u^2\otimes v_1, a_4 = u^2\otimes v_2$ of A and the analogous bases for B, C.

(2) Verify that the space of multilinear functions (2.3.2) is a vector space.

(3) Given $\alpha\in V^*$, $\beta\in W^*$, allow $\alpha\otimes\beta\in V^*\otimes W^*$ to act on $V\otimes W$ by, for $v\in V, w\in W$, $\alpha\otimes\beta(v\otimes w) = \alpha(v)\beta(w)$ and extending linearly. Show that this identification defines an isomorphism $V^*\otimes W^*\cong (V\otimes W)^*$.

(4) Show that $V\otimes\mathbb{C}\simeq V$.

(5) Show that for each $I\subset\{1,\dots,k\}$ with complementary index set I^c, that there are canonical identifications of $V_1^*\otimes\cdots\otimes V_k^*$ with the space of multi-linear maps $V_{i_1}\times\cdots\times V_{i_{|I|}}\to V_{i_1^c}^*\otimes\cdots\otimes V_{i_{k-|I|}^c}^*$.

(6) A bilinear map $f : U\times V \to \mathbb{C}$ is called a *perfect pairing* if $\mathrm{Lker}(f) = \mathrm{Rker}(f) = 0$. Show that if f is a perfect pairing it determines an identification $U \simeq V^*$.

(7) Show that the map $tr : V^* \times V \to \mathbb{C}$ given by $(\alpha, v) \mapsto \alpha(v)$ is bilinear and thus is well defined as a linear map $V^*{\otimes}V \to \mathbb{C}$. For $f \in V^*{\otimes}V$, show that $tr(f)$ agrees with the usual notion of the trace of a linear map, that is, the sum of the eigenvalues or the sum of the entries on the diagonal in any matrix expression of f.

(8) Show that $tr \in V{\otimes}V^*$, when considered as a map $V \to V$ is the identity map, i.e., that tr and Id_V are the same tensors. Write out this tensor with respect to any choice of basis of V and dual basis of V^*.

(9) A linear map $p : V \to V$ is a *projection* if $p^2 = p$. Show that if p is a projection, then $\mathrm{trace}(p) = \dim(\mathrm{image}(p))$.

(10) Show that a basis of V induces a basis of $V^{\otimes d}$ for each d.

(11) Determine $\dim(V_1{\otimes}\cdots\otimes V_k)$, in terms of $\dim V_i$.

2.4. The rank and border rank of a tensor

2.4.1. The rank of a tensor. Given $\beta_1 \in V_1^*, \ldots, \beta_k \in V_k^*$, define an element $\beta_1{\otimes}\cdots{\otimes}\beta_k \in V_1^*{\otimes}V_2^*{\otimes}\cdots{\otimes}V_k^*$ by

$$\beta_1{\otimes}\cdots{\otimes}\beta_k(u_1, \ldots, u_k) = \beta_1(u_1)\cdots\beta_k(u_k). \tag{2.4.1}$$

Definition 2.4.1.1. An element of $V_1^*{\otimes}V_2^*{\otimes}\cdots{\otimes}V_k^*$ is said to have *rank one* if it may be written as in (2.4.1).

Note that the property of having rank one is independent of any choices of basis. (The vectors in each space used to form the rank one tensor will usually not be basis vectors, but linear combinations of them.)

Definition 2.4.1.2. Define the *rank* of a tensor $T \in V_1{\otimes}V_2{\otimes}\ldots{\otimes}V_k$, denoted $\mathbf{R}(T)$, to be the minimal number r such that $T = \sum_{u=1}^{r} Z_u$ with each Z_u rank one.

Note that the rank of a tensor is unchanged if one makes changes of bases in the vector spaces V_i. Rank is sometimes called *outer product rank* in the tensor literature.

2.4.2. Exercises on ranks of tensors. In what follows, A, B, and C are vector spaces with bases respectively $a_1, \ldots, a_{\mathbf{a}}$, $b_1, \ldots, b_{\mathbf{b}}$, and $c_1, \ldots, c_{\mathbf{c}}$.

(1) Show that $V_1^*{\otimes}V_2^*{\otimes}\ldots{\otimes}V_k^*$ is spanned by its rank one vectors.

(2) Compute the ranks of the following tensors $T_1 = a_1{\otimes}b_1{\otimes}c_1 + a_1{\otimes}b_2{\otimes}c_1$, $T_2 = a_1{\otimes}b_1{\otimes}c_1 + a_2{\otimes}b_2{\otimes}c_1$, $T_3 = a_1{\otimes}b_1{\otimes}c_1 + a_2{\otimes}b_2{\otimes}c_2$.

(3) Show that for T_2 above, $\dim(T_2(C^*)) \neq \dim(T_2(B^*))$.

(4) For $T \in A{\otimes}B{\otimes}C$, show that $\mathbf{R}(T) \geq \dim T(A^*)$.

(5) Show that for all $T \in V_1{\otimes}\cdots{\otimes} V_k$, $\mathbf{R}(T) \leq \Pi_j(\dim V_j)$.

(6) Show that if $T \in A_1^*{\otimes}\cdots{\otimes} A_n^*$, then the multilinear rank $(\mathbf{b}_1, \dots, \mathbf{b}_n)$ of T satisfies $\mathbf{b}_i \leq \min(\mathbf{a}_i, \Pi_{j\neq i}\mathbf{a}_j)$ and that equality holds for generic tensors, in the sense that equality will hold for most small perturbations of any tensor.

By Exercise 2.4.2.(3), one sees that numbers that coincided for linear maps fail to coincide for tensors, i.e., the analog of the fundamental theorem of linear algebra is false for tensors.

2.4.3. $GL(V)$ **acts on** $V^{\otimes d}$**.** The group $GL(V)$ acts on $V^{\otimes d}$. The action on rank one elements is, for $g \in GL(V)$ and $v_1{\otimes}\cdots{\otimes} v_d \in V^{\otimes d}$,

$$g \cdot (v_1{\otimes}\cdots{\otimes} v_d) = (g \cdot v_1){\otimes}\cdots{\otimes} (g \cdot v_d) \tag{2.4.2}$$

and the action on $V^{\otimes d}$ is obtained by extending this action linearly.

Similarly, $GL(V_1) \times \cdots \times GL(V_k)$ acts on $V_1{\otimes}\cdots{\otimes} V_k$.

Exercises 2.4.3.1:

1. Let $\dim V = 2$ and give V basis e_1, e_2 and dual basis α^1, α^2. Let $g \in GL(V)$ be $g = \alpha^1{\otimes}(e_1+e_2)+\alpha^2{\otimes}e_2$. Compute $g\cdot(e_1{\otimes}e_1{\otimes}e_2 + e_1{\otimes}e_2{\otimes}e_2)$.
2. For fixed $g \in GL(V)$, show that the map $V^{\otimes d} \to V^{\otimes d}$ given on rank one elements by (2.4.2) is well defined (i.e., independent of our choice of basis).
3. Let $V = \mathbb{C}^2$ and let $d = 3$. The action of $GL(V)$ on $V^{\otimes 3}$ determines an embedding $GL(V) \subset GL(\mathbb{C}^8) = GL(V^{\otimes 3})$. Fix a basis of V, and write an 8×8 matrix expressing the image of $GL(V)$ in GL_8 with respect to the induced basis (as in Exercise 2.3.2.10).

2.4.4. Strassen's algorithm revisited. The standard algorithm for the multiplication of 2×2 matrices may be expressed in terms of tensors as follows. Let V_1, V_2, V_3 each denote the space of 2×2 matrices. Give V_1 the standard basis a^i_j for the matrix with a 1 in the (i, j)-th slot and zeros elsewhere and let α^i_j denote the dual basis element of V_1^*. Similarly for V_2, V_3. Then the standard algorithm is:

$$\begin{aligned} M_{2,2,2} =& \alpha^1_1{\otimes}\beta^1_1{\otimes}c^1_1 + \alpha^1_2{\otimes}\beta^2_1{\otimes}c^1_1 + \alpha^2_1{\otimes}\beta^1_1{\otimes}c^2_1 + \alpha^2_2{\otimes}\beta^2_1{\otimes}c^2_1 \\ &+ \alpha^1_1{\otimes}\beta^1_2{\otimes}c^1_2 + \alpha^1_2{\otimes}\beta^2_2{\otimes}c^1_2 + \alpha^2_1{\otimes}\beta^1_2{\otimes}c^2_2 + \alpha^2_2{\otimes}\beta^2_2{\otimes}c^2_2. \end{aligned} \tag{2.4.3}$$

Strassen's algorithm is

$$
\begin{aligned}
M_{2,2,2} =&(\alpha^1_1+\alpha^2_2)\otimes(\beta^1_1+\beta^2_2)\otimes(c^1_1+c^2_2)\\
&+(\alpha^2_1+\alpha^2_2)\otimes\beta^1_1\otimes(c^2_1-c^2_2)\\
&+\alpha^1_1\otimes(\beta^1_2-\beta^2_2)\otimes(c^1_2+c^2_2)\\
&+\alpha^2_2\otimes(-\beta^1_1+\beta^2_1)\otimes(c^2_1+c^1_1)\\
&+(\alpha^1_1+\alpha^1_2)\otimes\beta^2_2\otimes(-c^1_1+c^1_2)\\
&+(-\alpha^1_1+\alpha^2_1)\otimes(\beta^1_1+\beta^1_2)\otimes c^2_2\\
&+(\alpha^1_2-\alpha^2_2)\otimes(\beta^2_1+\beta^2_2)\otimes c^1_1.
\end{aligned}
\tag{2.4.4}
$$

Exercise 2.4.4.1: Verify that (2.4.4) and (2.4.3) are indeed the same tensors.

Remark 2.4.4.2. To present Strassen's algorithm this way, solve for the coefficients of the vector equation set each Roman numeral in (1.1.1) to a linear combination of the c^i_j and set the sum of the terms equal to (2.4.3).

Strassen's algorithm for matrix multiplication using seven multiplications is far from unique. Let $t \in \mathbb{R}$ be a constant. One also has

$$
\begin{aligned}
M_{2,2,2} =&(\alpha^1_1+\alpha^2_2+t\alpha^2_1)\otimes(\beta^1_1-t\beta^2_1+\beta^2_2)\otimes(c^1_1+c^2_2)\\
&+(\alpha^2_1+\alpha^2_2+t\alpha^2_1)\otimes(\beta^1_1-t\beta^2_1)\otimes(c^2_1-c^2_2)\\
&+\alpha^1_1\otimes(\beta^1_2-t\beta^2_2-\beta^2_2)\otimes(c^1_2+c^2_2)\\
&+(\alpha^2_2+t\alpha^2_1)\otimes(-\beta^1_1+t\beta^2_1+\beta^2_1)\otimes(c^2_1+c^1_1)\\
&+(\alpha^1_1+\alpha^1_2+t\alpha^1_1)\otimes\beta^2_2\otimes(-c^1_1+c^1_2)\\
&+(-\alpha^1_1+\alpha^2_1)\otimes(\beta^1_1-t\beta^2_1+\beta^1_2-t\beta^2_2)\otimes c^2_2\\
&+(\alpha^1_2+t\alpha^1_1-\alpha^2_2-t\alpha^2_1)\otimes(\beta^2_1+\beta^2_2)\otimes c^1_1.
\end{aligned}
\tag{2.4.5}
$$

In fact there is a nine parameter family of algorithms for $M_{2,2,2}$ each using seven multiplications. The geometry of this family is explained in §2.5.

An expression of a bilinear map $T \in V_1^*\otimes V_2^*\otimes V_3$ as a sum of rank one elements may be thought of as an algorithm for executing it. The number of rank one elements in the expression is the number of multiplications needed to execute the algorithm. The rank of the tensor T therefore gives an upper bound on the number of multiplications needed to execute the corresponding bilinear map using a best possible algorithm.

2.4.5. Border rank of a tensor. Chapter 4 is dedicated to the study of algebraic varieties, which are the zero sets of polynomials. In particular, if a sequence of points is in the zero set of a collection of polynomials, any limit point for the sequence must be in the zero set. In our study of tensors of a given rank r, we will also study limits of such tensors.

Consider the tensor

$$T = a_1 \otimes b_1 \otimes c_1 + a_1 \otimes b_1 \otimes c_2 + a_1 \otimes b_2 \otimes c_1 + a_2 \otimes b_1 \otimes c_1. \tag{2.4.6}$$

One can show that the rank of T is three, but T can be approximated as closely as one likes by tensors of rank two, as consider:

$$T(\epsilon) = \frac{1}{\epsilon}[(\epsilon - 1)a_1 \otimes b_1 \otimes c_1 + (a_1 + \epsilon a_2) \otimes (b_1 + \epsilon b_2) \otimes (c_1 + \epsilon c_2)] \tag{2.4.7}$$

Definition 2.4.5.1. A tensor T has *border rank* r if it is a limit of tensors of rank r but is not a limit of tensors of rank s for any $s < r$. Let $\underline{\mathbf{R}}(T)$ denote the border rank of T.

Note that $\mathbf{R}(T) \geq \underline{\mathbf{R}}(T)$.

For example, the sequence (2.4.7) shows that T of (2.4.6) has border rank at most two, and it is not hard to see that its border rank is exactly two.

The border rank admits an elegant geometric interpretation which I discuss in detail in §5.1. Intuitively, $T(\epsilon)$ is a point on the line spanned by the two tensors $a_1 \otimes b_1 \otimes c_1$ and $z(\epsilon) := (a_1 + \epsilon a_2) \otimes (b_1 + \epsilon b_2) \otimes (c_1 + \epsilon c_2)$ inside the set of rank one tensors. Draw $z(\epsilon)$ as a curve, for $\epsilon \neq 0$, $T(\epsilon)$ is a point on the secant line through $z(0)$ and $z(\epsilon)$, and in the limit, one obtains a point on the tangent line to $z(0) = a_1 \otimes b_1 \otimes c_1$

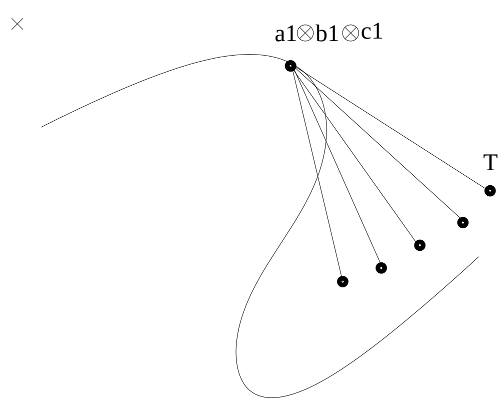


Figure 2.4.1. Unlabeled points are various T_ϵ's lying on secant lines to the curve

An especially important question is: *What is the border rank of the matrix multiplication operator?* All that is known is that $\underline{\mathbf{R}}(M_{2\times 2}) = 7$, that $14 \leq \underline{\mathbf{R}}(M_{3\times 3}) \leq 21$, and $\frac{5}{2}m^2 - 3m \leq \underline{\mathbf{R}}(M_{m\times m}) \leq m^{2.38}$, see Chapter 11.

2.5. Examples of invariant tensors

Certain tensors, viewed as multilinear maps, commute with the action of the group of changes of bases, i.e., as tensors, they are *invariant* with respect to the group action. Matrix multiplication is one such, as I explain below.

2.5.1. Contractions of tensors. There is a natural bilinear map

$$\mathrm{Con} : (V_1 \otimes \cdots \otimes V_k) \times (V_k^* \otimes U_1 \otimes \cdots \otimes U_m) \to V_1 \otimes \cdots \otimes V_{k-1} \otimes U_1 \otimes \cdots \otimes U_m$$

given by $(v_1 \otimes \cdots \otimes v_k, \alpha \otimes b_1 \otimes \cdots \otimes b_m) \to \alpha(v_k) v_1 \otimes \cdots \otimes v_{k-1} \otimes b_1 \otimes \cdots \otimes b_m$, is called a *contraction*. One can view the contraction operator Con as an element of

$$(V_1 \otimes \cdots \otimes V_k)^* \otimes (V_k^* \otimes U_1 \otimes \cdots \otimes U_m)^* \otimes (V_1 \otimes \cdots \otimes V_{k-1} \otimes U_1 \otimes \cdots \otimes U_m).$$

If $T \in V_1 \otimes \cdots \otimes V_k$ and $S \in U_1 \otimes \cdots \otimes U_m$, and for some fixed i, j there is an identification $V_i \simeq U_j^*$, one may contract $T \otimes S$ to obtain an element of $V_{\hat{i}} \otimes U_{\hat{j}}$ which is sometimes called the (i,j)*-mode product of* T *and* S.

Exercise 2.5.1.1: Show that if $f : V \to V$ is a linear map, i.e., $f \in V^* \otimes V$, then the trace of f corresponds to Con above.

In other words (recalling the convention that repeated indices are to be summed over) Exercise 2.5.1.1 says:

Con*,* Id*, and* trace *are all the same tensors. If* (a_i) *is a basis of* A *with dual basis* (α^i)*, then they all correspond to the tensor* $\alpha^i \otimes a_i$*.*

2.5.2. Matrix multiplication as a tensor. Let A, B, and C be vector spaces of dimensions $\mathbf{a}, \mathbf{b}, \mathbf{c}$, and consider the matrix multiplication operator $M_{\mathbf{a},\mathbf{b},\mathbf{c}}$ that composes a linear map from A to B with a linear map from B to C to obtain a linear map from A to C. Let $V_1 = A^* \otimes B$, $V_2 = B^* \otimes C$, $V_3 = A^* \otimes C$, so $M_{\mathbf{a},\mathbf{b},\mathbf{c}} \in V_1^* \otimes V_2^* \otimes V_3$. On rank one elements:

$$\begin{aligned} M_{\mathbf{a},\mathbf{b},\mathbf{c}} : (A^* \otimes B) \times (B^* \otimes C) &\to A^* \otimes C \\ (\alpha \otimes b) \times (\beta \otimes c) &\mapsto \beta(b) \alpha \otimes c \end{aligned} \tag{2.5.1}$$

and $M_{\mathbf{a},\mathbf{b},\mathbf{c}}$ is defined on all elements by extending the definition on rank one elements bi-linearly. In other words, as a tensor,

$$\begin{aligned} M_{\mathbf{a},\mathbf{b},\mathbf{c}} = \mathrm{Id}_A \otimes \mathrm{Id}_B \otimes \mathrm{Id}_C &\in (A^* \otimes B)^* \otimes (B^* \otimes C)^* \otimes (A^* \otimes C) \\ &= A \otimes B^* \otimes B \otimes C^* \otimes A^* \otimes C \end{aligned}$$

and it is clear from the expression (2.4.3), that it may be viewed as any of the three possible contractions. In other words, as a bilinear map $A^* \otimes B \times B^* \otimes C \to A^* \otimes C$, or as a bilinear map $A \otimes C^* \times B^* \otimes C \to A \otimes B^*$ or $A \otimes C^* \times A^* \otimes B \to C^* \otimes B$. When $A = B = C$, this gives rise to a symmetry under the action of the group of permutations on three elements $\mathfrak{S}_3$ which is often exploited in the study of the operator.

Exercise 2.5.2.1: Show that $M_{\mathbf{a},\mathbf{b},\mathbf{c}}$ viewed as a trilinear form in bases takes a triple of matrices (X, Y, Z) to $\mathrm{trace}(XYZ)$, and hence is invariant under changes in bases in A, B and C. The nine parameter family of algorithms for $M_{2,2,2}$ is the action of $SL(A)\times SL(B)\times SL(C)$ on the expression. (The action of the scalars times the identity will not effect the expression meaningfully as we identify $\lambda v{\otimes}w = v{\otimes}(\lambda w)$ for a scalar λ.)

Remark 2.5.2.2. The above exercise gives rise a nine parameter family of expressions for $M_{2,2,2}$ as a sum of seven rank one tensors. One could ask if there are any other expressions for $M_{2,2,2}$ as a sum of seven rank one tensors. In [**105**] it is shown that there are no other such expressions.

2.5.3. Another $GL(V)$-invariant tensor. Recall from above that as tensors Con, tr and Id_V are the same. In Chapter 6 we will see Id_V and its scalar multiples are the only $GL(V)$-invariant tensors in $V{\otimes}V^*$. The space $V{\otimes}V{\otimes}V^*{\otimes}V^* = \mathrm{End}(V{\otimes}V)$, in addition to the identity map $\mathrm{Id}_{V\otimes V}$, has another $GL(V)$-invariant tensor. As a linear map it is simply

$$\begin{aligned} \sigma : V{\otimes}V &\to V{\otimes}V \\ a{\otimes}b &\mapsto b{\otimes}a. \end{aligned} \tag{2.5.2}$$

2.6. Symmetric and skew-symmetric tensors

2.6.1. The spaces S^2V and Λ^2V. Recall the map (2.5.2), $\sigma : V^*{\otimes}V^* \to V^*{\otimes}V^*$. (Note that here we look at it on the dual space.) Consider $V^{\otimes 2} = V{\otimes}V$ with basis $\{v_i{\otimes}v_j,\ 1 \le i, j \le n\}$. The subspaces defined by

$$\begin{aligned} S^2V &:= \mathrm{span}\{v_i{\otimes}v_j + v_j{\otimes}v_i,\ 1 \le i, j \le n\} \\ &= \mathrm{span}\{v{\otimes}v \,|v \in V\} \\ &= \{X \in V{\otimes}V \mid X(\alpha,\beta) = X(\beta,\alpha)\ \forall \alpha,\beta \in V^*\} \\ &= \{X \in V{\otimes}V \mid X\circ\sigma = X\} \end{aligned}$$

$$\begin{aligned} \Lambda^2V &:= \mathrm{span}\{v_i{\otimes}v_j - v_j{\otimes}v_i,\ 1 \le i, j \le n\} \\ &= \mathrm{span}\{v{\otimes}w - w{\otimes}v \,|v, w \in V\}, \\ &= \{X \in V{\otimes}V \mid X(\alpha,\beta) = -X(\beta,\alpha)\ \forall \alpha,\beta \in V^*\} \\ &= \{X \in V{\otimes}V \mid X\circ\sigma = -X\} \end{aligned}$$

are respectively the spaces of *symmetric* and *skew-symmetric* 2-tensors of V. In the fourth lines we are considering X as a map $V^*{\otimes}V^* \to \mathbb{C}$. The second description of these spaces implies that if $T \in S^2V$ and $g \in GL(V)$, then (using (2.4.2)) $g\cdot T \in S^2V$ and similarly for $\Lambda^2 V$. That is, they are invariant under linear changes of coordinates, i.e., they are $GL(V)$-submodules of $V^{\otimes 2}$.

For $v_1, v_2 \in V$, define $v_1v_2 := \frac{1}{2}(v_1{\otimes}v_2 + v_2{\otimes}v_1) \in S^2V$ and $v_1 \wedge v_2 := \frac{1}{2}(v_1{\otimes}v_2 - v_2{\otimes}v_1) \in \Lambda^2V$.

2.6.2. Exercises.

(1) Show that the four descriptions of S^2V all agree. Do the same for the four descriptions of Λ^2V.

(2) Show that

$$V{\otimes}V = S^2V \oplus \Lambda^2 V. \tag{2.6.1}$$

By the remarks above, this direct sum decomposition is invariant under the action of $GL(V)$, c.f., Exercise 2.1.12. One says $V^{\otimes 2}$ *decomposes* as a $GL(V)$ module to $\Lambda^2V \oplus S^2V$.

(3) Show that the action of GL_2 on $\mathbb{C}^3$ of Example 2.2.1.2.5 is the action induced on $S^2\mathbb{C}^2$ from the action on $\mathbb{C}^2{\otimes}\mathbb{C}^2$.

(4) Show that no proper linear subspace of S^2V is invariant under the action of $GL(V)$; i.e., S^2V is an irreducible submodule of $V^{\otimes 2}$.

(5) Show that Λ^2V is an irreducible $GL(V)$-submodule of $V^{\otimes 2}$.

(6) Define maps

$$\begin{aligned} \pi_S : V^{\otimes 2} &\to V^{\otimes 2} \\ X &\mapsto \frac{1}{2}(X + X\circ\sigma) \end{aligned} \tag{2.6.2}$$

$$\begin{aligned} \pi_\Lambda : V^{\otimes 2} &\to V^{\otimes 2} \\ X &\mapsto \frac{1}{2}(X - X\circ\sigma) \end{aligned} \tag{2.6.3}$$

Show $\pi_S(V^{\otimes 2}) = S^2V$ and $\pi_\Lambda(V^{\otimes 2}) = \Lambda^2V$.

(7) What is $\ker \pi_S$?

Notational Warning. Above I used $\circ$ as composition. It is also used in the literature to denote symmetric product as defined below. To avoid confusion I reserve $\circ$ for composition of maps with the exception of taking the symmetric product of spaces, e.g., $S^dV \circ S^\delta V = S^{d+\delta}V$.

2.6.3. Symmetric tensors S^dV. Let $\pi_S : V^{\otimes d} \to V^{\otimes d}$ be the map defined on rank one elements by

$$\pi_S(v_1{\otimes}\cdots{\otimes}\, v_d) = \frac{1}{d!}\sum_{\sigma\in\mathfrak{S}_d} v_{\sigma(1)}{\otimes}\cdots{\otimes}v_{\sigma(d)}$$

where $\mathfrak{S}_d$ denotes the group of permutations of d elements.

Exercises 2.6.3.1:

1. Show that π_S agrees with (2.6.2) when $d = 2$.
2. Show that $\pi_S(\pi_S(X)) = \pi_S(X)$, i.e., that π_S is a projection operator (c.f. Exercise 2.3.2.9).

Introduce the notation $v_1 v_2 \cdots v_d := \pi_S(v_1 \otimes v_2 \otimes \cdots \otimes v_d)$.

Definition 2.6.3.2. Define

$$S^d V := \pi_S(V^{\otimes d})$$

the *d-th symmetric power of V.*

Note

$$\begin{aligned} S^d V &= \{X \in V^{\otimes d} \mid \pi_S(X) = X\} \\ &= \{X \in V^{\otimes d} \mid X \circ \sigma = X \ \forall \sigma \in \mathfrak{S}_d\}. \end{aligned} \tag{2.6.4}$$

Exercise 2.6.3.3: Show that in bases, if $u \in S^p\mathbb{C}^r$, $v \in S^q\mathbb{C}^r$, the symmetric tensor product $uv \in S^{p+q}\mathbb{C}^r$ is

$$(uv)^{i_1,\dots,i_{p+q}} = \frac{1}{(p+q)!}\sum u^I v^J$$

where the summation is over $I = i_1, \dots, i_p$ with $i_1 \le \cdots \le i_p$ and analogously for J.

Exercise 2.6.3.4: Show that $S^dV \subset V^{\otimes d}$ is invariant under the action of $GL(V)$.

Exercise 2.6.3.5: Show that if $e_1, \dots, e_{\mathbf{v}}$ is a basis of V, then $e_{j_1}e_{j_2}\cdots e_{j_d}$, for $1 \le j_1 \le \cdots \le j_d \le \mathbf{v}$ is a basis of S^dV. Conclude that $\dim S^d\mathbb{C}^{\mathbf{v}} = \binom{\mathbf{v}+d-1}{d}$.

2.6.4. S^kV^* as the space of homogeneous polynomials of degree k on V. The space S^kV^* was defined as the space of symmetric k-linear forms on V. It may also be considered as the space of homogeneous polynomials of degree k on V. Namely given a multi-linear form $\overline{Q}$, the map $x \mapsto \overline{Q}(x, \dots, x)$ is a polynomial mapping of degree k. The process of passing from a homogeneous polynomial to a multi-linear form is called *polarization.* For example, if Q is a homogeneous polynomial of degree two on V, define the bilinear form $\overline{Q}$ by the equation

$$\overline{Q}(x, y) = \frac{1}{2}[Q(x + y) - Q(x) - Q(y)].$$

For general symmetric multilinear forms, the polarization identity is

$$\overline{Q}(x_1,\dots,x_k)=\frac{1}{k!}\sum_{I\subset[k],I\neq\emptyset}(-1)^{k-|I|}Q\left(\sum_{i\in I}x_i\right). \tag{2.6.5}$$

Here $[k]=\{1,\dots,k\}$. Since Q and $\overline{Q}$ are really the same object, I generally will not distinguish them by different notation.

Example 2.6.4.1. For a cubic polynomial in two variables $P(s,t)$, one obtains the cubic form

$$\begin{aligned}\overline{P}((s_1,t_1),(s_2,t_2),(s_3,t_3)) =& \frac{1}{6}[P(s_1+s_2+s_3,t_1+t_2+t_3)\\ &-P(s_1+s_2,t_1+t_2)-P(s_1+s_3,t_1+t_3)-P(s_2+s_3,t_2+t_3)\\ &+P(s_1,t_1)+P(s_2,t_2)+P(s_3,t_3)\end{aligned}$$

so for, e.g., $P=s^2t$ on obtains $\overline{P}=\frac{1}{3}(s_1s_2t_3+s_1s_3t_2+s_2s_3t_1)$.

From this perspective, the contraction map is

$$\begin{aligned}V^*\times S^dV&\to S^{d-1}V\\ (\alpha,P)&\mapsto P(\alpha,\cdot)\end{aligned} \tag{2.6.6}$$

if one fixes α, this is just the partial derivative of P in the direction of α.

Exercise 2.6.4.2: Prove the above assertion by choosing coordinates and taking $\alpha=x_1$.

2.6.5. Polynomials and homogeneous polynomials. In this book I generally restrict the study of polynomials to homogeneous polynomials - this is no loss of generality, as there is a bijective map

$$S^d\mathbb{C}^m\to\{\text{polynomials of degree at most } d \text{ in } m-1 \text{ variables}\}$$

by setting $x_m=1$, i.e., let $I=(i_i,\dots,i_m)$ be a multi-index and write $|I|=i_1+\cdots+i_m$,

$$\sum_{|I|=d}a_{i_1,\dots,i_m}x_1^{i_1}\cdots x_{m-1}^{i_{m-1}}x_m^{i_m}\mapsto\sum_{|I|=d}a_{i_1,\dots,i_m}x_1^{i_1}\cdots x_{m-1}^{i_{m-1}}.$$

2.6.6. Symmetric tensor rank.

Definition 2.6.6.1. Given $\phi\in S^dV$, define the *symmetric tensor rank* of ϕ, $\mathbf{R}_S(\phi)$, to be the smallest r such that $\phi=v_1^d+\cdots+v_r^d$ for some $v_j\in V$. Define the *symmetric tensor border rank* of ϕ, $\underline{\mathbf{R}}_S(\phi)$ to be the smallest r such that ϕ is a limit of symmetric tensors of symmetric tensor rank r.

Exercise 2.6.6.2: Show that for any $\phi\in S^d\mathbb{C}^n$, $\mathbf{R}_S(\phi)\leq\binom{n+d-1}{d}$. ⊚

There is a natural inclusion $S^dV \subset S^sV \otimes S^{d-s}V$ given by partial polarization. Write $\phi_{s,d-s} \in S^sV \otimes S^{d-s}V$ for the image of $\phi \in S^dV$. Thinking of $S^sV \otimes S^{d-s}V$ as a space of linear maps $S^sV^* \to S^{d-s}V$, $\phi_{s,d-s}(\alpha_1 \cdots \alpha_s) = \overline{\phi}(\alpha_1, \ldots, \alpha_s, \cdot, \ldots, \cdot)$.

Exercise 2.6.6.3: Show that if $\underline{\mathbf{R}}_S(\phi) \leq k$, then $\operatorname{rank}(\phi_{s,d-s}) \leq k$ for all s.

Remark 2.6.6.4. Exercise 2.6.6.3 provides a test for symmetric tensor border rank that dates back to Macaulay [**221**].

Exercise 2.6.6.5: Considering $S^dV \subset V^{\otimes d}$, show that, for $\phi \in S^dV \subset V^{\otimes d}$, $\mathbf{R}_S(\phi) \geq \mathbf{R}(\phi)$ and $\underline{\mathbf{R}}_S(\phi) \geq \underline{\mathbf{R}}(\phi)$. P. Comon has conjectured that equality holds, see [**99**, §4.1] and §5.7.2.

More generally one can define the *partially symmetric rank* of partially symmetric tensors. We will not dwell much on this since this notion will be superceded by the notion of X-rank in Chapter 5. The term INDSCAL is used for the partially symmetric rank of elements of $S^2W \otimes V$.

2.6.7. Alternating tensors.

Define a map

$$\pi_\Lambda : V^{\otimes k} \to V^{\otimes k} \tag{2.6.7}$$

$$v_1 \otimes \cdots \otimes v_k \mapsto v_1 \wedge \cdots \wedge v_k := \frac{1}{k!} \sum_{\sigma \in \mathfrak{S}_k} (\operatorname{sgn}(\sigma)) v_{\sigma(1)} \otimes \cdots \otimes v_{\sigma(k)} \tag{2.6.8}$$

where $\operatorname{sgn}(\sigma) = \pm 1$ denotes the sign of the permutation σ (see Remark 2.9.2.3). Denote the image by $\Lambda^k V$, called the space of *alternating k-tensors* and note that it agrees with our previous definition of $\Lambda^2 V$ when $k = 2$.

In particular

$$\Lambda^k V = \{X \in V^{\otimes k} \mid X \circ \sigma = \operatorname{sgn}(\sigma) X \ \forall \sigma \in \mathfrak{S}_k\}.$$

Exercise 2.6.7.1: Show that $v_1 \wedge \cdots \wedge v_k = 0$ if and only if $v_1, \ldots, v_k$ are linearly dependent.⊚

2.6.8. Tensor, symmetric and exterior algebras.

Definition 2.6.8.1. For a vector space V, define $V^{\otimes} := \oplus_{k \geq 0} V^{\otimes k}$, the *tensor algebra of V*. The multiplication is given by defining the product of $v_1 \otimes \cdots \otimes v_s$ with $w_1 \otimes \cdots w_t$ to be $v_1 \otimes \cdots \otimes v_s \otimes w_1 \otimes \cdots w_t$ and extending linearly.

Definition 2.6.8.2. Define the *exterior algebra* $\Lambda^\bullet V = \oplus_k \Lambda^k V$ and the *symmetric algebra* $S^\bullet V := \oplus_d S^d V$ by defining the multiplications $\alpha \wedge \beta := \pi_\Lambda(\alpha \otimes \beta)$ for $\alpha \in \Lambda^s V$, $\beta \in \Lambda^t V$ and $\alpha\beta := \pi_S(\alpha \otimes \beta)$ for $\alpha \in S^s V$, $\beta \in S^t V$.

Note the following (i) $\Lambda^1 V = S^1 V = V$, (ii) the multiplication $S^s V \times S^t V \to S^{s+t} V$, when considering $S^k V$ as the space of homogeneous polynomials on V^*, corresponds to the multiplication of polynomials and (iii) these are both associative algebras with units respectively $1 \in S^0 V$, $1 \in \Lambda^0 V$.

2.6.9. Contractions preserve symmetric and skew-symmetric tensors. Recall (§2.5.1) the contraction

$$V^* \times V^{\otimes k} \to V^{\otimes k-1} \tag{2.6.9}$$
$$(\alpha, v_1 \otimes \cdots \otimes v_k) \mapsto \alpha(v_1) v_2 \otimes \cdots \otimes v_k.$$

Here we could have just as well defined contractions on any of the factors. This contraction preserves the subspaces of symmetric and skew-symmetric tensors, as you verify in Exercise 2.6.10.(2.6.13).

Remark 2.6.9.1. The *first fundamental theorem of invariant theory* (see, e.g., [**264**, p. 388]) states that the only $GL(V)$-invariant operators are of the form (2.6.9), and the only $SL(V)$-invariant operators are these and contractions with the volume form. (Here $SL(V)$ is the group of invertible endomorphisms of determinant one, see Exercises 2.6.12 and 2.6.13.)

For a pairing $V^* \times V \otimes W \to W$, I sometimes let $\alpha \lrcorner T$ denote the contraction of $\alpha \in V^*$ and $T \in V \otimes W$

2.6.10. Exercises.

(1) Show that the subspace $\Lambda^k V \subset V^{\otimes k}$ is invariant under the action of $GL(V)$.

(2) Show a basis of V induces a basis of $\Lambda^k V$. Using this induced basis, show that, if $\dim V = \mathbf{v}$, then $\dim \Lambda^k V = \binom{\mathbf{v}}{k}$. In particular, $\Lambda^{\mathbf{v}} V \simeq \mathbb{C}$, $\Lambda^l V = 0$ for $l > \mathbf{v}$ and, $S^3 V \oplus \Lambda^3 V \neq V^{\otimes 3}$ when $\mathbf{v} > 1$.

(3) Calculate, for $\alpha \in V^*$, $\alpha \lrcorner (v_1 \cdots v_k)$ explicitly and show that it indeed is an element of $S^{k-1} V$, and similarly for $\alpha \lrcorner (v_1 \wedge \cdots \wedge v_k)$.

(4) Show that the composition $(\alpha \lrcorner) \circ (\alpha \lrcorner) : \Lambda^k V \to \Lambda^{k-2} V$ is the zero map.

(5) Show that if $V = A \oplus B$ then there is an induced direct sum decomposition $\Lambda^k V = \Lambda^k A \oplus (\Lambda^{k-1} A \otimes \Lambda^1 B) \oplus (\Lambda^{k-2} A \otimes \Lambda^2 B) \oplus \cdots \oplus \Lambda^k B$ as a $GL(A) \times GL(B)$-module.

(6) Show that a subspace $A \subset V$ determines a well defined induced filtration of $\Lambda^k V$ given by $\Lambda^k A \subset \Lambda^{k-1} A \otimes \Lambda^1 V \subset \Lambda^{k-2} A \otimes \Lambda^2 V \subset \cdots \subset \Lambda^k V$. If $P_A := \{g \in GL(V) \mid g \cdot v \in A \ \forall v \in A\}$, then each filtrand is a P_A-submodule.

(7) Show that if V is equipped with a *volume form*, i.e. an nonzero element $\phi \in \Lambda^{\mathbf{v}} V$, then one obtains an identification $\Lambda^k V \simeq \Lambda^{\mathbf{v}-k} V^*$. ⊚

(8) Show that $V^* \simeq \Lambda^{\mathbf{v}-1} V \otimes \Lambda^{\mathbf{v}} V^*$ as $GL(V)$-modules.

2.6.11. Induced linear maps. Tensor product and the symmetric and skew-symmetric constructions are *functorial*. This essentially means: given a linear map $f : V \to W$ there are induced linear maps $f^{\otimes k} : V^{\otimes k} \to W^{\otimes k}$ given by $f^{\otimes k}(v_1 \otimes \cdots \otimes v_k) = f(v_1) \otimes \cdots \otimes f(v_k)$. These restrict to give well defined maps $f^{\wedge k} : \Lambda^k V \to \Lambda^k W$ and $f^{\circ k} : S^k V \to S^k W$.

Definition 2.6.11.1. Given a linear map $f : V \to V$, the induced map $f^{\wedge \mathbf{v}} : \Lambda^{\mathbf{v}} V \to \Lambda^{\mathbf{v}} V$ is called the *determinant* of f.

Example 2.6.11.2. Let $\mathbb{C}^2$ have basis e_1, e_2. Say $f : \mathbb{C}^2 \to \mathbb{C}^2$ is represented by the matrix $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ with respect to this basis, i.e., $f(e_1) = ae_1 + be_2$, $f(e_2) = ce_1 + de_2$. Then

$$\begin{aligned} f(e_1 \wedge e_2) &= (ae_1 + be_2) \wedge (ce_1 + de_2) \\ &= (ad - bc) e_1 \wedge e_2 \end{aligned}$$

Sometimes one fixes a given element of $\Lambda^{\mathbf{v}} V^*$ and calls it the determinant *det*. Which use of the word I am using should be clear from the context.

To see these induced maps explicitly, let $\alpha^1, \dots, \alpha^{\mathbf{v}}$ be a basis of V^* and $w_1, \dots, w_{\mathbf{w}}$ a basis of W. Write $f = f^s_j \alpha^j \otimes w_s$ in this basis. Then

(2.6.10)
$$f^{\otimes 2} = f^s_j f^t_k \alpha^j \otimes \alpha^k \otimes w_s \otimes w_t$$

(2.6.11)
$$f^{\wedge 2} = (f^s_j f^t_k - f^s_k f^t_j)(\alpha^j \wedge \alpha^k) \otimes (w_s \wedge w_t)$$

(2.6.12)
$$f^{\wedge p} = \sum_{\sigma \in \mathfrak{S}_p} \operatorname{sgn}(\sigma) f^{s_1}_{i_{\sigma(1)}} \cdots f^{s_p}_{i_{\sigma(p)}} \alpha^{i_{\sigma(1)}} \wedge \cdots \wedge \alpha^{i_{\sigma(p)}} \otimes w_{s_1} \wedge \cdots \wedge w_{s_p}.$$

The $GL(V)$-module isomorphism from Exercise 2.6.10.8 shows that a linear map $\phi : \Lambda^{\mathbf{v}-1} V \to \Lambda^{\mathbf{v}-1} W$, where $\dim V = \dim W = \mathbf{v}$ induces a linear map $V^* \otimes \Lambda^{\mathbf{v}} V \to W^* \otimes \Lambda^{\mathbf{v}} W$, i.e., a linear map $W \otimes \Lambda^{\mathbf{v}} W^* \to V \otimes \Lambda^{\mathbf{v}} V^*$. If $\phi = f^{\wedge(\mathbf{v}-1)}$ for some linear map $f : V \to W$, and f is invertible, then the induced linear map is $f^{-1} \otimes \det(f)$. If f is not invertible, then $f^{\wedge(\mathbf{v}-1)}$ has rank one. If $\operatorname{rank}(f) \leq \mathbf{v} - 2$, then $f^{\wedge(\mathbf{v}-1)}$ is zero. An advantage of $f^{\wedge \mathbf{v}-1}$ over f^{-1} is that it is defined even if f is not invertible, which is exploited in §7.6.

2.6.12. Exercises on induced linear maps and the determinant.

(1) Verify that if f has rank $\mathbf{v}-1$, then $f^{\wedge(\mathbf{v}-1)}$ has rank one, and if $\mathrm{rank}(f) \leq \mathbf{v}-2$, then $f^{\wedge(\mathbf{v}-1)}$ is zero. ⊚

(2) Show more generally that if f has rank r, then $\mathrm{rank}(f^{\wedge s}) = \binom{r}{s}$.

(3) Show that the eigenvalues of $f^{\wedge k}$ are the k-th elementary symmetric functions of the eigenvalues of f.

(4) Given $f : V \to V$, $f^{\wedge \mathbf{v}}$ is a map from a one-dimensional vector space to itself, and thus multiplication by some scalar. Show that if one chooses a basis for V and represents f by a matrix, the scalar representing $f^{\wedge \mathbf{v}}$ is the determinant of the matrix representing f.

(5) Assume $W = V$ and that V admits a basis of eigenvectors for f. Show that $\Lambda^k V$ admits a basis of eigenvectors for $f^{\wedge k}$ and find the eigenvectors and eigenvalues for $f^{\wedge k}$ in terms of those for f. In particular show that the k-th coefficient of the characteristic polynomial of f is $(-1)^k \mathrm{trace}\,(f^{\wedge k})$ where trace is defined in Exercise 2.3.2.7.

(6) Let $f : V \to W$ be invertible, with $\dim V = \dim W = \mathbf{v}$. Verify that $f^{\wedge \mathbf{v}-1} = f^{-1} \otimes \det(f)$ as asserted above.

(7) Fix $det \in \Lambda^{\mathbf{v}} V^*$. Let

$$SL(V) := \{g \in GL(V) \mid g \cdot det = det\}. \tag{2.6.13}$$

Show that $SL(V)$ is a group, called the *Special Linear group*. Show that if one fixes a basis of V^* $\alpha^1, \dots, \alpha^{\mathbf{v}}$ such that $det = \alpha^1 \wedge \cdots \wedge \alpha^{\mathbf{v}}$, and uses this basis and its dual to express linear maps $V \to V$ as $\mathbf{v} \times \mathbf{v}$ matrices, that $SL(V)$ becomes the set of matrices with determinant one (where one takes the usual determinant of matrices).

(8) Given n-dimensional vector spaces E, F, fix an element $\Omega \in \Lambda^n E^* \otimes \Lambda^n F$. Since $\dim(\Lambda^n E^* \otimes \Lambda^n F) = 1$, Ω is unique up to scale. Then given a linear map $f : V \to W$, one may write $f^{\wedge n} = c_f \Omega$, for some constant c_f. Show that if one chooses bases $e_1, \dots, e_n$ of E, $f_1, \dots, f_n$ of F such that $\Omega = e_1 \wedge \cdots \wedge e_n \otimes f_1 \wedge \cdots \wedge f_n$, and expresses f as a matrix M_f with respect to these bases, then $c_f = \det(M_f)$.

(9) Note that Ω determines a vector $\Omega^* \in \Lambda^n E \otimes \Lambda^n F^*$ by $\langle \Omega^*, \Omega \rangle = 1$. Recall that $f : V \to W$ determines a linear map $f^T : W^* \to V^*$. Use Ω^* to define $\det_{f^T}$. Show $\det_f = \det_{f^T}$.

(10) If $E = F$ then a volume form is not needed to define $\det_f$. Show that in this case $\det_f$ is the product of the eigenvalues of f.

2.6.13. The group $\mathfrak{S}_d$ acts on $V^{\otimes d}$. Let $\mathfrak{S}_d$ denote the symmetric group on d elements (see Definition 2.9.2.2). $\mathfrak{S}_d$ acts on $V^{\otimes d}$ by, for $v_1, \dots, v_d \in V$,

$$\sigma(v_1 \otimes \cdots \otimes v_d) = v_{\sigma(1)} \otimes \cdots \otimes v_{\sigma(d)}$$

and extending linearly. Looking at symmetric tensors from this perspective yields:

$$S^dV = \{T \in V^{\otimes d} \mid \sigma \cdot T = T \ \forall \ \sigma \in \mathfrak{S}_d\}.$$

This description is slightly different from (2.6.4) as before elements of $\mathfrak{S}_d$ acted on $V^{*\otimes d}$ and it was not explicitly mentioned that the elements were part of an $\mathfrak{S}_d$ action. In words:

S^dV is the subspace of $V^{\otimes d}$ whose elements are invariant under the action of $\mathfrak{S}_d$.

Exercise 2.6.13.1: Show that if $g \in GL(V)$ and $\sigma \in \mathfrak{S}_d$, then, for all $T \in V^{\otimes d}$, $g \cdot \sigma \cdot T = \sigma \cdot g \cdot T$.

Because it will be important in Chapter 6, I record the result of Exercise 2.6.13.1:

The actions of $GL(V)$ and $\mathfrak{S}_d$ on $V^{\otimes d}$ commute with each other.

2.7. Polynomials on the space of matrices

Consider homogeneous polynomials on the space of $\mathbf{a} \times \mathbf{b}$ matrices. We will be interested in how the polynomials change under changes of bases in $\mathbb{C}^{\mathbf{a}}$ and $\mathbb{C}^{\mathbf{b}}$.

More invariantly, we will study our polynomials as $GL(A) \times GL(B)$-modules. Let $V = A \otimes B$. We study the degree d homogeneous polynomials on $A^* \otimes B^*$, S^dV, as a $G := GL(A) \times GL(B)$-module.

Warning to the reader: I am identifying our base vector space of matrices $\mathbb{C}^{\mathbf{a}} \otimes \mathbb{C}^{\mathbf{b}}$ with $A^* \otimes B^*$, in order to minimize the use of $*$'s below.

Remark . In the examples that follow, the reader may wonder why I am belaboring such familiar polynomials. The reason is that later on we will encounter natural polynomials that are unfamiliar and will need ways of writing them down explicitly from an invariant description.

2.7.1. Quadratic polynomials. While S^2V is irreducible as a $GL(V)$-module, if $V = A \otimes B$, one expects to be able to decompose it further as a $G = GL(A) \times GL(B)$-module. One way to get an element of $S^2(A \otimes B)$ is simply to multiply an element of S^2A with an element of S^2B, i.e., given $\alpha \in S^2A$, $\beta \in S^2B$, $\alpha \otimes \beta$ is defined by $\alpha \otimes \beta(x_1 \otimes y_1, x_2 \otimes y_2) = \alpha(x_1, x_2)\beta(y_1, y_2)$, where $x_j \in A^*$, $y_j \in B^*$. Clearly $\alpha \otimes \beta(x_1 \otimes y_1, x_2 \otimes y_2) = \alpha \otimes \beta(x_2 \otimes y_2, x_1 \otimes y_1)$, so $\alpha \otimes \beta$ is indeed an element of $S^2(A \otimes B)$.

The space $S^2A{\otimes}S^2B$ is a G-invariant subspace of $S^2(A{\otimes}B)$, i.e., if $T \in S^2A{\otimes}S^2B$, then $g \cdot T \in S^2A{\otimes}S^2B$ for all $g \in G$. One may think of the embedding $S^2A{\otimes}S^2B \to S^2(A{\otimes}B)$ as the result of the composition of the inclusion

$$S^2A{\otimes}S^2B \to A{\otimes}A{\otimes}B{\otimes}B = (A{\otimes}B)^{\otimes 2}$$

with the projection π_S (see §2.6.3)

$$(A{\otimes}B)^{\otimes 2} \to S^2(A{\otimes}B).$$

Since S^2A is an irreducible $GL(A)$ module and S^2B is an irreducible $GL(B)$-module, $S^2A{\otimes}S^2B$ is an irreducible G-module.

On the other hand $\dim S^2V = \binom{ab+1}{2} = (ab+1)ab/2$ and $\dim(S^2A{\otimes}S^2B) = \binom{a+1}{2}\binom{b+1}{2} = (ab+a+b+1)ab/4$. So we have not found all possible elements of S^2V.

To find the missing polynomials, consider $\alpha \in \Lambda^2A$, $\beta \in \Lambda^2B$, and define $\alpha{\otimes}\beta(x_1{\otimes}y_1, x_2{\otimes}y_2) = \alpha(x_1,x_2)\beta(y_1,y_2)$. Observe that

$$\begin{aligned}\alpha{\otimes}\beta(x_2{\otimes}y_2, x_1{\otimes}y_1) &= \alpha(x_2,x_1)\beta(y_2,y_1)\\ &= (-\alpha(x_1,x_2))(-\beta(y_1,y_2))\\ &= \alpha{\otimes}\beta(x_1{\otimes}y_1, x_2{\otimes}y_2).\end{aligned}$$

Thus $\alpha{\otimes}\beta \in S^2(A{\otimes}B)$, and extending the map linearly yields an inclusion $\Lambda^2A{\otimes}\Lambda^2B \subset S^2(A{\otimes}B)$.

Now $\dim(\Lambda^2A{\otimes}\Lambda^2B) = (ab-a-b+1)ab/4$ so $\dim(\Lambda^2A{\otimes}\Lambda^2B) + \dim(S^2A{\otimes}S^2B) = S^2(A{\otimes}B)$, and thus

$$S^2(A{\otimes}B) = (\Lambda^2A{\otimes}\Lambda^2B) \oplus (S^2A{\otimes}S^2B)$$

is a decomposition of $S^2(A{\otimes}B)$ into $GL(A) \times GL(B)$-submodules, in fact into $GL(A) \times GL(B)$-irreducible submodules.

Exercise 2.7.1.1: Verify that the above decomposition is really a direct sum, i.e., that $(\Lambda^2A{\otimes}\Lambda^2B)$ and $(S^2A{\otimes}S^2B)$ are disjoint.

Exercise 2.7.1.2: Decompose $\Lambda^2(A{\otimes}B)$ as a $GL(A) \times GL(B)$-module.

2.7.2. Two by two minors. I now describe the inclusion $\Lambda^2A{\otimes}\Lambda^2B \to S^2(A{\otimes}B)$ in bases. Let (a_i), (b_s) respectively be bases of A, B, and let $(a_i{\otimes}b_s)$ denote the induced basis of $A{\otimes}B$ and $(\alpha^i{\otimes}\beta^s)$ the induced dual basis for $A^*{\otimes}B^*$. Identify $\alpha^i{\otimes}\beta^s$ with the matrix having a one in the (i,s)-entry and zeros elsewhere. Consider the following quadratic polynomial on $A^*{\otimes}B^*$, viewed as the space of $\mathbf{a} \times \mathbf{b}$ matrices with coordinates $x^{i,s}$, i.e., $X = \sum_{i,s} x^{i,s}a_i{\otimes}b_s$ corresponds to the matrix whose (i,s)-th entry is $x^{i,s}$:

$$P_{jk|tu}(X) := x^{j,t}x^{k,u} - x^{k,t}x^{j,u},$$

which is the two by two minor $(jk|tu)$. As a tensor

$$\begin{aligned} P_{jk|tu} &= (a_j\otimes b_t)(a_k\otimes b_u) - (a_k\otimes b_t)(a_j\otimes b_u) \\ &= \frac{1}{2}[a_j\otimes b_t\otimes a_k\otimes b_u + a_k\otimes b_u\otimes a_j\otimes b_t - a_k\otimes b_t\otimes a_j\otimes b_u - a_j\otimes b_u\otimes a_k\otimes b_t]. \end{aligned}$$

Exercise 2.7.2.1: Show that $A\otimes B$ is canonically isomorphic to $B\otimes A$ as a $GL(A)\times GL(B)$-module. More generally, for all $\sigma\in\mathfrak{S}_k$, $V_{\sigma(1)}\otimes\cdots\otimes V_{\sigma(k)}$ is isomorphic to $V_1\otimes\cdots\otimes V_k$ as a $GL(V_1)\times\cdots\times GL(V_k)$-module.

Now use the canonical isomorphism of Exercise 2.7.2.1 $A\otimes B\otimes A\otimes B\simeq A\otimes A\otimes B\otimes B$:

$$\begin{aligned} P_{jk|tu} &= \frac{1}{2}[a_j\otimes a_k\otimes b_t\otimes b_u + a_k\otimes a_j\otimes b_u\otimes b_t - a_k\otimes a_j\otimes b_t\otimes b_u - a_j\otimes a_k\otimes b_u\otimes b_t] \\ &= \frac{1}{2}[(a_j\otimes a_k - a_k\otimes a_j)\otimes b_t\otimes b_u + (a_k\otimes a_j - a_j\otimes a_k)\otimes b_u\otimes b_t] \\ &= (a_j\wedge a_k)\otimes b_t\otimes b_u + (a_k\wedge a_j)\otimes b_u\otimes b_t \\ &= 2(a_j\wedge a_k)\otimes(b_t\wedge b_u). \end{aligned}$$

Thus

A two by two minor of a matrix, expressed as a tensor in $S^2(A\otimes B)$, corresponds to an element of the subspace $\Lambda^2A\otimes\Lambda^2B\subset S^2(A\otimes B)$.

Compare this remark with (2.6.11).

Another perspective on the space of two by two minors is as follows: a linear map $x: A\to B^*$ has rank (at most) one if and only if the induced linear map

$$x^{\wedge 2}:\Lambda^2A\to\Lambda^2B^*$$

is zero. Now $x^{\wedge 2}\in\Lambda^2A^*\otimes\Lambda^2B^*$, and for any vector space U and its dual U^* there is a perfect pairing $U\times U^*\to\mathbb{C}$. Thus a way to test if x has rank one is if $x^{\wedge 2}$ pairs with each element of $\Lambda^2A\otimes\Lambda^2B$ to be zero.

In contrast, consider the induced map $x^2: S^2A\to S^2B^*$. This map is never identically zero if x is nonzero, so the set $\{x\in A^*\otimes B^*\mid \phi(x)=0\ \forall\ \phi\in S^2A\otimes S^2B\}$ is just the zero vector $0\in A^*\otimes B^*$.

2.7.3. Exercises on equations for the set of rank at most $k-1$ matrices. Let $\alpha^1,\dots,\alpha^k\in A^*$, $\beta^1,\dots,\beta^k\in B^*$ and consider $P:=(\alpha^1\wedge\cdots\wedge\alpha^k)\otimes(\beta^1\wedge\cdots\wedge\beta^k)\in\Lambda^kA^*\otimes\Lambda^kB^*$. By Exercise 2.7.3.2 one may consider $P\in S^k(A\otimes B)^*$, i.e., as a homogeneous polynomial of degree k on $A\otimes B$.

(1) Show that $\Lambda^2A\otimes\Lambda^2B\subset S^2(A\otimes B)$ is exactly the span of the collection of two by two minors (with respect to any choice of bases) considered as quadratic polynomials on $A^*\otimes B^*$.

(2) Show that $\Lambda^k A \otimes \Lambda^k B \subset S^k(A \otimes B)$, and that it corresponds to the span of the collection of $k \times k$ minors.

(3) Show that if $T \in A \otimes B$ is of the form $T = a_1 \otimes b_1 + \cdots + a_k \otimes b_k$ where $(a_1, \ldots, a_k)$ and $(b_1, \ldots, b_k)$ are linearly independent sets of vectors, then there exists $P \in \Lambda^k A^* \otimes \Lambda^k B^*$ such that $P(T) \neq 0$. Conclude that the set of rank at most $k-1$ matrices is the common zero locus of the polynomials in $\Lambda^k A^* \otimes \Lambda^k B^*$

(4) Given $T \in A \otimes B$, consider $T : A^* \to B$ and the induced linear map $T^{\wedge k} : \Lambda^k A^* \to \Lambda^k B$, i.e. $T^{\wedge k} \in \Lambda^k A \otimes \Lambda^k B$. Show that the rank of T is less than k if and only if $T^{\wedge k}$ is zero.

The perspectives of the last two exercises are related by noting the perfect pairing

$$(\Lambda^k A \otimes \Lambda^k B) \times (\Lambda^k A^* \otimes \Lambda^k B^*) \to \mathbb{C}.$$

2.7.4. The Pfaffian. Let E be a vector space of dimension $n = 2m$ and let $\Omega_E \in \Lambda^n E$ be a volume form. Let $x \in \Lambda^2 E$ and consider $x^{\wedge m} \in \Lambda^n E$. Since $\dim(\Lambda^n E) = 1$, there exists $c_x \in \mathbb{C}$ such that $x^{\wedge m} = c_x \Omega_E$. Define $\mathrm{Pf} \in S^m(\Lambda^2 E)$ by $\mathrm{Pf}(x) = \frac{1}{m!} c_x$. By its definition, Pf depends only on a choice of volume form and thus is invariant under $SL(E) = SL(E, \Omega)$ (which, by definition, is the group preserving Ω_E).

The Pfaffian is often used in connection with the orthogonal group because endomorphisms $E \to E$ that are "skew-symmetric" arise often in practice, where, in order to make sense of "skew-symmetric" one needs to choose an isomorphism $E \to E^*$, which can be accomplished, e.g., with a non-degenerate quadratic form $Q \in S^2 E^*$. It plays an important role in differential geometry as it is the key to Chern's generalization of the Gauss-Bonnet theorem, see e.g., [**291**, Chap. 13].

For example, if $n = 4$ and $x = \sum_{i,j} x^{ij} e_i \wedge e_j$, then

$$x \wedge x = \sum_{ijkl} x^{ij} x^{kl} e_i \wedge e_j \wedge e_k \wedge e_l = \sum_{\sigma \in \mathfrak{S}_4} x^{\sigma(1)\sigma(2)} x^{\sigma(3)\sigma(4)} e_1 \wedge e_2 \wedge e_3 \wedge e_4.$$

In particular $\mathrm{Pf}(x) = x^{12}x^{34} - x^{13}x^{24} + x^{14}x^{23}$.

2.7.5. Exercises on the Pfaffian.

(1) Let x be a skew-symmetric matrix. If $x^{ij} = 0$ for $i > j+1$, show that $\mathrm{Pf}(x) = x^{12}x^{34} \cdots x^{2m-1,2m}$. Note that the eigenvalues of x are $\pm\sqrt{-1}x^{j,j+1}$.

(2) Using the previous exercise, show that if x is skew-symmetric, $\mathrm{Pf}(x)^2 = \det(x)$.

(3) Fix a basis $e_1, \dots, e_n$ of E so $x \in \Lambda^2 E$ may be written as a skew-symmetric matric $x = (x^i_j)$. Show that in this basis

$$\mathrm{Pf}(x^i_j) = \frac{1}{2^n n!} \sum_{\sigma \in \mathfrak{S}_{2n}} sgn(\sigma) x^{\sigma(1)}_{\sigma(2)} \cdots x^{\sigma(2m-1)}_{\sigma(2m)} \tag{2.7.1}$$

$$= \frac{1}{2^n} \sum_{\sigma \in \mathcal{P}} sgn(\sigma) x^{\sigma(1)}_{\sigma(2)} \cdots x^{\sigma(2m-1)}_{\sigma(2m)} \tag{2.7.2}$$

where $\mathcal{P} \subset \mathfrak{S}_{2m}$ consists of the permutations such that $\sigma(2i-1) < \sigma(2i)$ for all $0 \le i \le m$. Note that these expressions are defined for arbitrary matrices and the $SL(V)$ invariance still holds as $S^n(\Lambda^2 V^*)$ is a $SL(V)$-submodule of $S^n(V^* \otimes V^*)$.

2.8. Decomposition of $V^{\otimes 3}$

When $d > 2$ there are subspaces of $V^{\otimes d}$ other than the completely symmetric and skew-symmetric tensors that are invariant under changes of bases. These spaces will be studied in detail in Chapter 6. In this section, as a preview, I consider $V^{\otimes 3}$.

Change the previous notation of $\pi_S : V^{\otimes 3} \to V^{\otimes 3}$ and $\pi_\Lambda : V^{\otimes 3} \to V^{\otimes 3}$ to $\rho_{\young(123)}$ and $\rho_{\young(1,2,3)}$ respectively.

Define the projections

$$\rho_{\young(1,2)} : V \otimes V \otimes V \to \Lambda^2 V \otimes V$$

$$v_1 \otimes v_2 \otimes v_3 \mapsto \frac{1}{2}(v_1 \otimes v_2 \otimes v_3 - v_2 \otimes v_1 \otimes v_3)$$

$$\rho_{\young(13)} : V \otimes V \otimes V \to V \otimes S^2 V$$

$$v_1 \otimes v_2 \otimes v_3 \mapsto \frac{1}{2}(v_1 \otimes v_2 \otimes v_3 + v_3 \otimes v_2 \otimes v_1)$$

which are also endomorphisms of $V^{\otimes 3}$. Composing them, gives

$$\rho_{\young(13,2)} = \rho_{\young(13)} \circ \rho_{\young(1,2)} : V \otimes V \otimes V \to S_{\young(13,2)} V$$

where $S_{\young(13,2)} V$ is defined as the image of $\rho_{\young(13,2)}$. The essential point here is:

The maps $\rho_{\young(13,2)}, \rho_{\young(13)}, \rho_{\young(1,2)}$ all commute with the action of $GL(V)$ on $V^{\otimes 3}$. Therefore the image of $\rho_{\young(13,2)}$ is a $GL(V)$-submodule of $V^{\otimes 3}$. Similarly define $S_{\young(12,3)} V$ as the image of $\rho_{\young(12,3)}$.

Warning: The image of $\rho_{\young(13)} : \Lambda^2 V \otimes V \to V^{\otimes 3}$ is no longer skew-symmetric in its first two arguments. Similarly the image of $\rho_{\young(1,2)} : V \otimes S^2 V \to V^{\otimes 3}$ is not symmetric in its second and third argument.

2.8.1. Exercises.

(1) Show that the sequence

$$S_{\young(12,3)}V \to V \otimes \Lambda^2 V \to \Lambda^3 V$$

is exact.

(2) Show that $S_{\young(12,3)}V \cap S_{\young(13,2)}V = (0)$. ⊚

(3) Show that there is a direct sum decomposition:

$$V^{\otimes 3} = S^3 V \oplus S_{\young(13,2)}V \oplus \Lambda^3 V \oplus S_{\young(12,3)}V \tag{2.8.1}$$
$$= S_{\young(123)}V \oplus S_{\young(12,3)}V \oplus S_{\young(1,2,3)}V \oplus S_{\young(13,2)}V$$

(4) Let $\dim V = 2$ with basis e_1, e_2. Calculate the images $S_{\young(13,2)}V$ and $S_{\young(12,3)}V$ explicitly.

(5) Now consider $S_{\young(31,2)}V$ to be the image of $\rho_{\young(31,2)}$. Show it is also a $GL(V)$ invariant subspace. Using the previous exercise, show that when $\dim V = 2$, that $S_{\young(31,2)}V \subset S_{\young(13,2)}V \oplus S_{\young(12,3)}V$.

(6) Now let $\dim V = 3$ with basis e_1, e_2, e_3 calculate the images $S_{\young(13,2)}V$ and $S_{\young(12,3)}V$ explicitly. Show that $S_{\young(31,2)}V \subset S_{\young(13,2)}V \oplus S_{\young(12,3)}V$. More generally, show that any of these spaces is contained in the span of any two others.⊚

(7) Explain why the previous exercise implies $S_{\young(31,2)}V \subset S_{\young(13,2)}V \oplus S_{\young(12,3)}V$ for V of arbitrary dimension.

(8) Consider the sequence of maps $d : S^p V \otimes \Lambda^q V \to S^{p-1} V \otimes \Lambda^{q+1} V$ given in coordinates by $f \otimes u \mapsto \sum \frac{\partial f}{\partial x^i} \otimes x_i \wedge u$. Show that $d^2 = 0$ so we have an exact sequence of $GL(V)$-modules ⊚:

$$0 \to S^d V \to S^{d-1} V \otimes V \to S^{d-2} V \otimes \Lambda^2 V \to \cdots \to V \otimes \Lambda^{d-1} V \to \Lambda^d V \to 0. \tag{2.8.2}$$

2.8.2. Isotypic decompositions. Recall that if a linear map $f : V \to V$ has distinct eigenvalues, there is a canonical decomposition of V into a direct sum of one dimensional eigenspaces. But if there are eigenvalues that occur with multiplicity, even if f is diagonalizable, there is no canonical splitting into eigenlines. The same phenomenon is at work here. Although our decomposition (2.8.1) is *invariant* under the action of $GL(V)$, it is not *canonical*, i.e., independent of choices. The space $S_{\begin{smallmatrix}1&3\\2\end{smallmatrix}}V \oplus S_{\begin{smallmatrix}1&2\\3\end{smallmatrix}}V$ is the analog of an eigenspace for an eigenvalue with multiplicity two. In fact, I claim that $S_{\begin{smallmatrix}3&1\\2\end{smallmatrix}}V, S_{\begin{smallmatrix}1&2\\3\end{smallmatrix}}V, S_{\begin{smallmatrix}1&3\\2\end{smallmatrix}}V$ are all isomorphic $GL(V)$-modules!

It is exactly that these modules are isomorphic which causes the decomposition to be non-canonical.

Exercise 2.8.2.1: Prove the claim. ⊚

Exercise 2.8.2.2: Calculate $\dim S_{\begin{smallmatrix}1&2\\3\end{smallmatrix}}V$ in terms of $n = \dim V$ by using your knowledge of $\dim V^{\otimes 3}, \dim S^3V, \dim \Lambda^3 V$ and the fact that $S_{\begin{smallmatrix}3&1\\2\end{smallmatrix}}V, S_{\begin{smallmatrix}1&2\\3\end{smallmatrix}}V$ are isomorphic vector spaces.

Definition 2.8.2.3. Let G be a group. G is said to be *reductive* if every G-module V admits a decomposition into a direct sum of irreducible G-modules.

For example $GL(V)$ is reductive but the group

$$N := \left\{ \begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix} \mid t \in \mathbb{C} \right\}$$

is not.

Exercise 2.8.2.4: Show that N is not reductive by showing that there is a line L in $\mathbb{C}^2$ preserved by N, but no N-invariant complement to L exists.

Remark 2.8.2.5. If one is working over finite fields, there are several different notions of reductive which coincide over $\mathbb{C}$, see e.g. [**113**].

Definition 2.8.2.6. Let G be reductive and let V be a G-module. Write the decomposition of V into irreducible G-modules as

$$V = W_{1,1} \oplus \cdots \oplus W_{1,m_1} \oplus W_{2,1} \oplus \cdots \oplus W_{2,m_2} \oplus \cdots \oplus W_{r,1} \oplus \cdots \oplus W_{r,m_r}$$

where the $W_{i,1}, \dots, W_{i,m_i}$'s are isomorphic G-modules and no $W_{i,j}$ is isomorphic to a $W_{k,l}$ for $k \neq i$. While this decomposition is invariant under the action of G, it is not canonical. If $U_i = W_{i,1} \oplus \cdots \oplus W_{i,m_i}$, then the decomposition $V = \oplus_i U_i$ is canonical and is called the *isotypic* decomposition of V as a G-module. The U_i are called *isotypic components*. We say m_j is the *multiplicity* of the irreducible module $W_{j,1}$ in V

Definition 2.8.2.7. Let $S_{21}V$ denote the irreducible $GL(V)$ module isomorphic to each of $S_{\begin{smallmatrix}1&2\\3\end{smallmatrix}}V, S_{\begin{smallmatrix}1&3\\2\end{smallmatrix}}V, S_{\begin{smallmatrix}2&3\\1\end{smallmatrix}}V$.

Exercise 2.8.2.8: Show that for $d > 3$ the kernel of the last nonzero map in Exercise 2.8.1(8) and the kernel of the second map give rise to different generalizations of $S_{21}V$. ⊚

In Chapter 6, the $GL(V)$ modules in $V^{\otimes d}$ will be studied for all d. They will be indexed by partitions of d. For example, the partitions of 3 give rise to the three modules $S_3V = S^3V$, $S_{111}V = \Lambda^3 V$, and our new friend $S_{21}V$.

2.8.3. Exercises.

(1) Show that there is an invariant decomposition

$$S^3(A\otimes B) = (S^3A\otimes S^3B) \oplus \pi_S(S_{\begin{smallmatrix}3&1\\2&\end{smallmatrix}}A\otimes S_{\begin{smallmatrix}3&1\\2&\end{smallmatrix}}B) \oplus \Lambda^3A\otimes\Lambda^3B$$

as a $GL(A)\times GL(B)$-module.

(2) Decompose $\Lambda^3(A\otimes B)$ as a $GL(A)\times GL(B)$-module.

(3) Let $\tilde{S}_{\begin{smallmatrix}1&2\\3&\end{smallmatrix}}A$ denote the kernel of the map $S^2A\otimes A \to S^3A$, so it represents a copy of the module $S_{21}A$ in $A^{\otimes 3}$.

Show that if $R \in \tilde{S}_{\begin{smallmatrix}1&2\\3&\end{smallmatrix}}A$, then $R(u,v,w) = R(v,u,w)$ for all $u,v,w \in A^*$ and

$$R(u,v,u) = -\frac{1}{2}R(u,u,v) \quad \forall u,v,\in A^*. \tag{2.8.3}$$

2.9. Appendix: Basic definitions from algebra

2.9.1. Linear algebra definitions. Vector spaces, dual spaces, linear maps and bilinear maps are defined in §1.2.1.

Definition 2.9.1.1. The *dimension* of a vector space V is the smallest number $\mathbf{v}$ such that there exist $e_1,\dots,e_{\mathbf{v}} \in V$ such that any $x \in V$ may be written $x = \sum c^ie_i$ for constants c^i. Such a set of vectors $\{e_j\}$ is called a *basis* of V.

Definition 2.9.1.2. The *rank* of a linear map $f : V \to W$ is $\dim(f(V))$.

Definition 2.9.1.3. Given a linear map $f : V \to V$, a nonzero vector $v \in V$ such that $f(v) = \lambda v$ for some $\lambda \in \mathbb{C}$ is called an *eigenvector*.

If V has basis $v_1,\dots,v_{\mathbf{v}}$, let $\alpha^1,\dots,\alpha^{\mathbf{v}}$ be the basis of V^* such that $\alpha^i(v_j) = \delta^i_j$. It is called the *dual basis* to $(v_1,\dots,v_{\mathbf{v}})$. Let W have basis $w_1,\dots,w_{\mathbf{w}}$. A linear map $f : V \to W$ is determined by its action on a basis. Write $f(v_j) = f^s_jw_s$. Then the matrix representing f with respect to these bases is (f^s_j).

Definition 2.9.1.4. Given vector spaces U, V, and W, and linear maps $f : U \to V$, and $g : V \to W$, one says the sequence

$$0 \to U \xrightarrow{f} V \xrightarrow{g} W \to 0$$

is an *exact sequence of vector spaces* if f is injective, g is surjective, and $\ker(g) = \operatorname{Image}(f)$.

Remark 2.9.1.5. Regarding dual spaces, the spaces V and V^* can be thought of as covariant and contravariant. There is no canonical isomorphism $V \cong V^*$. If V is endowed with a Hermitian inner product, there is then a canonical identification of V^* with $\overline{V}$, the complex conjugate. However in this book I will work with projectively invariant properties so Hermitian inner products will not play a role.

Definition 2.9.1.6. Given $f \in \operatorname{Hom}(V, W)$, define $f^T \in \operatorname{Hom}(W^*, V^*)$, called the *transpose* or *adjoint* of f, by $f^T(\beta)(v) = \beta(f(v))$.

2.9.2. Definitions regarding groups and rings.

Definition 2.9.2.1. A *group* is a set G, with a pairing $G \times G \to G$, $(a, b) \mapsto ab$ a preferred element (the identity) $\operatorname{Id} \in G$, such that $a(\operatorname{Id}) = (\operatorname{Id})a = a$ for all $a \in G$ and such that for each $a \in G$ there is an inverse element which is both a left and right inverse.

For example, a vector space is a group with the operation $+$, and the identity element 0.

One of the most important groups is the permutation group:

Definition 2.9.2.2 (The group $\mathfrak{S}_n$)**.** Given a collection of n ordered objects, the set of permutations of the objects forms a group, called the *symmetric group on n elements* or the *permutation group*, and is denoted $\mathfrak{S}_n$.

An important fact about elements $\sigma \in \mathfrak{S}_n$ is that they may be written as a product of transpositions, e.g. $(1, 2, 3) = (2, 3)(1, 2)$. This decomposition is not unique, nor is the number of transpositions used in the expression unique, but the parity is.

Definition 2.9.2.3. For a permutation σ, define the *sign* of σ, $\operatorname{sgn}(\sigma)$ to be $+1$ if an even number of transpositions are used and -1 if an odd number are used.

Definition 2.9.2.4. Let G, H be groups. A map $f : G \to H$ is called a *group homomorphism* if $f(g_1 g_2) = f(g_1)f(g_2)$ for all $g_1, g_2 \in G$.

Definition 2.9.2.5. A *ring* is a set equipped with an additive group structure and a multiplicative operation such that the two operations are compatible.

An *ideal* $I \subset R$ is a subset that is a group under addition, and strongly closed under multiplication in the sense that if $P \in I$ and $Q \in R$, then $PQ \in I$.

Definition 2.9.2.6. An *algebra* is a vector space V equipped with a multiplication compatible with the vector space structure, in other words, a ring that is also a vector space.

2.10. Appendix: Jordan and rational canonical form

Most linear algebra texts cover Jordan canonical form of a linear map $f : V \to V$ so I just state the result:

Let $f : V \to V$ be a linear map. If it has distinct eigenvalues it is diagonalizable under the action of $GL(V)$. If not, one can decompose V into *generalized eigenspaces* for f, i.e., say there are k distinct eigenvalues $\lambda_1, \dots, \lambda_k$. One may write

$$V = V_1 \oplus \cdots \oplus V_k$$

where if the minimal polynomial of f is denoted ϕ_f, one can write $\phi_f = \phi_1^{a_1} \cdots \phi_k^{a_k}$ where $\phi_j(\lambda) = (\lambda - \lambda_j)$. Then $g_j := f\,|_{V_j} - \lambda_j \operatorname{Id}_{V_j}$ is nilpotent on V_j and choosing bases with respect to generating vectors one obtains (noncanonically) blocks of sizes $s_1^j = a_j \geq s_2^j \cdots \geq s_{m_j}^j$, where each block in matrices looks like

$$\begin{pmatrix} \lambda_j & 1 & 0 & \cdots & 0 \\ 0 & \lambda_j & 1 & & \vdots \\ & \ddots & \ddots & \ddots & 0 \\ & & & \lambda_j & 1 \\ & & & 0 & \lambda_j \end{pmatrix}$$

In the end one obtains a matrix representing f whose entries are zero except on the diagonal, which have the eigenvalues, and some of the entries above the diagonal, which are ones.

To obtain the rational canonical form from the Jordan form, in each V_j, first take a vector $v_j = v_{j,1}$ such that $g_j^{\alpha_j - 1}(v_j) \neq 0$, let $w_1 = \sum v_{j,1}$. Let $W_1 \subset V$ be the subspace generated by successive images of w_1 under f. Note that the minimal polynomial of f restricted to W_1, call it ψ_1, is ϕ_f. From this one deduces that f restricted to W_1 has the form

$$\begin{pmatrix} 0 & & & -p_0 \\ 1 & \ddots & & \vdots \\ & \ddots & 0 & -p_{d-2} \\ & & 1 & -p_{d-1} \end{pmatrix}$$

where $\psi_1 = \phi_f = p_0 + p_1\lambda + \cdots + p_{d-1}\lambda^{d-1} + \lambda^d$.

Then in each V_j, in the complement of the subspace generated by the images of $v_{j,1}$ under g_j, take a vector $v_{j,2}$ such that the space $\langle g_j^s v_{j,2} \mid s \in \mathbb{N}\rangle$ has maximal dimension. Now let $w_2 = \sum v_{j,2}$ and consider the corresponding space W_2. The minimal polynomial of f restricted to W_2, call it ψ_2, divides ψ_1, and one obtains a matrix of the same form as above with respect to ψ_2 representing f restricted to W_2. One continues in this fashion. (For $u > m_j$, the contribution of V_j to the vector generating w_u is zero.) The polynomials ψ_u are called the *invariant divisors* of f and are independent of choices.

Note that ψ_{u+1} divides ψ_u and that, ignoring 1×1 blocks, the maximum number of Jordan blocks of size at least two associated to any eigenvalue of f is the number of invariant divisors.

Rational canonical form is described in [**133**, VI.6], also see, e.g., [**125**, §7.4]. For a terse description of rational canonical form via the Jordan form, see [**248**, Ex. 7.4.8].

2.11. Appendix: Wiring diagrams

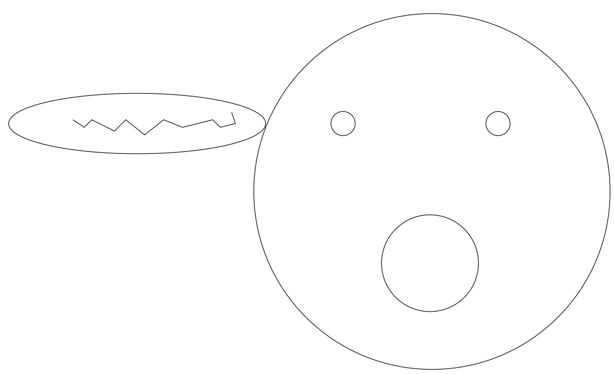


Figure 2.11.1. Wiring diagrams from unpublished notes of Clifford 1881

Wiring diagrams may be used to represent many tensors and tensor operations including contractions. Such diagrams date at least back to Clifford and were used by Feynman and Penrose [**260**] in physics, Cvitanović [**102**] (who calls them *birdtracks*) to study representation theory via invariant tensors, Kuperberg [**195**], Bar-Natan and Kontsevich [**16**], and Reshetikhin and Turaev [**271**] in knot theory/quantum groups, Deligne [**108**] and Vogel [**322, 321**] in their proposed categorical generalizations of Lie algebras, and many others.

A *wiring diagram* is a diagram that encodes a tensor as described below. A *natural wiring diagram*, or more precisely a *G-natural wiring diagram* is a diagram that encodes a G-invariant tensor $T \in A_1^{\otimes d_1} \otimes A_1^{*\otimes \delta_1} \otimes \cdots \otimes A_n^{\otimes d_n} \otimes A_n^{*\otimes \delta_n}$, where $G = GL(A_1) \otimes \cdots \otimes GL(A_n)$. Often such invariant tensors will be viewed as natural operations on other tensors. For example $tr = \mathrm{Id}_V \in V \otimes V^*$ is a $GL(V)$-invariant tensor.

When a diagram is viewed as an operation, it is to be read top to bottom. Strands going in and coming out represent vector spaces, and a group of strands represents the tensor product of the vector spaces in the group. If there are no strands going in (resp. coming out) we view this as inputing a scalar (resp. outputing a scalar).

For example, encode a linear map $f : A \to A$ by the diagram in Figure 2.11.2 (a). The space of tensors designated in a box at the top and bottom of the figure indicate the input and output of the operator. More generally, if a diagram contains a T with a circle around it with k strands with arrows going out, and ℓ strands with arrows going in, then $T \in V^{\otimes k} \otimes V^{*\otimes \ell}$

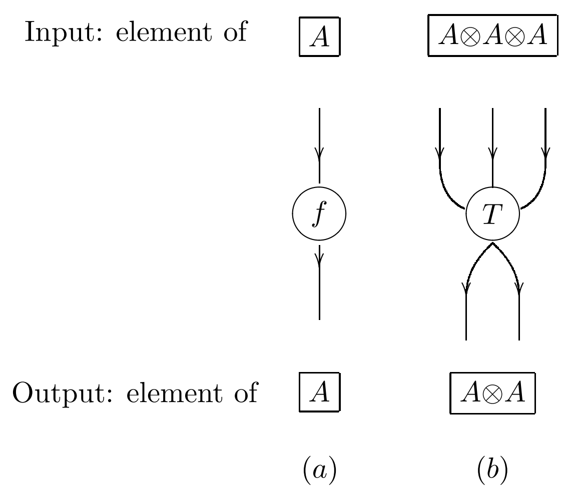


Figure 2.11.2. Wiring diagram of (a) a map $f : A \to A$, i.e. of $f \in A^* \otimes A$ and (b) a tensor $T \in (A^*)^{\otimes 3} \otimes A^{\otimes 2}$.

Recalling that one may also view $f \in A^* \otimes A$ as a map $A^* \to A^*$, or as a bilinear map $A \times A^* \to \mathbb{C}$, one may write the diagram in the three ways depicted in Figure 2.11.3. Note that the *tensor* is the same in all four cases; one just changes how one views it as an operator.

Of particular importance is the identity map $\mathrm{Id} : A \to A$, the first of the four diagrams in Figure 2.11.4. In these diagrams I indicate the identity by omitting the letter f.

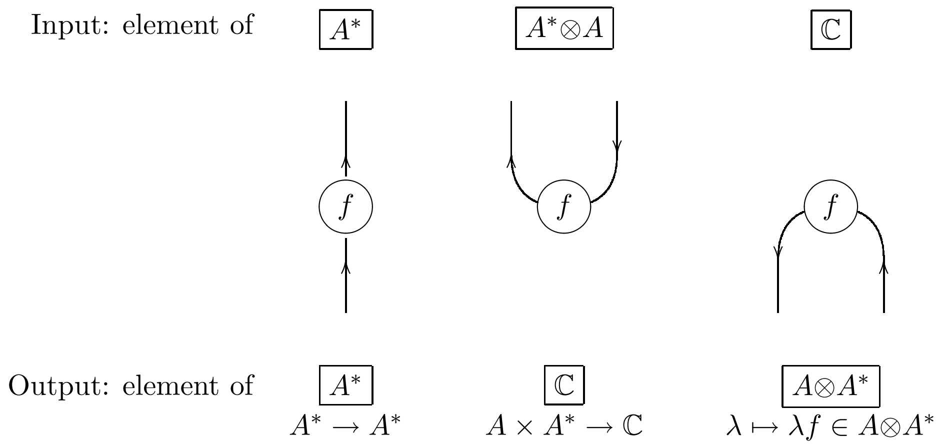


Figure 2.11.3. Three views of the same tensor as maps.

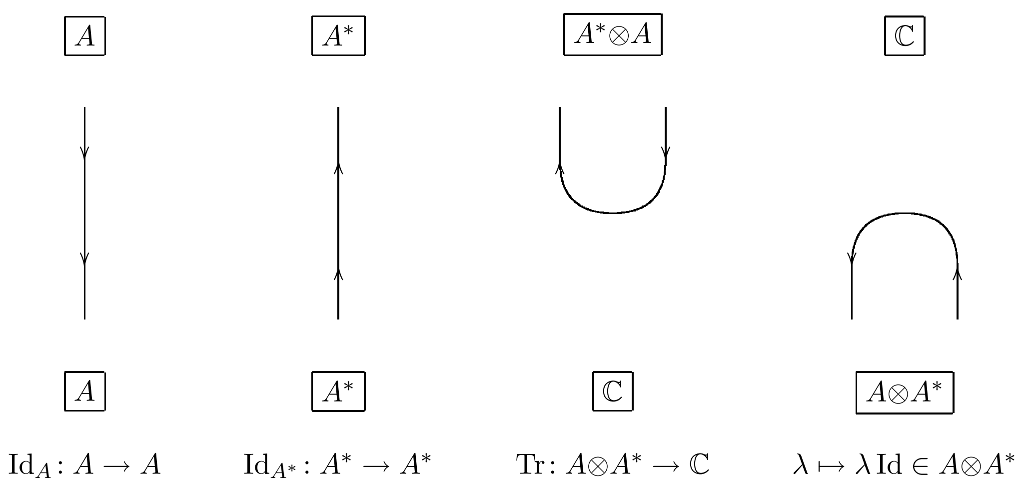
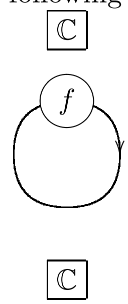


Figure 2.11.4. $\mathrm{Id} \in A^*\otimes A$ viewed as: a map $A \to A$, a map $A^* \to A^*$, a map $A^*\otimes A \to \mathbb{C}$ (i.e. a linear map maps to its trace), and a map $\mu_A : \mathbb{C} \to A^*\otimes A$ $(\lambda \mapsto \lambda\, \mathrm{Id}_A)$

.

If an arrow is present it points from a vector space to its dual. The identity and dualizing operators are depicted in Figure 2.11.

Exercise 2.11.0.7: Show that the following diagram represents the scalar $Tr(f)$, for a linear map $f : V \to V$.



In particular, if $f : V \to V$ is a projection, then the diagram represents $\dim \mathrm{Image}(f)$. (Compare with Exercise 2.3.2.(9).)

2.11.1. A wiring diagram for the dimension of V. If we compose the maps $\mu_V : \mathbb{C} \to V^*{\otimes}V$ of figure 2.11.4 and $Tr_V : V^*{\otimes}V \to \mathbb{C}$, we obtain a map $Tr_V \circ \mu_V : \mathbb{C} \to \mathbb{C}$ that is multiplication by some scalar.

Exercise 2.11.1.1: Show that the scalar is $\dim V$.

Thus the picture

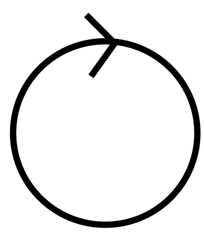


Figure 2.11.5. Symbolic representation of $\dim V$

should be interpreted as the scalar $\dim V$. Similarly, since $\mathrm{Id}_{V^{\otimes d}}$ has trace equal to $(\dim V)^d$, the union of d disjoint circles should be interpreted as the scalar $(\dim V)^d$.

Below I will take formal sums of diagrams, and under such a formal sum one adds scalars and tensors.

Recall the linear map

$$\sigma : V{\otimes}V \to V{\otimes}V$$
$$a{\otimes}b \mapsto b{\otimes}a.$$

It has the wiring diagram in Figure 2.11.6.

Exercise 2.11.1.2: Show pictorially that $\mathrm{trace}\,\sigma = \dim V$ by composing the picture with σ with the picture for $\mathrm{Id}_{V\otimes V}$. (Of course one can also

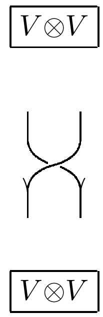


Figure 2.11.6. A wiring diagram for the map $a \otimes b \mapsto b \otimes a$.

obtain the result by considering the matrix of σ with respect to the basis $e_i \otimes e_j$.)

In Chapter 6 I show that *all* $GL(V)$-natural wiring diagrams in $V^{\otimes d} \otimes V^{* \otimes \delta}$ are built from Id_V and σ.

2.11.2. A wiring diagram for matrix multiplication. Matrix multiplication is depicted as a wiring diagram in Figure 2.11.7.

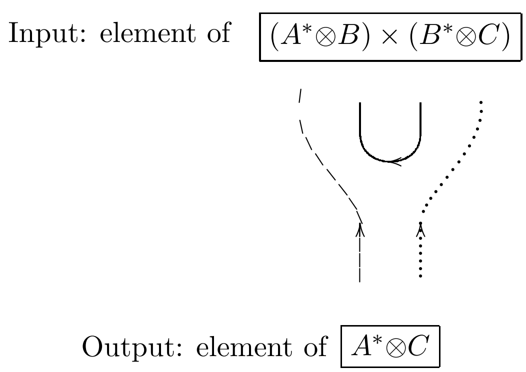


Figure 2.11.7. Matrix multiplication as an operator $(A^* \otimes B) \times (B^* \otimes C) \to A^* \otimes C$.

Exercise 2.11.2.1: Show that the diagram in Figure 2.11.7 agrees with the matrix multiplication you know and love.

Encode the tensor π_S of §2.6.2 with the white box shorthand on the left hand side. It is one half the formal sum of the wiring diagrams for σ and $\mathrm{Id}_{V^{\otimes 2}}$, as on the right hand side of Figure 2.11.8.

Encode $\pi_S : V^{\otimes d} \to V^{\otimes d}$ by a diagram as in Figure 2.11.10 with d strands. This is $\frac{1}{d!}$ times the formal sum of diagrams corresponding to all permutations. Recall that since each permutation is a product of transposi-

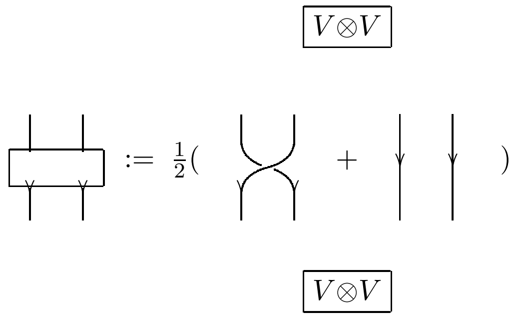


Figure 2.11.8. The wiring diagram $\frac{1}{2}(\sigma + \mathrm{Id}_{V^{\otimes 2}})$.

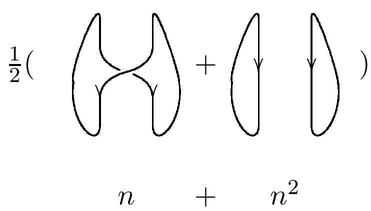


Figure 2.11.9. Symbolic representation of $\dim(S^2V)$.

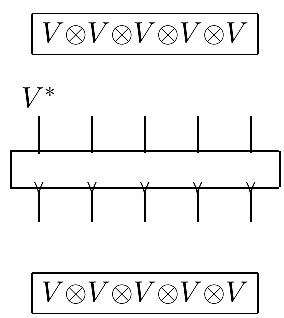


Figure 2.11.10. The symmetrizing wiring diagram π_S for $d = 5$.

tions, a wiring diagram for any permutation can be written as a succession of σ's acting on different pairs of factors, for example one can write the diagram for the cyclic permutation on three factors as in Figure 2.11.11.

2.11.3. Exercises.

(1) Reprove that $\pi_S : V^{\otimes 2} \to V^{\otimes 2}$ is a projection operator by showing that the concatenation of two wiring diagrams for π_S yields the diagram of π_S.

(2) Show that $\dim(S^2V) = \binom{n+1}{2}$ pictorially by using the picture 2.11.9.

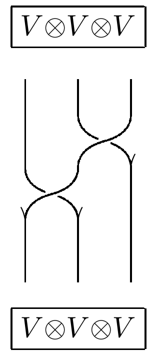


Figure 2.11.11. The cyclic permutation on three factors.

(3) Define and give a wiring diagram for $\pi_\Lambda : V^{\otimes 2} \to V^{\otimes 2}$, the projection operator to $\Lambda^2 V$ and use the diagram to compute the dimension of $\Lambda^2 V$.

(4) Show that $V^{\otimes 2} = S^2V \oplus \Lambda^2 V$ by showing that the diagram of $\mathrm{Id}_{V^{\otimes 2}}$ is the sum of the diagrams for S^2V and $\Lambda^2 V$.

(5) Calculate the trace of the cyclic permutation on three factors $V{\otimes}V{\otimes}V \to V{\otimes}V{\otimes}V$ using wiring diagrams. Verify your answer by using bases.

(6) Show diagrammatically that $V^{\otimes 3} \neq \Lambda^3 V \oplus S^3 V$ by showing that the diagram for $\mathrm{Id}_{V^{\otimes 3}}$ is not the sum of the diagrams for π_S and π_Λ.

(7) Use wiring diagrams to calculate $\dim S^3V$ and $\dim \Lambda^3 V$.

The three maps $\rho_{\begin{smallmatrix}1&3\end{smallmatrix}}$, $\rho_{\begin{smallmatrix}1\\2\end{smallmatrix}}$, and $\rho_{\begin{smallmatrix}1&3\\2\end{smallmatrix}}$ are depicted visually in the three wiring diagrams of Figure 2.11.12, which add new components represented by the black and white boxes. Each of these diagrams, just as in the definitions above, shows what to do with a rank one element of $V{\otimes}V{\otimes}V$, which is extended linearly. The three vertical lines in each diagram correspond to the three tensor factors of a rank one element $v_1{\otimes}v_2{\otimes}v_3$. A black box indicates skew-symmetrization, and a white box symmetrization. The factor of $\frac{1}{b!}$, where b is the number of wires passing through a box, is implicit.

$\rho_{\begin{smallmatrix}1&2\\3\end{smallmatrix}}$, may be described by the wiring diagram in Figure 2.11.13.

Exercise 2.11.3.1: Prove the direct sum decomposition (2.8.1) diagrammatically.

Exercise 2.11.3.2: Calculate $\dim S_{\begin{smallmatrix}1&2\\3\end{smallmatrix}}V$ in terms of $n = \dim V$ by using wiring diagrams.⊚

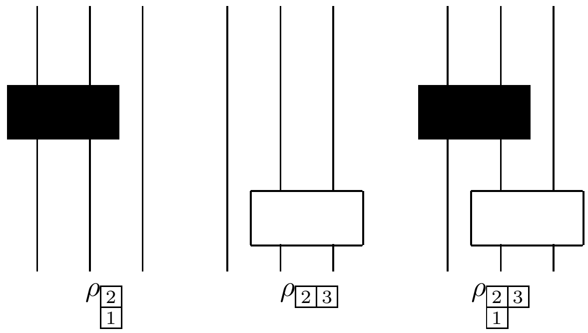


Figure 2.11.12. The wiring diagrams for three projection operators

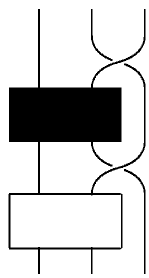


Figure 2.11.13. A wiring diagram for the projection map $\rho_{12,3}$.

.

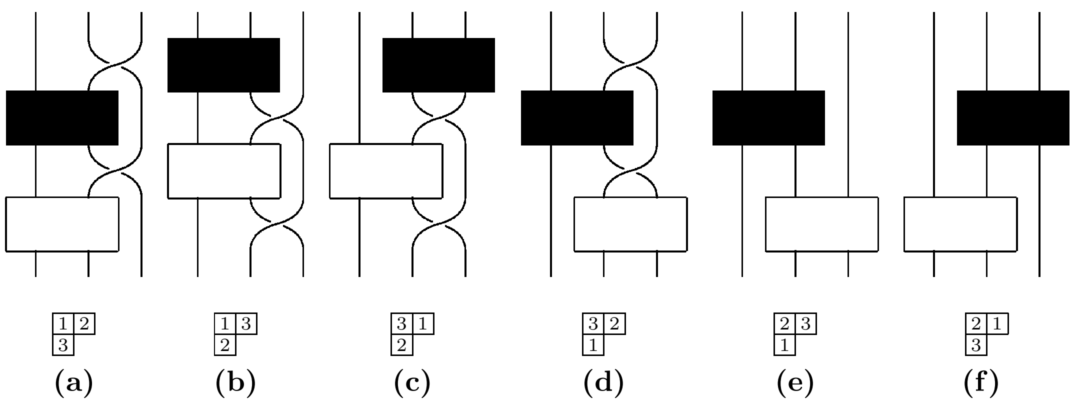


Figure 2.11.14. 6 natural realizations of $S_{21}V$