Structured Tensors and the Geometry of Data

by

Anna Leah Seigal

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Mathematics

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Bernd Sturmfels, Chair
Professor David Eisenbud
Professor James Demmel
Professor Hernan Garcia-Melan

Spring 2019

Structured Tensors and the Geometry of Data

Abstract

Structured Tensors and the Geometry of Data

by

Anna Leah Seigal

Doctor of Philosophy in Mathematics

University of California, Berkeley

Professor Bernd Sturmfels, Chair

We analyze data to build a quantitative understanding of the world. Linear algebra is the foundation to algorithms, dating back one hundred years, for extracting structure from data. Modern technologies provide an abundance of multi-dimensional data, in which multiple variables or factors can be compared simultaneously. To organize and analyze such datasets we can use a tensor, the higher order analogue of a matrix. However, many theoretical and practical challenges arise in extending linear algebra to the setting of tensors.

In the first part of this thesis, I study and develop the algebraic theory of tensors. Tensors of low real rank, as well as singular vectors and singular values of tensors, parametrize semi-algebraic sets, defined by polynomial equations and inequalities. I give exact algebraic characterizations of sets of tensors of interest, using real algebraic geometry, polyhedral geometry and computer algebra. I obtain a membership test for the set of real rank two tensors, I describe the variety of singular vectors of orthogonally decomposable tensors, and I obtain inequalities relating the singular values of the flattenings of a tensor. I show that rank and symmetric rank coincide for tensors of low symmetric rank, bounded by seven. I conclude by describing tensor hypernetworks, a flexible framework for decomposing and approximating tensors in application-specific contexts. Throughout these results, an in-depth study of small tensors is used to set up the general theory. The theoretical results explain pitfalls of existing tensor algorithms, and also suggest new approaches for finding structure in a tensor.

In the second part of this thesis, I present three algorithms for tensor data. The algorithms use algebraic and geometric structure to give guarantees of optimality. Tensors have a close connection to multivariate distributions in statistics. I obtain the first non-trivial instance of an exact maximum likelihood estimate for a model with hidden variables. I give a numerical algorithm to recover paths from their third order signature tensors, an inverse problem from stochastic analysis. I also give an algorithm to cluster multi-dimensional data, with structured clusters encoded via algebraic constraints in a tensor. The structure facilitates the interpretation of clusters in a dataset of cancer cell lines.

Contents

List of Figures

List of Tables

Acknowledgments

I would like to thank my advisor Bernd Sturmfels for his generosity with his time, his boundless enthusiasm, and his reliable practical advice. Thank you Lior Pachter for the enjoyable semester I spent with your group, and thank you Caroline Uhler and Lek-Heng Lim for your guidance throughout my time in grad school. Thank you to my collaborators and friends: Hirotachi Abo, Carlos Améndola, Miriam Barlow, Mariano Beguerisse-Díaz, Heather Harrington, Kathlén Kohn, Portia Mira, Guido Montúfar, Max Pfeffer, and Elina Robeva. I learned a huge amount from working with you. Thank you David Eisenbud, Jim Demmel, and Hernan Garcia for taking the time to be on my thesis committee. Thank you Lynn Frank for nourishing my early interest in maths, and Mike Levin, Denise Sheer, and Sergei Yakovenko for facilitating my positive early experiences of research.

I would like to thank my friends, especially Aliya, Charlie, Rachael, Isaac, Joe, Vino, and Alex, who traveled a long way to visit me in California. Thank you Ceci and Henry who first made San Francisco a home away from home. Thank you Noble and Salil for wonderful trips to Yosemite. Thank you Anna, Sasha, and Yael for sharing the twists and turns of Berkeley life with me. Thank you James Walsh and Maddie Brandt for being the best office mates I could ask for. Thank you Vicky, Judie and Marsha for your help with every administrative hurdle.

An enormous thank you to my parents for your encouragement, for the opportunities you have made available to me, and for your unstoppable enthusiasm to visit. Thank you to my brother Louis for your loyalty and thoughtfulness, and for making me smile from the beginning. Thank you Adele for being an inspiring creative force and for your sustained inquisitiveness in me and everything. Thank you Kurt for your enthusiasm for academic pursuits and your interest in my research. Thank you Elsa for your practical perspective and for the creative skills you taught me. Thank you Harry for the reverberations of your kindness and consideration. Thank you to my cousins for having me to stay and for visiting me. Thank you to Salil's family for sharing your wisdom and life perspective with me, and for supporting my dreams. Most of all, thank you Salil for the inspiring discussions, for the adventures so far, and for the ones to come.

Chapter 1

From matrices to tensors

In this chapter, I describe key ingredients in extending the theory of linear algebra to the multi-dimensional setting of tensors. I describe the importance of matrices with a focus on the singular value decomposition, arguably the central result in linear algebra, and I give proofs of results in linear algebra, which give a preview to the tensor methods of later chapters. I describe similarities and differences between matrices and tensors, and highlight the algebraic structure of small tensors as a case study to motivate the general theory.

But first, I would like to welcome you to my thesis. Whether you are a mathematician or not, a practitioner in the study of multi-dimensional data or just looking for a statement of new results, thank you for being here. Please see below for an introduction to this thesis that might be catered to your interest.

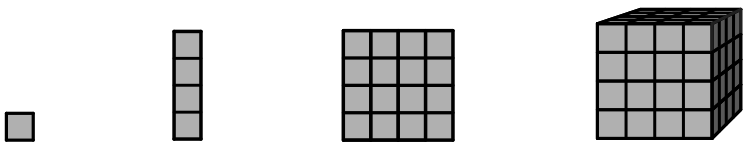


Figure 1.1: Scalar, vector, matrix, tensor.

For everyone: The matrix is a two-dimensional grid of numbers, and linear algebra is the theory of matrices. Matrix data allows the comparison of two changing variables or factors: one for the rows and one for the columns. Humans are quite good at understanding the relationship between two variables: be it correlation, causation, independence, or something else. But modern datasets can compare multiple variables simultaneously, and this presents the opportunity to build an understanding of complex systems. Such data can be organized into a tensor, a grid of numbers with a larger number of dimensions. It is more difficult to make sense of the relationship between three or more variables simultaneously, i.e. to extract interpretable structure from tensors of data. The linear algebra theory that was foundational to the study of matrix data cannot be directly applied in the higher dimensional setting of tensors. This hinders design of tensor algorithms, and consequently limits our ability to

detect higher-order structure in data. In this thesis, I study the theory of tensors, and use it to design algorithms for finding structure in tensor data.

For a mathematician: This thesis is in applied algebra. I study the structure of tensors, using tools from real algebraic geometry, polyhedral geometry and computer algebra. I discuss how linear algebra extends to the multi-linear (i.e. non-linear) setting of tensors, with a focus on the exact algebraic characterization of sets of tensors of interest. I apply my theoretical results to tensors of data, via exact and numerical algorithms. To place this thesis in the context of the theory of tensors, I describe the three main textbooks in the area. The first, [105] from 2012, focuses on using methods from algebraic geometry and representation theory to find the rank of a tensor with complex entries, and discusses applications to complexity theory. The textbook [77], also from 2012, outlines the functional analysis of tensors and gives a numerical treatment of hierarchical tensor decompositions, with applications to solving partial differential equations. The more recent textbook [145], from 2017, studies the spectral theory of tensors, with a focus on eigenvectors and eigenvalues, with applications to hypergraph theory, and discusses notions of positivity for tensors.

For a practitioner: There is a range of algorithms in data analysis that are based around the tensor. There are specialized numerical toolboxes for tensors, such as the MATLAB Tensor Toolbox [15] and Tensorlab [182]. Moreover, the machine learning software library TensorFlow organizes its data structures around tensors [180]. Tensor methods have been successfully applied to applications in computer science, statistics, physics, and biology, as we will see in this thesis. However, tensor algorithms suffer from a lack of interpretability and most are not guaranteed to find the global optimal solution to an optimization problem. In this thesis, we will see how the algebraic and geometric structure of tensors can be used to find interpretable signals in tensor data, and to understand and overcome practical challenges in tensor algorithms. We will see examples of tensor algorithms with guarantees, and apply them to real and simulated data.

For a list of new results: A statement of the contributions made in this thesis, with references to published articles, is given in Section 1.6.

1.1 What is a tensor?

In this section, I give two definitions of a tensor and discuss why they are equivalent. I define the multiplication of tensors and the rank of a tensor. I give examples of tensors arising in various contexts that I return to later in this thesis.

A tensor is a grid of numbers organized by multiple indices. Each entry of the tensor is specified by fixing values for the indices. In this thesis, a tensor will usually be denoted by the letter X. The tensor has entries $x_{i_1 \cdots i_d}$ that are numbers, specified by fixing a value for each index $i_1, \ldots, i_d$.

The order of a tensor is the number of indices. For example, a vector with entries x_i is a tensor of order one. A matrix with entries x_{ij} is a tensor of order two. An order three tensor has entries x_{ijk} given by three indices. See Figure 1.1 for a cartoon of an order zero, one, two, and three tensor: each small cube in the figure represents one entry of the tensor.

The format of a tensor is a product of numbers, giving the range of values that each index can take. For example, the tensor X has format $4 \times 4 \times 4$ if its entries are x_{ijk} for $i, j, k \in \{1, 2, 3, 4\}$, see the right of Figure 1.1. A general tensor X, with entries $x_{i_1,\ldots,i_d}$, has format $n_1 \times \cdots \times n_d$ if the index i_j lies in the set $\{1, \ldots, n_j\}$ for all indices $j \in \{1, \ldots, d\}$.

Tensors are not just grids of numbers. Like matrices, they are algebraic objects. We can multiply tensors by vectors, matrices, and other tensors of compatible formats. For example, we can multiply a tensor X of format $n_1 \times n_2 \times n_3$ by a tuple of matrices $(A^{(1)}, A^{(2)}, A^{(3)})$, where the matrix $A^{(i)}$ has format $m_i \times n_i$ and (j, k) entry denoted by $a^{(i)}_{jk}$. We obtain a new tensor of format $m_1 \times m_2 \times m_3$ which I denote by $[\![X; A^{(1)}, A^{(2)}, A^{(3)}]\!]$. See Figure 1.2 for a cartoon of this tensor matrix multiplication. The (i, j, k) entry of the new tensor is

$$\sum_{\gamma=1}^{n_3} \sum_{\beta=1}^{n_2} \sum_{\alpha=1}^{n_1} x_{\alpha\beta\gamma} a^{(1)}_{i\alpha} a^{(2)}_{j\beta} a^{(3)}_{k\gamma}. \tag{1.1}$$

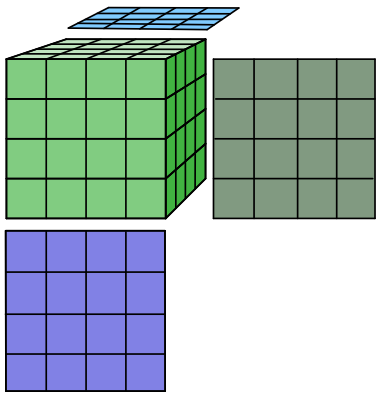


Figure 1.2: Multiplying a tensor by a tuple of matrices.

Important structure of a tensor is often unchanged under multiplication by a certain class of matrices. For example, in Chapter 2 we study the real rank of a tensor, unchanged under multiplication by real invertible matrices. In Chapter 4, we study singular values of tensors, unchanged under multiplication by orthogonal matrices. In Chapter 8 we study how to recover the matrix A from the tensor $[\![X; A, A, A]\!]$.

The tensor matrix multiplication in Equation (1.1) can be extended to multiplication along other combinations of indices. We can represent a large tensor of interest by multiplying together an interrelated collection of smaller tensors. This leads to tensor hypernetworks, which represent tensors based on the adjacency of hypergraphs, discussed in Chapter 6.

Tensors can also be defined as elements of a tensor product space.

Definition 1.1 (The tensor product)**.** *Let V_i be vector spaces over a field $\mathbb{K}$. The tensor product $V_1 \otimes \cdots \otimes V_d$ consists of the linear span over $\mathbb{K}$ of all elements*

$$v^{(1)} \otimes \cdots \otimes v^{(d)}, \qquad \text{where } v^{(i)} \in V_i.$$

The tensor product notation $\otimes$ denotes the quotient of ordered tuples $(v^{(1)}, \ldots, v^{(d)})$, by relations

$$(\lambda v^{(1)}) \otimes v^{(2)} \otimes \cdots \otimes v^{(d)} = \lambda(v^{(1)} \otimes v^{(2)} \otimes \cdots \otimes v^{(d)}) \tag{1.2}$$

and

$$(v^{(1)} + w^{(1)}) \otimes v^{(2)} \otimes \cdots \otimes v^{(d)} = v^{(1)} \otimes v^{(2)} \otimes \cdots \otimes v^{(d)} + w^{(1)} \otimes v^{(2)} \otimes \cdots \otimes v^{(d)} \tag{1.3}$$

where the vector $v^{(i)} \in V_i$, the vector $w^{(1)} \in V_1$, and the scalar $\lambda \in \mathbb{K}$. The analogous equations on the other vector space $V_2, \ldots, V_d$, also hold.

The tensor product in Definition 1.1 relates to the definition of a tensor as a grid of numbers by fixing a basis. Combining bases $\{e^{(i_j)}\}$ for each vector space V_j gives a basis of $V_1 \otimes \cdots \otimes V_d$. Denote the coefficient of X with respect to the basis vector $e^{(i_1)} \otimes \cdots \otimes e^{(i_d)}$ by $x_{i_1 \ldots i_d}$. If the vector space V_j has dimension n_j, then the index i_j runs over the set $\{1, \ldots, n_j\}$ and we obtain a grid of numbers of format $n_1 \times \cdots \times n_d$. Equation (1.2) means that scaling each row of an tensor by λ has the effect of scaling the entire tensor by λ. Equation (1.3) considers two tensors whose slices, obtained by fixing one index, are scalar multiplies of each other. The sum of the two tensors is a new tensor, obtained by summing the scalar multiples. The relations in Equations (1.2) and (1.3) hold for grids of numbers. Conversely, these conditions are all that distinguishes a tensor from an abstract sum of tuples of vectors.

Definition 1.2 (Rank)**.** *A tensor in $V_1 \otimes \cdots \otimes V_d$ has* rank one *if it can be written as*

$$v^{(1)} \otimes \cdots \otimes v^{(d)}, \quad \text{where } v^{(j)} \in V_j.$$

A general tensor X is a sum of rank one tensors

$$X = \sum_{i=1}^{r} v_i^{(1)} \otimes \cdots \otimes v_i^{(d)}, \quad \text{where } v_i^{(j)} \in V_j.$$

The smallest number of rank one tensors that sum to X is called the rank *of X.*

Definition 1.2 specializes to the usual matrix rank in the case $d = 2$. Tensors, and tensor products, arise in many contexts across mathematics. Definition 1.1 can be extended to tensor products of structures other than vector spaces.

I now give three examples of tensors arising in the study of multi-dimensional data, which I will return to in the second part of this thesis. I also describe how a familiar function in mathematics – a polynomial – can be viewed as a tensor.

Example 1.3 (A tensor of probabilities)**.** *In April 2017 there was a workshop on Algebraic Statistics at the Mathematisches Forschungsinstitut Oberwolfach (MFO). I wrote an article for the Oberwolfach Snapshot series, in which I analyzed survey data from the participants of the workshop using algebraic statistical methods [162].*

At the workshop, the weather started off cold with intermittent rain showers. The middle of the week saw snow and hail, and there were two sunny days at the end. Fifty out of the 52 participants at the workshop responded to a survey, detailing if they liked the weather, if it was their first time visiting Oberwolfach, and if they played a game during their stay (the favorite games were Carom Billiards, Hanabi, and Resistance). Here were the responses as proportions.

		First time at MFO	*Visited MFO before*
Games	*Liked weather*	*0.24*	*0.1*
	Disliked weather	*0.1*	*0.08*
No games	*Liked weather*	*0.14*	*0.18*
	Disliked weather	*0.06*	*0.1*

There are eight entries in the table, the joint probability distribution (or probability mass function) of three binary random variables

$$X = \begin{cases} 0 & \textit{liked weather,} \\ 1 & \textit{disliked weather,} \end{cases}$$

$$Y = \begin{cases} 0 & \textit{not visited MFO before,} \\ 1 & \textit{visited MFO before,} \end{cases}$$

$$Z = \begin{cases} 0 & \textit{played a game,} \\ 1 & \textit{played no game.} \end{cases}$$

For example, a workshop participant selected uniformly at random was at MFO for the first time, played a game, and enjoyed the weather with probability 0.24*. A random participant liked the weather with probability* $0.24 + 0.1 + 0.14 + 0.18 = 0.66$*.*

We can represent the probability distribution by a $2 \times 2 \times 2$ *tensor*

$$P = \left[\begin{array}{cc||cc} p_{000} & p_{010} & p_{001} & p_{011} \\ p_{100} & p_{110} & p_{101} & p_{111} \end{array}\right] = \left[\begin{array}{cc||cc} 0.24 & 0.1 & 0.14 & 0.18 \\ 0.1 & 0.08 & 0.06 & 0.1 \end{array}\right],$$

where the entry p_{ijk} *of the tensor* P *represents the probability* $P(X = i, Y = j, Z = k)$*. Properties of the tensor* P*, such as its non-negative rank, translate to statistical properties of the distribution. I will return to algebraic methods to analyze probability distributions in Chapter 7. See also [162, 163].*

Example 1.4 (A tensor of biological data)**.** *Modern biological experiments seek to understand the relation between multiple changing variables or factors. For example, the data set*

depicted in Figure 1.3, and introduced in [131], describes the response of breast cancer cell lines to different ligands. The measurements of the experiment are the temporal phosphorylation levels of two proteins that are involved in cellular decisions and fates. One goal of studying data such as this is drug discovery. The ligands can be thought of as experimental conditions, or stand-ins for possible drugs, and the cell lines can be thought of as patients for whom the drug may be suitable.

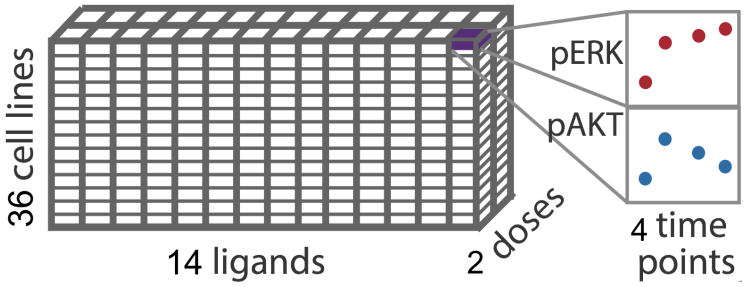


Figure 1.3: A tensor of biological experiments measuring the response of breast cancer cell lines to ligands.

The dataset is a tensor of order 5 and format $36 \times 14 \times 2 \times 4 \times 2$*, corresponding to the 36 cell lines, 14 ligands, 2 doses, 4 time points, and 2 proteins. Properties of the tensor reflect structure in the data, which we hope to interpret biologically. I will return to this data set in Chapter 9. See also [166].*

Example 1.5 (A tensor of integrals)**.** *Consider the 'skyline path', a piecewise linear path in* $\mathbb{R}^2$ *with steps given by the columns of the matrix*

$$A = \begin{bmatrix} 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & -1 & 0 & 2 & 0 & -2 & 0 & 1 & 0 & -1 & 0 \end{bmatrix}.$$

A path can be encoded by its signature [115], an infinite series of tensors whose entries are iterated integrals in the coordinates of the path. For example, the third order signature tensor X *has entries*

$$x_{ijk} = \int_0^1 \left(\int_0^{t_3} \left(\int_0^{t_2} \mathrm{d}\psi_i(t_1) \right) \mathrm{d}\psi_j(t_2) \right) \mathrm{d}\psi_k(t_3) \quad \text{for } 1 \le i, j, k \le 2.$$

Evaluating these integrals for the skyline path gives the $2 \times 2 \times 2$ *tensor*

$$\frac{1}{6} \left[\begin{array}{cc||cc} 343 & 0 & -84 & 18 \\ 84 & 18 & -36 & 0 \end{array} \right].$$

Many other paths will have the same third order signature tensor as the skyline path. Some examples are shown in Figure 1.4. The shortest path is approximated by taking a length-constrained piecewise linear path with a large number of steps.

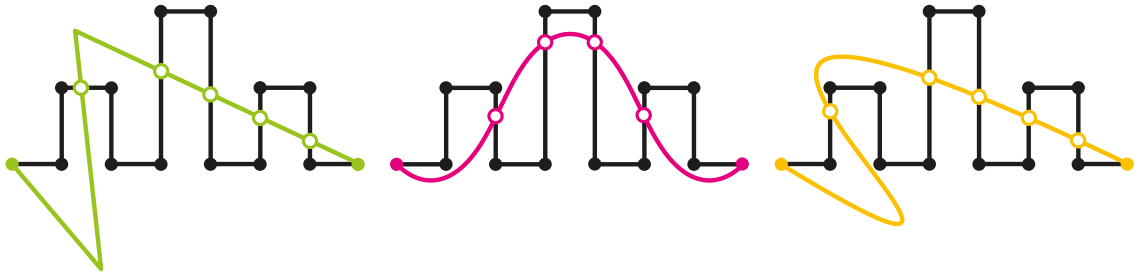


Figure 1.4: Three paths with the same third order signature tensor as the skyline path (black), a path with three steps (left), the shortest path (middle), and a polynomial path of degree three (right).

The first order signature tensor encodes the increment of the path, the endpoint minus the start-point. The second order signature tensor encodes the signed area between the path and the segment connecting the endpoints. A direct geometric interpretation for the higher order signature tensors is not known, although they can be thought of as certain iterated areas [60]. Nonetheless, there is a well-understood equivalence class of tensors whose infinite series of signature tensors are the same [47, 79]. In Chapter 8, I describe identifiability properties and recovery algorithms for learning a path from its third order signature tensor. See also [141].

There are many other contexts and applications in which tensors arise that we will see throughout this thesis. I conclude this section by showing that polynomials can be viewed as symmetric tensors. But first, I describe the notation I use in this thesis.

Tensors are denoted by capital letters, usually X, and their entries are lowercase letters $x_{i_1,\dots,i_d}$. Indices are usually denoted with the letters i and j, with the boldface $\mathbf{i}$ to denote a tuple of indices $(i_1,\dots,i_d)$. When possible, I will denote the indices by single letters, e.g. (i,j,k) instead of (i_1,i_2,i_3). Otherwise the indices themselves have subscripts, and a tower of multiple $i_{n_{d_{i_{ces}}}}$ can make formulae more difficult to read. I say that a tensor X has format $n \times \cdots \times n$ (d times) as a simpler notation for d factors that are all of dimension n,

$$\underbrace{n \times \cdots \times n}_{d \text{ times}},$$

i.e. $X \in V^{\otimes d}$, where V is a vector space of dimension n (usually $\mathbb{C}^n$ or $\mathbb{R}^n$). Matrices are also denoted by capital letters, usually M, and their entries are m_{ij}. Vectors are denoted by lowercase letters, e.g. v with entries v_i. Scalars are also lowercase letters, usually from the Greek alphabet. For a collection of vectors that range across an indexing set, I will use the notation $v^{(i)}$, $i \in \{1,\dots,n\}$. If two indexes are required, e.g. in a tensor decomposition, I will use $v_i^{(j)}$. For $X \in V_1 \otimes \cdots \otimes V_d$, I call d the order of the tensor and not the dimension,

to distinguish it from the dimension of the vector spaces V_i. When $V_i = \mathbb{K}^{n_i}$ I abbreviate the tensor product space to $\mathbb{K}^{n_1 \times \cdots \times n_d}$. When displaying the entries of a tensor, the double line $\|$ denotes the separation between the slices.

The multiplication of the tensor X by the vector $v^{(j)}$ along the jth index for all $j \in \{1, \ldots, d\}$ is denoted by $[\![X; v^{(1)}, \ldots, v^{(d)}]\!]$, and similarly for multiplication by matrices, following the notation in [93]. Multiplication is sometimes carried out in some indices but not in others. Indices in which no multiplication occurs are denoted by a "$\cdot$" in that location, e.g. $[\![X; \cdot, \ldots, \cdot, v]\!]$ denotes the multiplication of the tensor X by the vector v along the dth index. Calligraphic letters $\mathcal{X}, \mathcal{M}$ etc. are for algebraic objects such as semi-algebraic sets and varieties.

Definition 1.6 (Symmetric tensors)**.** *A tensor X of format $n \times \cdots \times n$ (d times) is* symmetric *if its entries are unchanged under permuting the indices, i.e. $x_{\mathbf{i}} = x_{\sigma(\mathbf{i})}$ for permutations σ of the tuple of indices $\mathbf{i} = (i_1, \ldots, i_d)$.*

For example, a matrix $M \in \mathbb{K}^{n \times n}$ is symmetric if $M = M^{\mathsf{T}}$, or $m_{ij} = m_{ji}$ for all $i, j \in \{1, \ldots, n\}$. An order three tensor $X \in \mathbb{K}^{n \times n \times n}$ with entries x_{ijk} is symmetric if

$$x_{ijk} = x_{jik} = x_{kji} = x_{ikj} = x_{jki} = x_{kij}, \qquad \text{for all } i, j, k \in \{1, \ldots, n\}. \tag{1.4}$$

Example 1.7 (Polynomials are symmetric tensors)**.** *A polynomial in n variables is a finite sum of monomials,*

$$f(z_1, \ldots, z_d) = \sum_{j_1, \ldots, j_d} c_{j_1, \ldots, j_d} z_1^{j_1} \cdots z_d^{j_d}.$$

Like a tensor, a polynomial has multi-indexed structure.

A symmetric matrix $M \in \mathbb{K}^{n \times n}$ encodes a quadratic form, a polynomial of degree two. The quadratic form is obtained from the matrix by $M \mapsto z^{\mathsf{T}} M z$, where z is the vector of variables $(z_1, \ldots, z_n)$. Similarly, symmetric tensors of format $n \times n \times \cdots \times n$ (d times) are in bijection with homogeneous polynomials of degree d in n variables. The bijection is obtained by multiplying the tensor on each side by the vector of variables z,

$$X \qquad \leftrightarrow \qquad f(z_1, \ldots, z_n) = [\![X; z, \ldots, z]\!] = \sum_{i_1 \ldots, i_d = 1}^{n} x_{i_1 \ldots i_d} z_{i_1} \cdots z_{i_d}. \tag{1.5}$$

The entries of the tensor combine to make the coefficients of the polynomial. Since the tensor is symmetric, this process can be inverted to recover the tensor from the polynomial. Under the correspondence between symmetric tensors and polynomials, we can use tensors as a lens through which to study questions from classical algebraic geometry, see Chapter 5.

1.2 The singular value decomposition

Linear algebra gives decomposition theorems for matrices that allow structure to be extracted from matrix data. A complicated matrix can be approximated by a simpler one

that captures the most important information. In this section, I discuss the singular value decomposition (SVD), the eigen-decomposition, and principal component analysis (PCA). This sets the scene for studying methods to extract structure from tensors in later chapters. I give a proof of the existence of the SVD of a matrix and the Eckart-Young theorem on low rank approximation. I prove these results using Lagrange multipliers, which reduce the optimization problem to a system of polynomial equations. This algebraic formulation is a preview to the algebraic approaches to tensor optimization that I describe in later chapters. I also discuss how polynomials, such as the characteristic polynomial, give an algebraic encoding of the structure of a matrix, a preview to the algebraic approaches to tensor structure that I describe in later chapters.

The SVD is a central result from linear algebra, the key to practical methods to find low rank structure in matrices. The matrices (and tensors) that arise in data applications often have real entries. I begin this section by recalling the existence of the SVD for a real matrix.

Theorem 1.8 (The singular value decomposition)**.** *A real matrix M of format $n_1 \times n_2$ can be written in the form $M = U\Sigma V^{\mathsf{T}}$ or, in tensor product notation,*

$$M = \sum_{i=1}^{n} \sigma_i u^{(i)} \otimes v^{(i)}, \tag{1.6}$$

where the matrix U is orthogonal of format $n_1 \times n_1$ with ith column equal to $u^{(i)}$, the matrix V is orthogonal of format $n_2 \times n_2$ with ith column equal to $v^{(i)}$, and the matrix Σ is diagonal of format $n_1 \times n_2$ with non-negative diagonal entries $\sigma_1 \geq \ldots \geq \sigma_n$, where $n = \min\{n_1, n_2\}$.

The vectors $u^{(i)}$ and $v^{(i)}$ are called the ith left and right singular vectors of M, respectively, and the scalar σ_i is the ith singular value. Equation (1.6) implies the following equations for a matrix and its singular vectors and singular values:

$$Mv^{(i)} = \sigma_i u^{(i)} \quad \text{and} \quad M^{\mathsf{T}} u^{(i)} = \sigma_i v^{(i)}. \tag{1.7}$$

This is an alternative definition of the singular vectors and singular values of a matrix M: the vectors u and v are left and right singular vectors with singular value σ if

$$Mv = \sigma u \quad \text{and} \quad M^{\mathsf{T}} u = \sigma v. \tag{1.8}$$

A third way to define the singular vectors and singular values of a matrix is via rank one approximation. A rank one matrix has the form $\sigma u \otimes v$ where u and v are vectors of norm one, $\|u\|^2 = \sum_{i=1}^{n} u_i^2 = 1$, and σ is a scalar which can be chosen to be non-negative. Consider the closest rank one matrix to a given matrix M, with respect to the Euclidean norm. The best rank one approximation can be found by maximizing the inner product $\langle M, u \otimes v \rangle$ over vectors u and v of norm one. The scalar multiple σ is the inner product $\langle M, u \otimes v \rangle$. Hence the best rank one approximation is the global maximum of the optimization problem

$$\text{maximize}_{\,u,v} \langle M, u \otimes v \rangle \qquad \text{subject to } \frac{1}{2}(1 - \|u\|^2) = 0 \text{ and } \frac{1}{2}(1 - \|v\|^2) = 0. \tag{1.9}$$

In particular, the best rank one approximation is one of the critical points of this optimization problem, where the partial derivatives with respect to all variables vanish. We can find the critical points using Lagrange multipliers. We introduce auxiliary variables λ_1 and λ_2, and define the functional

$$L(u, v, \lambda_1, \lambda_2) = \sum_{i,j} M_{ij} u_i v_j + \frac{\lambda_1}{2} \left(1 - \sum_i u_i^2\right) + \frac{\lambda_2}{2} \left(1 - \sum_i v_i^2\right).$$

The critical points occur when the partial derivatives of L with respect to the variables $u, v, \lambda_1, \lambda_2$ vanish. The vanishing of the partial derivatives gives the system of equations

$$\textstyle\sum_j M_{ij} v_j = \lambda_1 u_i \left(\sum_j v_j^2\right) = \lambda_1 u_i, \qquad \sum_i M_{ij} u_i = \lambda_2 v_j \left(\sum_i u_i^2\right) = \lambda_2 v_j.$$

Hence we have the conditions $Mv = \lambda_1 u$ and $M^\mathsf{T} u = \lambda_2 v$. Since $\lambda_1 = u^\mathsf{T} M v = v^\mathsf{T} M^\mathsf{T} u = \lambda_2$, the two auxiliary variables λ_i are equal, and we obtain the condition from Equation (1.8) for (u, v) to be a singular vector pair. The best rank one approximation is obtained by choosing the critical point whose inner product $\langle M, u \otimes v \rangle$ is largest.

We can use the critical points of the distance to the rank one matrices to give a proof of the SVD, as follows. We assume for convenience that the matrix M is generic and square.

Proof of Theorem 1.8. Construct the critical points of the optimization function in Equation (1.9). There are finitely many critical points, given by vectors $(u^{(i)}, v^{(i)})$. Fix the signs of vectors $u^{(i)}$ and $v^{(i)}$ such that the inner product $\sigma_i := \langle M, u^{(i)} \otimes v^{(i)} \rangle$ is non-negative. Concatenate the vectors $u^{(i)}$ to form the columns of the matrix U, concatenate the vectors $v^{(i)}$ to form the columns of V, and set σ_i to be the diagonal entries of Σ. Since Equation (1.7) holds for all tuples $(u^{(i)}, v^{(i)}, \sigma_i)$, we have

$$MV = U\Sigma, \quad M^\mathsf{T} U = V\Sigma.$$

The vector $u^{(i)}$ is an eigenvector of MM^T with eigenvalue σ_i^2. Since M is generic, the eigenvalues are distinct and strictly positive, so the vectors $u^{(i)}$ are linearly independent, and hence U is invertible. Rearranging $M^\mathsf{T} U = V\Sigma$ gives $M = U^{-1}\Sigma V^\mathsf{T}$ and hence $VV^\mathsf{T} = I$. Then $MV = U\Sigma$ implies that $M = U\Sigma V^\mathsf{T}$. □

If two or more singular values of a matrix take the same value, say $\sigma_i = \sigma_{i+1} = \cdots = \sigma_j$, the SVD still exists but is not unique. There are infinitely many singular vector pairs, given by the choices of orthogonal bases for the vector spaces $\langle u^{(i)}, u^{(i+1)}, \ldots, u^{(j)} \rangle$ and $\langle v^{(i)}, v^{(i+1)}, \ldots, v^{(j)} \rangle$.

The uses of the SVD are extensive, as we will see later in this section. The decomposition possesses several useful properties. The rank of a matrix M can be read directly from the SVD: it is the number of non-vanishing singular values. A matrix of higher rank can be approximated by one of lower rank using the SVD, as follows.

Theorem 1.9 (The Eckart-Young theorem [63]). *The best rank r approximation to a matrix is given by truncating its SVD to the top r singular values.*

The Eckart-Young theorem is hugely important for computing low rank approximations of matrices. The theorem implies that the best rank rank r approximation of a matrix can be found by solving r successive best rank one approximation problems, and this is helpful in practice. For a numerical treatment of matrices and linear algebra, see [58]. For more historical details on the singular value decomposition, see [176]. The Eckart-Young Theorem does not hold in general for tensors, as we will see later in this chapter. This is one reason behind problems to compute low rank decompositions or approximations of a tensor: while efficient algorithms exist for rank one approximation, see [44], they cannot be extended to give optimal approximations of higher rank.

Proof of Theorem 1.9. We seek a rank r approximation of a matrix M. A matrix of rank r possesses an SVD, and the set of matrices of rank at most r is closed. This means the best rank r approximation will be a critical point of the functional

$$\|M - \sum_{i=1}^{r} \gamma_i a^{(i)} \otimes b^{(i)}\|^2 + \sum_{i=1}^{r} \alpha_i \left(\|a^{(i)}\|^2 - 1\right) + \sum_{i=1}^{r} \beta_i \left(\|b^{(i)}\|^2 - 1\right),$$

where we can assume that the vectors $a^{(i)}$ satisfy $\langle a^{(i)}, a^{(j)} \rangle = \delta_{ij}$, the $b^{(i)}$ satisfy $\langle b^{(i)}, b^{(j)} \rangle = \delta_{ij}$, and the γ_i are non-negative. The orthogonality condition means when we take the partial derivatives of the functional, we obtain the condition that the pairs of vectors $(a^{(i)}, b^{(i)})$ are singular vectors of M. Hence the critical points are given by linear combinations of r tensor products of singular vectors. From among the critical points, choosing the singular vectors corresponding to the r largest singular values minimizes the distance to the original matrix. □

In light of the useful properties of the SVD, and of the prevalence of tensor data coming from applications, it is a topic of major interest to extend the SVD to tensors. It is arguably even more crucial to find a low rank approximation of a tensor than it is for a matrix: the higher order makes tensors in their original form especially computationally intractable. However, there are theoretical and practical challenges associated with extending the SVD to tensors.

1. *The set of low rank tensors may be not be closed.* This means the best low rank approximation of a tensor may not exist. If we consider the closest tensor in the closure of the low rank tensors, we can obtain a tensor of higher rank. Such situations are not boundary cases, but can occur with positive probability. I explore the geometry of the low real rank approximation problem for tensors in Chapter 2.

2. *Not all tensors are orthogonally decomposable.* The SVD writes a matrix as a sum of rank one matrices $\sigma_i u^{(i)} \otimes v^{(i)}$ such that the vectors $u^{(i)}$ are orthogonal and the vectors

$v^{(i)}$ are orthogonal. General tensors do not have a decomposition into rank one terms of orthogonal vectors. Tensors that can be written in this form are called orthogonally decomposable. I investigate the singular vectors of orthogonally decomposable tensors in Chapter 3.

3. *The best way to define the singular value of a tensor is not clear.* For the most practical definition, it is not known how to construct a tensor with prescribed singular values, see Chapter 4.

I now describe how the SVD can be specialized to symmetric matrices.

The eigen-decomposition

Recall that a real matrix M of format $n \times n$ is symmetric if $M = M^{\mathsf{T}}$. If a matrix is symmetric, a decomposition into symmetric rank one terms can be found,

$$M = V\Lambda V^{\mathsf{T}} = \sum_{i=1}^{n} \lambda_i v^{(i)} \otimes v^{(i)},$$

where V is an $n \times n$ orthogonal matrix with ith column equal to $v^{(i)}$, and the λ_i are real scalars. This is called the eigen-decomposition. The coefficient λ_i is the eigenvalue corresponding to eigenvector $v^{(i)}$.

The Lagrange multiplier construction of the best rank one approximation of a matrix can be applied to show that the best rank one approximation of a symmetric matrix is $\lambda_1 v^{(1)} \otimes v^{(1)}$, where λ_1 is the eigenvalue of largest magnitude and $v^{(1)}$ its eigenvector. Since the best rank r decomposition can be found by successively finding the best rank one approximation, the best rank r approximation is given by truncating the eigen-decomposition to the singular vectors corresponding to the r eigenvalues of largest magnitude.

The eigen-decomposition shows that the best rank r approximation of a symmetric matrix can be chosen to be symmetric, and that a decomposition of a rank r symmetric matrix can be chosen to be a sum of symmetric rank one terms. I study the analogous problem for tensors in Chapter 5.

Principal component analysis

Given a collection of n data points $x^{(i)}$ in $\mathbb{R}^p$, we can construct the $n \times p$ matrix X whose ith row is the vector $x^{(i)}$. The entries of $x^{(i)}$ are measurements with respect to p different sensors. Finding a low rank approximation of the data matrix X gives us information about the structure that is present in the data. For example, a rank one approximation gives the uncoupling of the data points from the sensors that best approximates the data. The best rank r approximation gives the most accurate approximation of the data that decomposes it into r uncoupled terms.

Often, a low rank approximation is not sought of the matrix X directly, but rather of the $p \times p$ matrix $X^\mathsf{T}X$. If the rows of X are mean-centered, then $X^\mathsf{T}X$ is proportional to the empirical covariance matrix of the data. The eigenvectors of $X^\mathsf{T}X$ are directions (called principal components) which are important for the data: the eigenvectors corresponding to large eigenvalues are directions along which a large proportion of the variation between the sensors is exhibited.

Principal component analysis (PCA) finds a low rank approximation of $X^\mathsf{T}X$. The rank one terms are linear combinations of sensors that capture the variation in the data. PCA is based on the idea that we seek a sensor that discriminates between data points: the method combines existing sensors to obtain new coordinates, with respect to which there is large variability in the data.

The matrix $X^\mathsf{T}X$ is symmetric and positive semi-definite, so the eigen-decomposition exists and the eigenvalues are non-negative. The eigenvector $v^{(i)}$ corresponding to the ith largest eigenvalue is called the ith principal component.

The idea of projecting data to its largest principal components is widespread, appearing in applications ranging from biological data analysis to political advertising to movie preferences. For example, the algorithm used by a search engine to measure the importance of webpages uses a principal component, and there are estimated to be over 70,000 internet searches completed each second [88, 87].

On a personal level, my first encounter with PCA was the paper [134] from 2008. The authors plot the top two principal components of genetic data from individuals of European ancestry. The principal components are seen to be a close fit to a geographic map of Europe. That is, the linear combinations of genetic information that maximize variance are, in this setting, a good proxy for the spacial coordinates of a person's ancestry.

Matrix structure via polynomials

We can compute polynomials in the entries of a matrix, such as $\det(M) = m_{00}m_{11} - m_{01}m_{10}$ for a matrix M of format 2×2. These polynomials encode the algebraic structure of the matrix. A matrix is rank deficient if and only if its determinant vanishes. A matrix has rank at most $r-1$ if and only if all $r \times r$ minors vanish. The characteristic polynomial of the Gram matrix $M^\mathsf{T}M$ is

$$p(t) = \det(M^\mathsf{T}M - tI).$$

Another exact rank test for a matrix, rather than checking all minors of given size, is obtained from the coefficients of $p(t)$ considered as a univariate polynomial in t. The matrix M has rank at most r if and only if the coefficients of $t^{n-r}, \ldots, t, 1$ all vanish.

Recall that the rank of a matrix is also the number of non-zero singular values. The singular values of a matrix are orthogonal invariants, unchanged by orthogonal transformations. A basis of polynomial orthogonal invariants of M is given as follows. The coefficients of $p(t)$ are polynomial expressions in the entries of M. For example, the constant term is $\det(M^\mathsf{T}M) = \det(M)^2$. The polynomial $p(t)$ can also be factorized, using the singular values

of M, as

$$p(t) = \prod_{i=1}^{n} (t - \sigma_i^2).$$

From this factorization, we see that the coefficients of $p(t)$ are the elementary symmetric polynomials evaluated at the squares of the singular values. For example, the constant term is $\prod_{i=1}^{n} \sigma_i^2 = \det(M)^2$. Hence the coefficients of $p(t)$ are orthogonal invariants, and they are polynomials in the entries of M.

1.3 Similarities between matrices and tensors

From their definitions, matrices and tensors are similar: a matrix is an array of numbers in which each entry is described by two indices, while a tensor is an array of numbers in which each entry is described by any number of indices. It is unsurprising, then, that there are other similarities between them.

In this section, I discuss the main similarities between matrices and tensors: settings in which tensor structure can be studied using linear algebra. I begin by comparing change of basis operations for matrices and tensors, and discussing ways to consider a tensor as a multi-linear map, just like a matrix is a linear map. Then I study a tensor via its flattenings, the ways to reshape its entries into a matrix. It turns out that matrices and tensors can be studied similarly with regard to finding rank one structure. I prove the well-known extension of the Eckart-Young theorem for finding the best rank one approximation of a tensor. The theoretical and numerical properties of tensors differ from matrices at higher ranks, as I describe in Section 1.4.

Change of basis

For both matrices and tensors, it is helpful to apply a change of basis operation. For a matrix $M \in V_1 \otimes V_2$ we change basis by applying a linear transformation to the vector spaces V_i. In coordinates, we apply the transformation $M \mapsto [\![M; A^{(1)}, A^{(2)}]\!] = {A^{(1)}}^{\mathsf{T}} M A^{(2)}$. For an order three tensor X we change basis by applying a transformation $X \mapsto [\![X; A^{(1)}, A^{(2)}, A^{(3)}]\!]$ as in Equation (1.1). Similarly, we can extend change of basis operations to higher order tensors, $X \mapsto [\![X; A^{(1)}, \ldots, A^{(d)}]\!]$.

Given a matrix $M \in V_1 \otimes V_2$, we can find subspaces $W_i \subseteq V_i$ of minimal dimension such that $M \in W_1 \otimes W_2$. In bases, this translates to finding invertible matrices $A^{(i)}$ such that ${A^{(1)}}^{\mathsf{T}} M A^{(2)}$ has a block of non-zero entries of format $\dim(W_1) \times \dim(W_2)$, and all other entries zero. Similarly, the subspace representation writes a tensor as an array of smallest possible format. This is important for compressing the tensor.

Definition 1.10 (Tensor subspace representation, see [77, Chapter 8])**.** *Given a tensor $X \in V_1 \otimes \cdots \otimes V_d$, the subspace representation writes X as an element of the tensor product space $W_1 \otimes \cdots \otimes W_d$, where the vector subspaces $W_i \subseteq V_i$ are of minimal dimension. Consider a basis*

for V_i consisting of a basis for W_i plus a basis for its orthogonal complement. With respect to this basis, the tensor X has a block of non-zero entries of format $\dim(W_1) \times \cdots \times \dim(W_d)$, and all other entries zero.

Multi-linear maps

A matrix can be viewed as a linear or bilinear map. Consider the matrix $M \in V_1 \otimes V_2$, where the vector spaces V_i are over a field $\mathbb{K}$. The matrix M represents a linear map $V_2^* \to V_1$, or a linear map $V_1^* \to V_2$, or a bilinear map $V_1^* \otimes V_2^* \to \mathbb{K}$, where V_i^* denotes the vector space dual to V_i. Similarly, a tensor $X \in V_1 \otimes \cdots \otimes V_d$ represents various multi-linear maps on the vector spaces V_i and their duals, such as the linear map

$$\begin{array}{ccll} V_d^* & \to & V_1 \otimes \cdots \otimes V_{d-1} & \\ v & \mapsto & [\![X; \cdot, \ldots, \cdot, v]\!] & = \sum_j x_{i_1,\ldots,i_{d-1},j} v_j. \end{array}$$

or the linear map

$$\begin{array}{ccll} V_2^* \times \cdots \times V_d^* & \to & V_1 & \\ (v^{(2)}, \ldots, v^{(d)}) & \mapsto & [\![X; \cdot, v^{(2)}, \ldots, v^{(d)}]\!] & = \sum_{j_2,\ldots,j_d} x_{i,j_2,\ldots,j_d} v^{(2)}_{j_2} \cdots v^{(d)}_{j_d}. \end{array}$$

We study such maps in the context of singular vectors of tensors in Chapter 3. Interpolating between these examples, we see that a tensor can be viewed as many other multi-linear maps. If the number of possible values taken by an index of a tensor agrees with the length of a vector, we can multiply the tensor and the vector together, summing over that index.

Flattenings

The flattenings of a tensor are matrices, obtained by reshaping the tensor by reindexing. We can study some of the structure of a tensor via the linear algebra of its flattenings. I will first describe what it means to flatten a tensor into a matrix, with the aid of pictures, and I will then describe decompositions of tensors based on flattenings.

Recall that a matrix can be vectorized, or reshaped into a vector, as in Figure 1.5. We can choose to concatenate the rows of the matrix into a vector, or to concatenate the columns, and this gives two possible vectorizations of the matrix.

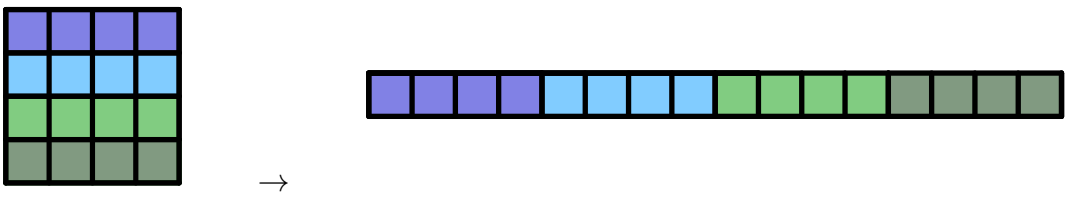


Figure 1.5: A matrix can be flattened into a vector.

A tensor has many possible flattenings. One way to flatten a tensor into a matrix is shown in Figure 1.6. Consider the tensor X with entries $x_{i_1,\dots,i_d}$ and format $n_1 \times \cdots \times n_d$. Each flattenings is given by a non-empty subset $S \subsetneq \{1,\dots,d\}$, as follows. We define two multi-indices $\mathbf{i} = (i_k : k \in S)$ and $\mathbf{j} = (i_k : k \notin S)$. We let the multi-indices $\mathbf{i}$ label the rows of the matrix, and the multi-indices $\mathbf{j}$ label the columns, and we obtain a matrix of format $\prod_{k\in S} n_k \times \prod_{k\notin S} n_k$. All entries of the original tensor X appear in the flattening: the $(\mathbf{i},\mathbf{j})$ entry of the matrix is $x_{i_1,\dots,i_d}$. For more details, see [77, Section 5.2].

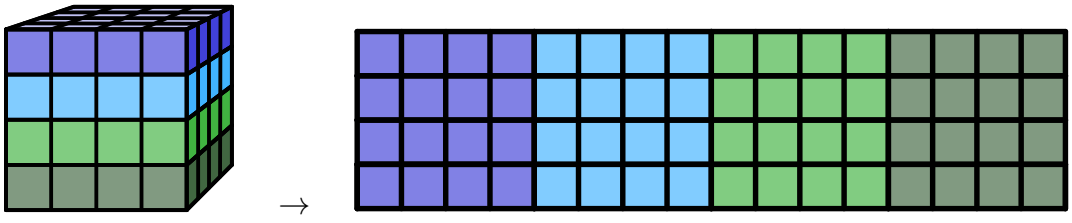


Figure 1.6: A tensor can be flattened into a matrix.

A tensor contains more information than any of its flattenings. For example, some pairs of rows of the flattening matrix are more similar than others, depending on the similarity of the multi-indices $\mathbf{i} = (i_1,\dots,i_d)$ and $\mathbf{i}' = (i'_1,\dots,i'_d)$ that label them. The two multi-indices could differ only at the dth index but be the same at all other indices. However, applying any linear algebra method to the matrix does not see the finer structure in the multi-indices. One approach to get around this problem is to consider several flattening matrices at once, which considers a tensor as an interrelated collection of matrices.

A popular choice of flattenings are called the *principal flattenings*. For a tensor of format $n_1 \times \cdots \times n_d$, the principal flattenings are d matrices, each of format $n_i \times \prod_{j\neq i} n_j$, which use a single index for the rows, and combine all other indices to form the columns. Setting r_i to be the rank of the ith principal flattening, we can write a tensor in the form

$$X = [\![C; A^{(1)},\dots,A^{(d)}]\!] = \sum_{j_1=1}^{r_1} \cdots \sum_{j_d=1}^{r_d} c_{j_1,\dots,j_d} a^{(1)}_{j_1 i_1} \cdots a^{(d)}_{j_d i_d}, \tag{1.10}$$

where C is a core tensor of format $r_1 \times \cdots \times r_d$. Equation (1.10) is the Tucker decomposition of X, see [77, Chapter 8]. Note the similarity with the subspace representation from Definition 1.10. The Tucker decomposition of a tensor represents it via the adjacency structure of a graph. In the order three case, the graph is shown in Figure 1.7. The central vertex corresponds to the core tensor, of format $m_1 \times m_2 \times m_3$. As the values of the m_i increase, more tensors can be represented by this tensor network. When $m_i = n_i$, all tensors factor according to the tensor network. For small values of m_i, we get a subvariety of tensors. For more on tensor networks, see Chapter 6.

A special choice of Tucker decomposition can be chosen, which imposes orthogonality on the matrices $A^{(i)}$. In this case the core tensor C has properties akin to matrix singular values, as I now describe.

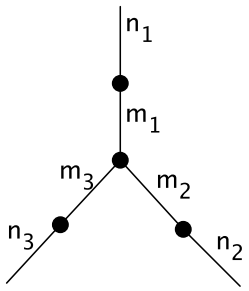


Figure 1.7: The Tucker decomposition, as a tensor network, for an order three tensor.

Definition 1.11 (The higher-order singular values [106])**.** *A tensor $X \in \mathbb{R}^{n_1 \times \cdots \times n_d}$ has d principal flattenings. Denote the ith flattening by $X^{(i)}$. It is a matrix of format $n_i \times \prod_{j \neq i} n_j$. The higher-order singular values of X are a list of vectors $\{\sigma^{(1)}, \ldots, \sigma^{(d)}\}$, where $\sigma^{(i)}$ is the vector of singular values of $X^{(i)}$, a vector of length n_i.*

Theorem 1.12 (The higher-order singular value decomposition [106, Theorem 2])**.** *A tensor $X \in \mathbb{R}^{n_1 \times \cdots \times n_d}$ can be written in the form*

$$X = \sum_{j_1=1}^{n_1} \cdots \sum_{j_d=1}^{n_d} c_{j_1,\ldots,j_d} a^{(1)}_{j_1 i_1} \cdots a^{(d)}_{j_d i_d},$$

where each $A^{(i)}$ is an orthogonal $n_i \times n_i$ matrix. Denote the ith principal flattening of the core tensor $C \in \mathbb{R}^{n_1 \times \cdots \times n_d}$ by $C^{(i)}$. Then each Gram matrix $C^{(i)}(C^{(i)})^{\mathsf{T}}$ is diagonal, with diagonal entries the squared singular values of the principal flattening $X^{(i)}$.

Definition 1.11 extends the notion of singular values from a matrix to a tensor. The higher-order singular values generalize many useful properties of the usual matrix singular values, see [106]. On the practical side, since the principal flattenings are matrices, the higher-order singular values can be computed using methods from linear algebra. However, the higher-order singular values also have some drawbacks.

The higher-order singular value decomposition does not give an optimal approximation of a tensor. If we keep only the top r_i singular values of the ith flattening, setting the others to zero, we obtain an approximation to a subspace representation of format $r_1 \times \cdots \times r_d$. The error introduced by this approximation can be given in terms of sums of squares of higher-order singular values, as in the matrix case. However, unlike for matrices, a subspace approximation of a tensor given by truncating its higher-order singular values is not optimal, see [106, Example 5]. Moreover, the higher-order singular value decomposition is a high rank decomposition of a tensor in general. It expresses a tensor as a sum of $\prod_{i=1}^{d} r_i$ rank one terms, where r_i is the rank of the ith principal flattening. In general, this will be much higher than the rank of the tensor.

Many algorithms for tensors are based around its flattenings. For example, the algorithm in [112], referred to as a multilinear principal component analysis algorithm, finds projections of a tensor in its different flattenings.

When we view a tensor as an interrelated collection of matrices, several questions of theoretical and practical interest arise. What is the best collection of flattening matrices to use? To what extent does the structure of the tensor appear in its flattenings? How is the structure of the different flattenings interrelated? These questions are algebraic and geometric. For example, in [78] the authors posed the following question.

Problem 1.13 ([78, Problem 1.4])**.** *What are the feasible numbers that can arise as the singular values in the principal flattenings of a tensor?*

The set of feasible singular values of the principal flattenings of a tensor is a semi-algebraic set, defined by polynomial equations and inequalities. This set remains ill-understood. We study it, and describe it in some cases, in Chapter 4.

The singular values of a flattening of a tensor are unchanged under orthogonal changes of basis, like the singular values of a matrix. However, unlike for matrices, the singular values of all the flattenings do not specify a tensor up to orthogonal equivalence. There exist tensors with the same singular values in each flattening, but without an orthogonal change of basis that maps one to the other. This can be seen by a dimension counting argument. A first example of this phenomenon was given in [78]. In Chapter 4, we see how in one setting a full set of orthogonal invariants for a tensor can be obtained by taking the singular values of flattenings in combination with the hyperdeterminant polynomial, which is invariant under changes of coordinates via the special linear group.

There are many ways to represent a tensor using its flattenings. The Tucker decomposition from Equation (1.10) is just one example. One popular approach is to associate a tensor decomposition to a graph. This is called a tensor network. Weights are attached to the edges, and each edge determines an index that is summed over in the decomposition of the tensor. Trees (graphs without cycles) give tensor networks known as hierarchical tensor decompositions [77, Chapter 11], which recursively divide the total set of indices $\{i_1, \ldots, i_d\}$ into subsets, until we reach sets consisting of a single index. The set of tensors that factor according to a given tensor network, with fixed weights on the edges of the graph, give tensor network ranks, studied in [190]. I will return to tensor networks in Chapter 6, where I introduce a broad framework, *tensor hypernetworks*, to study sets of tensors with notions of low rank that can be specialized to those arising in an application. Usual tensor networks, as well as the usual tensor rank, occur as special cases in this framework. See [77] for a numerical treatment of tensor networks.

Rank one approximation

Recall that the tensor $X \in V_1 \otimes \cdots \otimes V_d$ is rank one if $X = v^{(1)} \otimes \cdots \otimes v^{(d)}$, where $v^{(i)} \in V_i$. In coordinates, the tensor X is rank one if

$$x_{i_1,\ldots,i_d} = v^{(1)}_{i_1} \cdots v^{(d)}_{i_d}.$$

A rank one tensor has the interpretation that the different factors, or indices, are de-coupled from one another. A general tensor is a sum of rank one terms, as described in Definition 1.2. The best rank one approximation of a tensor is the closest tensor of the form $v^{(1)} \otimes \cdots \otimes v^{(d)}$, measured with respect to a metric such as the Euclidean distance. In this subsection, I describe the best rank one approximation problem for tensors. As we saw in Theorem 1.9, the best rank one approximation of a matrix is given by a singular vector pair. The best rank one approximation of a tensor is given by a singular vector tuple.

Definition 1.14 (Singular vector tuple)**.** *A singular vector tuple of a tensor $X \in \mathbb{R}^{n_1 \times \cdots \times n_d}$ is a d-tuple of nonzero vectors $(v^{(1)}, \ldots, v^{(d)}) \in \mathbb{C}^{n_1} \times \cdots \times \mathbb{C}^{n_d}$ such that*

$$[\![X; v^{(1)}, \ldots, v^{(k-1)}, \cdot, v^{(k+1)}, \ldots, v^{(d)}]\!] \text{ is parallel to } v^{(k)}, \text{ for all } k = 1, \ldots, d. \quad (1.11)$$

The left side of Equation (1.11) is the vector with ith coordinate

$$\sum_{j_d=1}^{n_d} \cdots \sum_{j_{k+1}=1}^{n_{k+1}} \sum_{j_{k-1}=1}^{n_{k-1}} \cdots \sum_{j_1=1}^{n_1} x_{j_1,\ldots,j_{k-1},i,j_{k+1},\ldots,j_d} v^{(1)}_{j_1} \cdots v^{(k-1)}_{j_{k-1}} v^{(k+1)}_{j_{k+1}} \cdots v^{(d)}_{j_d}.$$

Definition 1.15 (Singular value of a tensor)**.** *Consider a singular vector tuple of norm one vectors $(v^{(1)}, \ldots, v^{(d)})$. The singular value of the tuple is the scalar σ obtained by multiplying the tensor X by the vector $v^{(j)}$ in the jth direction for all j,*

$$\sigma := [\![X; v^{(1)}, \ldots, v^{(d)}]\!].$$

Definitions 1.14 and 1.15 specialize to the usual singular vectors and singular values of a matrix when $d = 2$. Any singular vector tuple can be rescaled to make the vectors norm one. For a tensor, if we do not fix the norm of the vectors $v^{(j)}$ the singular value is not well-defined. Unlike for matrices, the condition in Equation (1.11) is not homogeneous in the entries of the vectors $v^{(j)}$, so rescaling the tuple of vectors will change the singular value. In some settings, the condition in Equation (1.11) is homogenized by taking the $(d-1)$th power of each entry of the vector $v^{(k)}$, in order to have singular values that are unchanged by rescaling the vector, see e.g. [44], but the disadvantage is that such singular values are not orthogonal invariants, see [145]. Note that for a general tensor the singular values are not the same as the higher-order singular values from Definition 1.11.

The notion of a singular vector tuple allows us to extend the Eckart-Young theorem from matrices to tensors, for best rank one approximation.

Theorem 1.16 (Rank one approximation of tensors)**.** *The best rank one approximation of a tensor is given by the singular vector tuple corresponding to the largest singular value.*

This result appears as [67, Lemma 4] and [145, Theorem 2.19], see also further references in [145]. I include a proof here using Lagrange multipliers. For notational simplicity, I work with an order three tensor, though the proof extends directly to tensors of higher order.

Proof of Theorem 1.16. The best rank one approximation of a tensor X of order three is given by $\sigma u \otimes v \otimes w$ where, without loss of generality, we can assume the vectors have norm one, and that σ is non-negative. As for the matrix case in Equation (1.9), we obtain the best rank one approximation by maximizing the inner product $\langle X, u \otimes v \otimes w \rangle$ subject to the conditions

$$\frac{1}{2}(1 - \|u\|^2) = 0, \quad \frac{1}{2}(1 - \|v\|^2) = 0, \quad \text{and} \quad \frac{1}{2}(1 - \|w\|^2) = 0.$$

We use Lagrange multipliers to formulate the optimization problem as the solution to a system of polynomial equations. Unlike the matrix case, these are non-linear equations. The critical points of the functional

$$\sum_{i,j,k} x_{ijk} u_i v_j w_k + \frac{\lambda_1}{2}\left(1 - \sum_i u_i^2\right) + \frac{\lambda_2}{2}\left(1 - \sum_i v_i^2\right) + \frac{\lambda_3}{2}\left(1 - \sum_i w_i^2\right)$$

occur when

$$\sum_{j,k} x_{ijk} v_j w_k = \lambda_1 u_i, \qquad \sum_{i,k} x_{ijk} u_i w_k = \lambda_2 v_j, \qquad \sum_{i,j} x_{ijk} u_i v_j = \lambda_3 w_k.$$

Multiplying each of these expressions by the vector on the right hand side, we see that $\sigma := \lambda_1 = \lambda_2 = \lambda_3$, and we obtain the condition for the triple (u, v, w) to be a singular vector tuple of X with singular value σ. The best rank one approximation is the critical point whose singular value σ is largest. □

The higher rank version of Theorem 1.16 does not hold: we cannot use singular vector tuples to give a low rank approximation of a tensor in general. In this sense, the spectral theory of tensors and the theory of low rank approximations diverge from one another for ranks exceeding one. For more details on singular vectors of tensors as an optimization problem, see [109].

1.4 Differences between matrices and tensors

The first thing that sets tensors apart from matrices is that they facilitate the comparison of multiple variables, or factors, simultaneously. Their properties, both theoretical and numerical, also differ substantially from those of matrices. Decompositions such as the SVD

	Matrix $M \in \mathbb{R}^{n_1 \times n_2}$	Tensor $X \in \mathbb{R}^{n_1 \times \cdots \times n_d}$	See
Real rank	equals complex rank	real and complex ranks may differ; multiple real ranks can occur generically	Example 1.27, Chapter 2
Border rank	the set of matrices of rank $\leq r$ is closed	limits of tensors can have higher rank than the ranks in the sequence	Example 1.24, Chapters 2, 5
Symmetric rank	equals non-symmetric rank by eigen-decomposition	symmetric and non-symmetric ranks may differ	Chapter 5, [168]
Generic rank	same as maximum rank, $\min(n_1, n_2)$	can be less than maximum rank	Example 1.24, Chapter 5
Computing rank	use e.g. SVD, minors	no test in general, can use flattenings for rank 1	Chapters 2, 5
Best low rank approximation	truncate the SVD	use singular vectors for rank one, but not necessarily for higher ranks	Theorems 1.9 and 1.16, Chapter 3

Table 1.1: Summary of differences between matrices and tensors.

do not exist for general tensors. In this section, I describe the various notions of rank, and how to compute the rank, for a tensor. We see significant differences compared to the case of matrices, see Table 1.1.

Definition 1.17 (Complex rank, real rank, symmetric rank, border rank)**.** *Let X be a tensor in $\mathbb{K}^{n_1 \times \cdots \times n_d}$, where $\mathbb{K} = \mathbb{R}$ or $\mathbb{C}$.*

1. Complex rank: *the smallest r such that there exists a decomposition into r rank one terms, $X = \sum_{i=1}^{r} v_i^{(1)} \otimes \cdots \otimes v_i^{(d)}$, where the vectors in the decomposition have complex entries, $v_i^{(j)} \in \mathbb{C}^{n_j}$.*

2. Real rank: *the smallest r such that there exists a decomposition $X = \sum_{i=1}^{r} v_i^{(1)} \otimes \cdots \otimes v_i^{(d)}$, where $v_i^{(j)} \in \mathbb{R}^{n_j}$.*

3. Symmetric rank: *the smallest r such that there exists a decomposition $X = \sum_{i=1}^{r} v_i^{\otimes d}$ where $v_i \in \mathbb{C}^n$ (or $\mathbb{R}^n$ for real symmetric rank).*

4. Border rank: *the smallest r such that $X = \lim_{\epsilon \to 0} X_\epsilon$ where each X_ϵ has rank r. Each notion of rank has a corresponding notion of border rank, the smallest r such that X lies in the closure of the rank r tensors.*

I give some well-known examples of tensors for which the different ranks differ, in Section 1.5. The examples involve the smallest higher order tensors, those of format $2 \times 2 \times 2$. Real rank applies to tensors with real entries, similarly symmetric rank applies to tensors with symmetric entries. The set of tensors of rank $\leq r$ may not be closed when it is a proper subset of the space of tensors and $r > 1$. The same holds for the space of symmetric tensors of symmetric rank $\leq r$, see [50]. In these situations, there exist tensors whose rank and border rank differ. The notions of rank in Definition 1.17 can specialized to matrices, where they all give the usual rank of a matrix.

Another key difference between matrices and tensors is the uniqueness properties of a decomposition. A matrix always has infinitely many decompositions into rank one terms, but in some cases a tensor has a unique such decomposition, and this can be very useful in practice, see [93, Section 3.2] and [98]. I discuss the identifiability of tensors under congruence action in Chapter 8.

I now come to a comparison of the rank and symmetric rank of a tensor. A conjecture from [50] says the following.

Conjecture 1.18 (Comon's conjecture)**.** *The rank and symmetric rank of a symmetric tensor agree.*

The conjecture is true in some special cases: for matrices (by the eigen-decomposition), when the symmetric rank of a tensor is at most two [50], when the rank is less than or equal to the order [193], and when the rank is at most the flattening rank plus one [68]. Furthermore, the conjecture has been proved to generically hold in certain families of tensors [17]. However, in [168], it was shown that the conjecture is false: the author constructs a tensor whose symmetric and non-symmetric ranks differ. It is a large tensor of format $800 \times 800 \times 800$. I refer the reader to [168] to read more about the example.

Many aspects of Conjecture 1.18 remain unknown. It is not known whether rank and symmetric rank are the same generically, i.e. whether for practical purposes in the study of tensor data they can be considered the same. Moreover, there are no known counter-examples over the real numbers, and an example in which the border rank and symmetric border rank differ has not yet been obtained. I will discuss rank and symmetric rank further in Chapter 5, where I prove that the rank and symmetric rank coincide for all tensors of symmetric rank at most seven, improving on previous best lower bounds from [68]. This gives a partial explanation for why counter-examples to Conjecture 1.18 seem hard to find: there are very few tools for computing the rank of a tensor whose rank is high (i.e. exceeding seven), as I now describe.

Recall that the rank of a matrix can be read directly from the singular value decomposition, see Section 1.2. The rank of a matrix M can also be computed algebraically, as we saw in the discussion of matrix structure via polynomials on Page 13. Computing the rank of a tensor is much more difficult. We can test if a tensor has rank one using the flattenings. If all the flattenings of a tensor have rank one, then the tensor has rank one. If all the flattenings of a tensor have rank two, then the tensor has border rank two [101]. This does not extend

to higher ranks: the ranks of the flattenings will only give a lower bound on the rank of the tensor.

Many algorithms have been developed to compute tensor rank and tensor decompositions. Roughly speaking, the algorithms fall into two categories. Algebraic methods give exact solutions, but are not tractable for large tensors. Numerical methods are more scalable, but with fewer guarantees of finding the optimal solution. Numerical algebraic geometry algorithms, see e.g. [83], interpolate between the two in terms of scalability and guarantees.

For a general tensor, there is no known way to compute its rank. The problem of computing the rank is algebraic, as I now describe. I begin by considering the symmetric rank of a tensor, and how it relates to a question of classical interest in algebraic geometry. Then I discuss the complex rank, and its connection to secant varieties, and finally I discuss real rank and its connection to semi-algebraic sets.

Symmetric rank

Recall that a tensor of format $n \times \cdots \times n$ (d times) corresponds to a homogeneous polynomial of degree d in n variables, see Example 1.7. Under this correspondence, a symmetric rank one tensor $v^{\otimes d}$ corresponds to a power of a linear form l^d. The vector $v \in \mathbb{K}^n$ lists the coefficients of the linear form l. Hence a decomposition of X into a sum of rank one symmetric terms is a decomposition of the corresponding polynomial into a sum of powers of linear forms,

$$X = \sum_{i=1}^{r} \lambda_i v_i^{\otimes d} \qquad \leftrightarrow \qquad f = \sum_{i=1}^{r} \lambda_i l_i^d, \tag{1.12}$$

where each $v_i \in \mathbb{K}^n$ is the vector of coefficients of the linear form l_i and the λ_i are scalars. The minimal number of summands in such a decomposition is called the *symmetric rank* of X, or the *Waring rank* of f, over $\mathbb{K}$. We have the following classical result about cubic surfaces.

Theorem 1.19 (Sylvester's Pentahedral Theorem (1851), see [159, §84])**.** *A generic cubic surface f can be decomposed uniquely as the sum of five cubes $f = l_1^3 + l_2^3 + l_3^3 + l_4^3 + l_5^3$, where each $l_i \in \mathbb{C}[x_1, x_2, x_3, x_4]$ is a linear form.*

The theorem says that a generic $4 \times 4 \times 4$ symmetric tensor has complex rank five. Or, the complex border rank of a symmetric $4 \times 4 \times 4$ tensor is at most five. We will see more about ranks of cubic surfaces in Chapter 5.

Perhaps the first example of an algorithm to compute the rank of a symmetric tensor is Sylvester's algorithm from 1886 for binary forms, see [37] and references therein. Approaches for larger tensors involve algebraic tools such as the apolarity lemma, which relates the existence of a decomposition to the structure of the apolar ideal. In principle, the rank of a tensor can be computed directly from solving an elimination problem: by eliminating the variables v_i from the decomposition on the left in Equation (1.12), where the entries of X are set to the tensor of interest. However, such eliminations do not terminate in practice.

Another close connection between tensor rank and algebraic tools begins with the following definition.

Definition 1.20 (Varieties of symmetric rank r tensors)**.** *The set of symmetric tensors of format $n \times \cdots \times n$ (d times) of complex rank one, considered up to scale, is the Veronese variety $\nu_d(\mathbb{P}^{n-1})$. The set of tensors whose complex symmetric border rank is at most r, considered up to scale, is the rth secant variety of the Veronese, denoted by $\sigma_r(\nu_d(\mathbb{P}^{n-1}))$.*

The following theorem justifies calling this set an algebraic variety. It is the specialization of a classical result from algebraic geometry to the setting of tensors, and it holds in the case of both symmetric and non-symmetric tensors.

Theorem 1.21. *The set of tensors of complex (symmetric) border rank at most r is an algebraic variety.*

Proof. The Zariski closure of a set S is the smallest set containing S that is defined by the vanishing of polynomials. A locally closed set is one that is Zariski open in its Zariski closure, and a constructible set is the finite union of locally closed sets.

The set of tensors of complex (symmetric) rank at most r is a constructible set, by Chevalley's Theorem [181, p. 7.4.2], since it is the image of the vectors occurring in a rank r (symmetric) decomposition. Hence the set of tensors of complex rank r is a finite union of sets that are open in their Zariski closures. Taking the Euclidean closure of each set gives the Euclidean closure of the complex rank r tensors. Since each set is locally closed, the Euclidean closure equals the Zariski closure. In fact, since secants varieties of irreducible varieties are irreducible [191], it is enough to take the closure of the largest locally closed set. □

Computing the border rank of a symmetric tensor means testing the vanishing of the defining equations of a certain secant variety, i.e. we need to compute finitely many polynomials in the entries of the tensor. However, these equations are known in very few cases. Since the rank of a tensor is invariant under general linear changes of coordinates in each factor, the defining equations will be modules of the general linear group, which we make use of in Chapter 5. See [105, Chapter 6].

Although the symmetric rank of a given tensor cannot be found in general, the rank of a generic tensor can be computed. For tensors of fixed format there is one complex rank that occurs generically, and it is called the *generic rank* [105]. The value of the generic rank follows from the dimensions of the secant varieties. The secant varieties of the Veronese variety have an expected dimension, given by counting parameters. The Alexander-Hirschowitz Theorem [4], see also [105, Theorem 3.2.2.4], says that the true dimension of the secant variety is given by the expected dimension, in all but a few exceptional cases.

Complex rank

Consider the set of tensors of complex border rank at most r, i.e. the set of tensors that can be written as a limit $X = \lim_{\epsilon \to 0} X_\epsilon$, where each tensor X_ϵ has complex rank r. As in the symmetric case, we have the following definition.

Definition 1.22 (Varieties of rank r tensors)**.** *The set of tensors of format $n_1 \times \cdots \times n_d$ of rank one, considered up to scale, is the Segre variety* $\mathrm{Seg}(\mathbb{P}^{n_1-1} \times \cdots \times \mathbb{P}^{n_d-1})$*. The set of tensors of complex border rank at most r, considered up to scale, is the rth secant variety of the Segre, denoted* $\sigma_r(\mathrm{Seg}(\mathbb{P}^{n_1-1} \times \cdots \times \mathbb{P}^{n_d-1}))$*.*

As in the symmetric case, in principle we can test if a tensor has complex border rank at most r by testing the vanishing of the defining equations for the secant variety. The generic rank is the smallest r such that the rth secant variety fills the space of tensors. It is conjectured that the secant varieties of the Segre variety have the expected dimension, except in a few exceptional cases [1]. However, unlike for the symmetric case, to date a full characterization remains unproved.

Real rank

A *basic semi-algebraic set* is a set of points upon which a finite list of polynomials vanishes and a finite list of polynomials is strictly positive. A *semi-algebraic set* is the union of finitely many basic semi-algebraic sets. Real algebraic geometry is based around the study of semi-algebraic sets, and the following theorem is central.

Theorem 1.23 (The Tarski-Seidenberg theorem/quantifier elimination [32, Chapter 5])**.** *The projection of a semi-algebraic set is semi-algebraic.*

A decomposition of a tensor X as a sum of r real rank one terms has the form $X = \sum_{i=1}^{r} v_i^{(1)} \otimes \cdots \otimes v_i^{(d)}$, where $v_i^{(j)} \in \mathbb{R}^{n_j}$. Consider the set of all tensors which can be written in the above form, as the vectors $v_i^{(j)}$ range over $\mathbb{R}^{n_j}$. This is a semi-algebraic set in the entries of the tensor X and the entries of all the vectors $v_i^{(j)}$. Projecting this set to the entries of X gives the set of tensors X for which *there exists* a real rank $\leq r$ decomposition. The Tarski-Seidenberg theorem says that this projected set is semi-algebraic, i.e. it is a finite union of sets each of which is given by the signs of certain polynomials in the entries of the tensor. In principle, this gives a membership test for the tensors of real rank $\leq r$. We can test if a tensor has real rank r by evaluating the list of polynomials and seeing if they have one of the required sign patterns. We can repeat the test for different ranks r to find the rank of a tensor.

Although Theorem 1.23 gives the existence of a semi-algebraic membership test for the set of tensors of fixed real rank, it is a different question to actually obtain the equations and inequalities. The problem of finding a membership test for the real rank $\leq r$ tensors quickly becomes computationally intractable as the size of the tensor increases, as quantifier elimination is doubly-exponential in the number of variables [55].

I give the equations for the real rank membership problem for real rank two tensors in Chapter 2. The set of tensors of fixed non-negative rank is also a semi-algebraic set, see the discussion of non-negative rank on Page 30. See Chapter 7 for more about membership tests for non-negative rank r tensors.

Unlike in the complex case, the real rank of a tensor does not take a single value generically. When sampling uniformly in the space of real tensors of fixed format, multiple real ranks occur with positive probability. These are called the *typical ranks*, see [28]. The typical ranks of a given tensor format are not known in general.

1.5 Small data

The smallest higher order tensors have format $2 \times 2 \times 2$ and consist of eight entries x_{ijk}, for $0 \leq i, j, k \leq 1$. The entries can be arranged at the vertices of a three-dimensional cube, see Figure 1.8. In this section, I describe some known results, and also some new contributions, in the setting of $2 \times 2 \times 2$ tensors. The results are a preview of the contributions of later chapters. I begin by discussing the flattenings, then I give exact algebraic algorithms to compute the complex, real, and non-negative rank. Algorithms to compute the real and complex ranks in this setting are well-known, but there was previously no exact method to compute the non-negative rank.

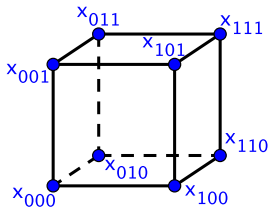


Figure 1.8: Tensors of format $2 \times 2 \times 2$ consist of eight entries arranged at the vertices of a three-dimensional cube.

Flattenings

A tensor X of format $2 \times 2 \times 2$ has three flattenings, each of format 2×4:

$$\begin{bmatrix} x_{000} & x_{001} & x_{010} & x_{011} \\ x_{100} & x_{101} & x_{110} & x_{111} \end{bmatrix}, \quad \begin{bmatrix} x_{000} & x_{001} & x_{100} & x_{101} \\ x_{010} & x_{011} & x_{110} & x_{111} \end{bmatrix}, \quad \begin{bmatrix} x_{000} & x_{010} & x_{100} & x_{110} \\ x_{001} & x_{011} & x_{101} & x_{111} \end{bmatrix}. \tag{1.13}$$

In the ith flattening, the ith index labels the rows of the matrix. For each flattening M, we can compute the 2×2 matrix MM^{T} and then take its determinant, $\det(MM^{\mathsf{T}})$. I call this the ith Gram determinant of X, and denote it by g_i.

By the Cauchy-Binet formula, each of the three Gram determinants is the sum of squares of the six minors from a flattening. For example,

$$g_1 = \begin{array}{c}(x_{000}x_{101} - x_{001}x_{100})^2 + (x_{000}x_{110} - x_{010}x_{100})^2 + (x_{000}x_{111} - x_{011}x_{100})^2 + \\ (x_{001}x_{110} - x_{010}x_{101})^2 + (x_{001}x_{111} - x_{011}x_{101})^2 + (x_{010}x_{111} - x_{011}x_{110})^2.\end{array} \tag{1.14}$$

Each minor is supported on four vertices in the cube. The faces of the cube in Figure 1.8 are the support of minors which appear in two Gram determinants. They have monomials $x_{\mathbf{i}}x_{\mathbf{j}}$ where $\mathbf{i}$ and $\mathbf{j}$ are multi-indices in $\{0,1\}^3$ that differ in two indices. The remaining six minors, supported on the shaded squares in Figure 1.9, are unique to a single flattening. They have monomials $x_{\mathbf{i}}x_{\mathbf{j}}$ where $\mathbf{i}$ and $\mathbf{j}$ differ in all three coordinates.

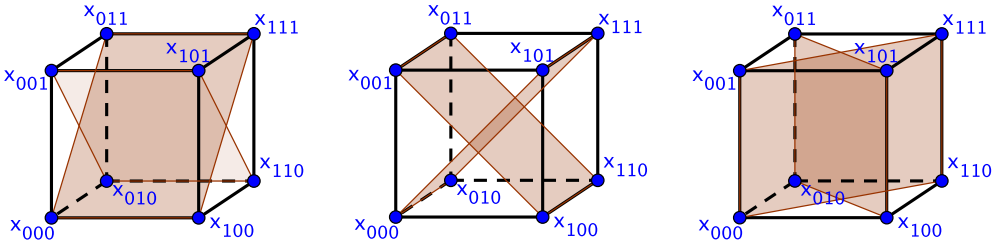


Figure 1.9: The minors unique to flattenings one, two, and three, respectively.

The following inequalities between the Gram determinants hold:

$$\begin{aligned} g_1 &\leq g_2 + g_3, \\ g_2 &\leq g_1 + g_3, \\ g_3 &\leq g_1 + g_2. \end{aligned}$$

I prove these inequalities by finding a sum of squares certificate for $g_2 + g_3 - g_1$ in Example 4.3. The inequalities satisfied by the Gram determinants result in relations between the higher-order singular values. I find relations between the Gram determinants of larger tensors, and use them to make conclusions about the higher-order singular values, in Chapter 4.

Complex rank

The complex rank of a $2 \times 2 \times 2$ tensor X is the smallest number r such that it can be written

$$X = \sum_{l=1}^{r} a_l \otimes b_l \otimes c_l, \quad \text{for some} \quad a_l, b_l, c_l \in \mathbb{C}^2.$$

The possible complex ranks of a $2 \times 2 \times 2$ tensor are $\{1, 2, 3\}$.

Consider a $2 \times 2 \times 2$ tensor up to scale as a point in projective space $\mathbb{P}^7$. The rank one tensors parametrize the Segre variety, $\text{Seg}(\mathbb{P}^1 \times \mathbb{P}^1 \times \mathbb{P}^1)$. The closure of the complex rank

two tensors is the secant variety to the Segre, denoted $\sigma_2(\text{Seg}(\mathbb{P}^1 \times \mathbb{P}^1 \times \mathbb{P}^1))$. The secant variety fills the space $\mathbb{P}^7$. This means that a generic $2 \times 2 \times 2$ tensor has complex rank two, and that any $2 \times 2 \times 2$ tensor can be arbitrarily well-approximated by one of complex rank two, i.e. the complex border rank of a $2 \times 2 \times 2$ tensor is at most two.

The complex rank three tensors parametrize a subvariety of $\mathbb{P}^7$, given by the vanishing of the *hyperdeterminant* [42]

$$\begin{aligned} x_{000}^2x_{111}^2 + x_{001}^2x_{110}^2 + x_{010}^2x_{101}^2 + x_{011}^2x_{100}^2 + 4x_{000}x_{011}x_{101}x_{110} + 4x_{001}x_{010}x_{100}x_{111} \\ -2x_{000}x_{001}x_{110}x_{111} - 2x_{000}x_{010}x_{101}x_{111} - 2x_{000}x_{011}x_{100}x_{111} \\ -2x_{001}x_{010}x_{101}x_{110} - 2x_{001}x_{011}x_{100}x_{110} - 2x_{010}x_{011}x_{100}x_{101}. \end{aligned} \tag{1.15}$$

The hyperdeterminant is the only polynomial (up to scale) in the entries of a $2 \times 2 \times 2$ tensor that is invariant under change of basis by the group $SL_2 \times SL_2 \times SL_2$. The hyperdeterminant arises here because the rank of a tensor is also an invariant with respect to this group action.

Example 1.24 (Border rank vs. rank)**.** *The tensor*

$$X = \left[\begin{array}{cc||cc} 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 \end{array}\right] \tag{1.16}$$

can be arbitrarily well-approximated by a sum of two rank one tensors,

$$X = \lim_{\epsilon \to 0} \frac{1}{2\epsilon} \left(\left[\begin{array}{cc||cc} 1 & \epsilon & \epsilon & \epsilon^2 \\ \epsilon & \epsilon^2 & \epsilon^2 & \epsilon^3 \end{array}\right] + \left[\begin{array}{cc||cc} -1 & \epsilon & \epsilon & -\epsilon^2 \\ \epsilon & -\epsilon^2 & -\epsilon^2 & \epsilon^3 \end{array}\right] \right).$$

However, to express X exactly as a sum of rank one tensors, three terms are required. The tensor X has rank three and border rank two, and its hyperdeterminant is zero.

Example 1.25 (Generic rank vs. maximum rank)**.** *The rank of a generic $2 \times 2 \times 2$ tensor is two, hence the tensor in Equation* (1.16) *is an example of a tensor whose rank exceeds the generic rank.*

We will also see the role of the hyperdeterminant in computing the real rank of a tensor. A cartoon of the different possibilities for the real rank and complex rank of a $2 \times 2 \times 2$ tensor is given in Figure 1.10.

In summary, we have the following algebraic recipe to find the complex rank of a $2 \times 2 \times 2$ tensor.

Algorithm 1.26 (Find the complex rank of a $2 \times 2 \times 2$ tensor)**.** *First, test if all 2×2 minors of flattenings vanish, see Equation* (1.13)*. If all minors all vanish, the tensor has rank one. Otherwise, test if the hyperdeterminant vanishes, see Equation* (1.15)*. If it vanishes, the tensor has complex rank three. Otherwise the complex rank is two.*

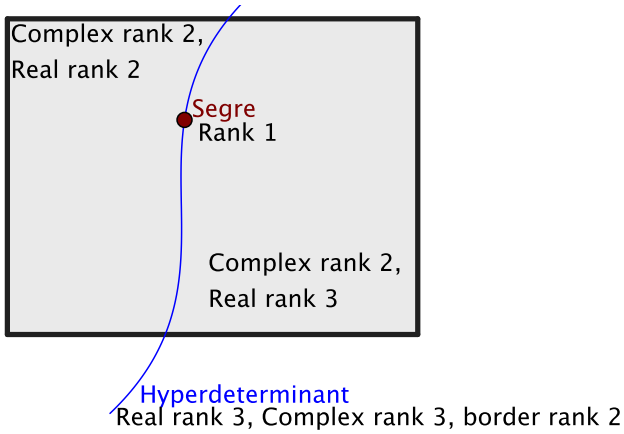


Figure 1.10: The hyperdeterminant divides the $2 \times 2 \times 2$ tensors according to their real rank.

Real rank

The real rank of a $2 \times 2 \times 2$ tensor is the smallest number r such that it can be written

$$X = \sum_{l=1}^{r} a_l \otimes b_l \otimes c_l, \quad \text{for some} \quad a_l, b_l, c_l \in \mathbb{R}^2.$$

A $2 \times 2 \times 2$ tensor X has real rank one if $X = a \otimes b \otimes c$, for some real vectors a, b, c of length two. If X is real, and rank one, then the vectors a, b, c in such an expression can always be chosen to be real. Hence a tensor is real rank one if and only if it is a real point in the cone over the Segre variety $\text{Seg}(\mathbb{P}^1 \times \mathbb{P}^1 \times \mathbb{P}^1)$. For higher ranks, a tensor with real entries can have real rank strictly larger than its complex rank, as the following example shows.

Example 1.27 (Complex rank vs. real rank)**.** *Let X be the complex rank two tensor $z \otimes z \otimes z + \overline{z} \otimes \overline{z} \otimes \overline{z}$ where $z = \begin{bmatrix} 1 & i \end{bmatrix}^{\mathsf{T}}$ and $\overline{z} = \begin{bmatrix} 1 & -i \end{bmatrix}^{\mathsf{T}}$ is its complex conjugate. The tensor X has real entries*

$$X = \left[\begin{array}{cc||cc} 2 & 0 & 0 & -2 \\ 0 & -2 & -2 & 0 \end{array}\right]. \tag{1.17}$$

However, its real rank exceeds two. We can show that there no way to write X as a sum of two real rank one terms, using Macaulay2 [76]. The following code constructs the ideal I consisting of all equations obtained by setting the sum of two rank one tensors to be equal to the tensor X in Equation (1.17)*.*

```
R = QQ[a_(0,0)..a_(1,1),b_(0,0)..b_(1,1),c_(0,0)..c_(1,1)];
for i to 1 do ( for j to 1 do (  for k to 1 do (
    x_(i,j,k) = a_(i,0)*b_(j,0)*c_(k,0) + a_(i,1)*b_(j,1)*c_(k,1); )));
X_(0,0,0) = 2; X_(0,0,1) = 0; X_(0,1,0) = 0; X_(1,0,0) = 0;
X_(0,1,1) = -2; X_(1,0,1) = -2; X_(1,1,0) = -2; X_(1,1,1) = 0;
```

```
I = ideal();
for i to 1 do ( for j to 1 do ( for k to 1 do (
    I = I + ideal(X_(i,j,k) - x_(i,j,k)) )));
decompose I
```

The output of this computation shows that the ideal I contains terms

$$a_{01}^2 + a_{11}^2 \qquad a_{00}^2 + a_{10}^2.$$

Over $\mathbb{R}$, these are zero only if $a_{00} = a_{01} = a_{10} = a_{11} = 0$. But if all entries a_{ij} are zero, the overall decomposition would be a tensor of zeros, which is not equal to X. Hence X has no real rank two decomposition.

The matrix analogue of the above example is the rank two matrix $z \otimes z + \overline{z} \otimes \overline{z}$, which can be decomposed into real rank one terms using elementary basis vectors. Why is rank equal to complex rank for a matrix? One way to see this is via the singular value decomposition: each left and right singular vector pair is the best rank one approximation to a real matrix, and hence is real.

The set of tensors of fixed real rank is a semi-algebraic set. A real tensor $X \in \mathbb{R}^{2\times 2\times 2}$ lies in the (topological) closure of the real rank two tensors if and only if the hyperdeterminant from Equation (1.15) is non-negative. Multiple real ranks occur with positive probability when sampling uniformly in the space of real $2 \times 2 \times 2$ tensors. I generalize the description of real ranks from Figure 1.10 to general tensor formats in Chapter 2.

We have the following algebraic recipe for finding the real rank of a $2 \times 2 \times 2$ tensor.

Algorithm 1.28 (Find the real rank of a $2\times2\times2$ tensor)**.** *Test if all three Gram determinants vanish, see Equation* (1.14)*. If so, the tensor has real rank one. Otherwise, compute the hyperdeterminant. If it is strictly positive, the tensor has real rank two. If it is not positive, the tensor has real rank three.*

Non-negative rank

In a statistical setting, we seek tensor decompositions whose rank one terms can be given a probabilistic interpretation, see Chapter 7. This means the entries of the rank one tensors must be non-negative. We have the following definition.

Definition 1.29 (Non-negative rank)**.** *The non-negative rank of a non-negative tensor X is the minimal r such that X can be written as a sum of r rank one terms with non-negative entries,*

$$X = \sum_{i=1}^{r} v_i^{(1)} \otimes \cdots \otimes v_i^{(d)}, \qquad \text{for some} \quad v_i^{(j)} \in \mathbb{R}^{n_j}_{\geq 0}.$$

Specializing the above definition to the case $d = 2$ gives the non-negative rank of a matrix, see [45]. The rank and non-negative rank of a matrix agree for 2×2 matrices, but the ranks are not the same for larger matrices.

The set of tensors of fixed non-negative rank is a semi-algebraic set by Theorem 1.23, since it is a projection of a semi-algebraic set. The possible non-negative ranks of a $2 \times 2 \times 2$ tensor are $\{1, 2, 3, 4\}$. A membership test for the non-negative ranks was only previously known for non-negative ranks one and two [6]. In Chapter 7, I give a membership test for the non-negative rank three tensors of this format. The non-negative rank of a $2 \times 2 \times 2$ tensor with positive entries can therefore be determined algebraically, via the following algorithm.

Algorithm 1.30 (Find the non-negative rank of a strictly positive $2\times2\times2$ tensor)**.** *Compute the signs* $\{+, -, 0\}$ *of the six binomials*

$$\begin{array}{lll} x_{000}x_{011} - x_{010}x_{001}, & x_{000}x_{101} - x_{100}x_{001}, & x_{000}x_{110} - x_{100}x_{010}, \\ x_{100}x_{111} - x_{110}x_{101}, & x_{010}x_{111} - x_{110}x_{011}, & x_{001}x_{111} - x_{101}x_{011}. \end{array} \tag{1.18}$$

If all six binomials vanish, the tensor has non-negative rank one. Otherwise, if the sign pattern of the 2×3 *grid of determinants is any of*

$$\begin{array}{llll} + \ + \ + & + \ - \ - & - \ + \ - & - \ - \ + \\ + \ + \ +, & + \ - \ -, & - \ + \ -, & - \ - \ +, \end{array}$$

the tensor has non-negative rank two. Otherwise, if the sign pattern is any of

$$\begin{array}{llllll} + \ * \ * & - \ * \ * & * \ + \ * & * \ - \ * & * \ * \ + & * \ * \ - \\ + \ * \ *, & - \ * \ *, & * \ + \ *, & * \ - \ *, & * \ * \ +, & * \ * \ -, \end{array}$$

the tensor has non-negative rank three, where $*$ *indicates that any sign is allowed. Otherwise, the tensor has non-negative rank four.*

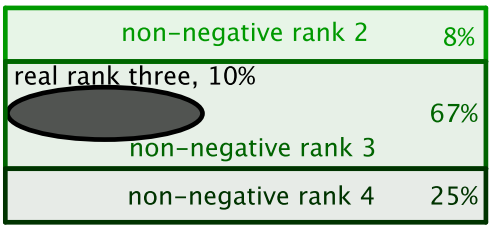


Figure 1.11: The possible non-negative ranks and real ranks of a $2 \times 2 \times 2$ tensor, with percentages estimated by sampling uniformly in the space of non-negative tensors with entries summing to one.

The algorithm be extended to test the non-negative rank of a tensor with zero entries, by taking the closure of the semi-algebraic sets given by the sign patterns above. The set of

non-negative tensors with entries summing to one is the probability simplex Δ_7. Previous results showed that the non-negative rank ≤ 3 tensors occupy a volume of at most 96.4% of this set [125]. Using Algorithm 1.30, we can estimate the true volume to be 75.3%, see Figure 1.11 and Chapter 7. The following example gives a $2 \times 2 \times 2$ tensor with non-negative entries, whose real and non-negative ranks differ.

Example 1.31. *The following non-negative tensor has real rank 2 and non-negative rank 4.*

$$\left[\begin{array}{cc||cc} 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{array}\right] = \frac{1}{2}\left[\begin{array}{cc||cc} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{array}\right] + \frac{1}{2}\left[\begin{array}{cc||cc} 1 & -1 & -1 & 1 \\ -1 & 1 & 1 & -1 \end{array}\right]$$

The hyperdeterminant of this tensor, from Equation (1.15)*, is four. The sign pattern of the binomials from Equation* (1.18) *is* $\begin{smallmatrix} + & + & + \\ - & - & - \end{smallmatrix}$.

1.6 Statement of contributions

The geometry of structured tensors

In *Chapter 2: Real rank geometry*, I present an implicit description, or membership test, for the set of real rank two tensors. Previously, such a description was only known in the rank one case where it is a classical object, the real points on the Segre variety. I consider alternative notions of rank, which use building blocks other than rank one tensors, and I use this geometric framework to study the boundary of the real rank two tensors. This chapter is based on joint work with Bernd Sturmfels, published in the Journal of Algebra [164].

In *Chapter 3: Singular vectors*, I describe the singular vectors of orthogonally decomposable tensors as a variety in a product of projective spaces. This chapter is based on joint work with Elina Robeva, published in Linear and Multilinear Algebra [153].

In *Chapter 4: Singular values*, I study the higher-order singular values of a tensor via polynomial orthogonal invariants, the determinants of flattenings. By finding relations between the determinants, I answer a question raised in [78] concerning the tensors whose higher-order singular values take extremal values. This chapter is based on a single-authored project, published in Linear Algebra and its Applications [160].

In *Chapter 5: Rank vs. symmetric rank*, I introduce a test for high rank tensors via discriminant loci, and use it to study questions from classical algebraic geometry. I prove that cubic surfaces with finitely many singular points are a sum of at most six cubic powers of linear forms, generalizing a classical result from [159]. It is known that rank and symmetric rank need not agree for symmetric tensors [168], but in many practical situations equivalence of rank and symmetric rank seems to hold. I use the discriminant characterization of high rank tensors to prove that rank and symmetric rank coincide for all tensors of symmetric rank at most seven, improving on previous best lower bounds from [68]. This chapter is based on a single-authored project, available as a preprint [161].

In *Chapter 6: Tensor hypernetworks*, I study the generalization of tensor networks from graphs to hypergraphs. The generalization enables usual tensor rank to be realized in the context of tensor network ranks, and connections to be made to statistics. I show that tensor hypernetworks are dual to graphical models, multivariate statistical models on based on graphs. One consequence is the equivalence of two algorithms: the belief propagation algorithm for marginalizing probability distributions, and an algorithm for tensor network contraction from quantum physics. This chapter is based on joint work with Elina Robeva, published in Information and Inference: A Journal of the IMA [152].

Algorithms for tensor data

In *Chapter 7: Semi-algebraic statistics*, I obtain the semi-algebraic description of two statistical models, a mixture model and a restricted Boltzmann machine. I use this description to give a closed-form formula for the maximum likelihood estimate to the model. For models with hidden variables, maximum likelihood estimates are usually found using local optimization methods with no guarantee of global convergence. The algebraic approach gives the first non-trivial instance of an exact maximum likelihood solution for a model with hidden variables. This chapter is based on joint work with Guido Montúfar, published in the Journal of Algebraic Statistics.

In *Chapter 8, Learning paths from signature tensors*, I study the extension of matrix congruence to the setting of tensors. Given a tensor in the orbit of another tensor, I seek the matrix that transforms one to the other. I give identifiability results, both exact and numerical, for this recovery problem. The motivation is an inverse problem from stochastic analysis: the recovery of paths from their third order signature tensors. I give a numerical optimization algorithm for path recovery from inexact data and an algorithm to approximate the shortest path with a given signature tensor. This chapter is based on joint work with Max Pfeffer and Bernd Sturmfels, published in the SIAM Journal on Matrix Analysis and Applications [141].

In *Chapter 9: Tensor clustering with algebraic constraints*, I study the problem of clustering, i.e. partitioning data into subsets related in ways that are not directly measured. I give a structured clustering algorithm for multi-dimensional data. The algorithm encodes algebraic constraints in a tensor, and then partitions the data using integer optimization, finding the globally optimal partition of the data with respect to an objective function. I apply the algorithm to cluster a dataset of breast cancer cell lines, to find similarities consistent with a mechanism for signal transduction in two pathways that are known to dysfunction in cancer. This chapter is based on joint work with Mariano Beguerisse-Díaz, Birgit Schoeberl, Mario Niepel and Heather Harrington, published in the Journal of the Royal Society Interface [166].

Part I

The geometry of structured tensors

Chapter 2

Real rank geometry

The governing idea behind matrix methods, such as principal component analysis and the singular value decomposition, is that much of the useful information in a matrix can be captured by a relatively small number of rank one components. In a similar way, low rank approximation of tensors is central to extracting structure from multi-dimensional data. The approximation compresses the information in the original tensor. In each of the rank one components, the variables have been de-coupled from one another, and this facilitates interpretation in the context of an application: the data is a mixture of rank one signals. Despite the importance of low rank approximations for compression and interpretation of tensor data, many theoretical results about tensor rank and low rank approximation of tensors remain unknown [93, 170].

The best rank r approximation of a matrix is given by truncating the singular value decomposition to the largest r singular values, as we saw in Section 1.2. In particular, a low rank decomposition of a matrix can be found by computing successive best rank one approximations. Moreover there are algebraic tests for a matrix to have rank at most r, given by evaluating polynomials in the entries of the matrix. For tensors, no exact test to compute the rank of a tensor is known in general, as we saw in Section 1.4. On the numerical side, while numerical linear algebra is well-established [58], and numerical multi-linear algebra is a fast-growing area [77], numerical algorithms to compute low rank approximations of tensors have drawbacks and challenges [85, 93, 170]. Computing a low rank approximation of a tensor can lead to counter-intuitive pitfalls. For example, subtracting a best rank one approximation from a tensor may increase the rank [174].

The real and complex rank of a real tensor may differ, as we saw in Example 1.27. The distinction between real and complex rank is important because, for a real tensor, a decomposition involving real numbers may be required in order to interpret the rank one terms in the context of an application. For example, the rank of the tensor encoding a linear operator is the number of multiplications it requires [105, 177], and the real rank gives the number of multiplication over the real numbers. In functional magnetic resonance imaging (fMRI) the tensor consists of spacial and temporal axes. The rank is used to count the signals [3], and the entries of the tensor are relative levels of blood oxygenation, which are

real-valued. In signal processing, the measurement is the amplitude of a real (e.g. sound) signal [169].

Challenges associated with low rank approximation of real tensors become clearer after understanding the geometry of sets of low rank tensors. The space of all tensors of a given format, say $n_1 \times \cdots \times n_d$, is a high-dimensional vector space $\mathbb{K}^{n_1 \times \cdots \times n_d}$. The tensors that possess a certain property form a subset of this space. Often we consider tensors up to scale as points in projective space $\mathbb{P}^{N-1}$ where $N = n_1 \cdots n_d$. For example, the rank one tensors up to scale parametrize the Segre variety $\text{Seg}(\mathbb{P}^{n_1-1} \times \cdots \times \mathbb{P}^{n_d-1})$, the zero set of the 2×2 minors of the flattenings. The rank one tensors themselves parametrize the affine cone over the Segre variety. The rank one tensors are an algebraic variety: they are defined by the vanishing of certain polynomials.

Other sets of structured tensors are the zero-set of some polynomials in the entries of the tensor. This underlies methods to study tensors and tensor rank via algebraic geometry, see [105]. Tensors of complex rank r are the rth secant variety of the Segre variety, up to closure, see Definitions 1.20 and 1.22. Methods to obtain defining equations of these secant varieties are discussed in [105]. Less is known about the tensors of fixed real rank. The set of tensors of real rank at most r is a semi-algebraic set in $\mathbb{R}^{n_1 \times \cdots \times n_d}$, defined by polynomial equations and inequalities, see Theorem 1.23. The real and complex situations are similar for rank one: the real rank one tensors are the real points on the Segre variety. For higher ranks, it is not true that the real rank of a tensor can be determined by computing its complex rank and checking whether the entries are real. The semi-algebraic set of real rank r tensors is full-dimensional, but not full volume, in the set of real tensors of complex rank r. The equations and inequalities for the set of tensors of real rank at most r give a membership test for a tensor to lie in the set.

In this chapter, I give a membership test for the set of real rank two tensors, based on evaluating polynomial equations and inequalities. I give a description of the boundary of the set, which I use to shed light on numerical issues arising in low rank tensor approximation. I also give a method to lower bound the real rank of a tensor. Some situations require building blocks other than rank one terms. I describe the geometric framework of ranks with respect to other building blocks, and focus on real rank with respect to a curve in three-dimensional space. This chapter is based on joint work with Bernd Sturmfels, published in the Journal of Algebra [164]. Section 2.2 is from my preprint [161].

2.1 Real rank two tensors

In this section, I give a semi-algebraic description of the real rank two tensors. It is given by the signs of some polynomials in the entries of the tensors.

Definition 2.1 (The real rank r locus)**.** *The* real rank r locus *is the topological closure of the set of real rank r tensors, i.e. the set of tensors of real border rank at most r. It consists of all tensors that can be written as a limit $\lim_{\epsilon \to 0} X_\epsilon \to X$ where each X_ϵ has real rank r.*

A membership test for a $2 \times 2 \times 2$ tensor to lie in the real rank two locus is given on Page 29. The main result of this section is a membership test for the real rank two locus for tensors of arbitrary format. The statement of the theorem requires a generalization of the hyperdeterminant from Equation (1.15) to a tensor X of general format $n_1 \times \cdots \times n_d$. A $2 \times 2 \times 2$ sub-tensor of X is obtained by choosing pairs for three indices and fixing the remaining $d-3$ indices. I call the hyperdeterminants of these sub-tensors the $2 \times 2 \times 2$ *sub-hyperdeterminants* of X. The number of sub-hyperdeterminants is

$$\frac{1}{8} \cdot n_1 n_2 n_3 \cdots n_d \cdot \sum_{1 \leq i < j < k \leq d} (n_i - 1)(n_j - 1)(n_k - 1). \tag{2.1}$$

Note that the sub-hyperdeterminants of X differ from its hyperdeterminant [135].

Theorem 2.2. *A real tensor is in the real rank two locus if and only if its flattenings all have rank ≤ 2 and its $2 \times 2 \times 2$ sub-hyperdeterminants are all non-negative.*

Proof. We begin by assuming X has real border rank ≤ 2. Then every $2 \times 2 \times 2$ sub-tensor X' also has real border rank ≤ 2. We can approximate X' by a sequence of tensors X'' that have real rank two. The entries x''_{ijk} of any tensor in the approximating sequence can be written as $x''_{ijk} = a_i b_j c_k + d_i e_j f_k$, where the parameters are real. The hyperdeterminant of X'' can be written in terms of this decomposition. It is

$$(a_1 d_2 - a_2 d_1)^2 (b_1 e_2 - b_2 e_1)^2 (c_1 f_2 - c_2 f_1)^2.$$

This expression is non-negative since all parameters are real. By continuity, we conclude that all $2 \times 2 \times 2$ sub-hyperdeterminants of the original tensor X are non-negative.

Conversely, suppose that X is a real tensor whose $2 \times 2 \times 2$ sub-hyperdeterminants are all non-negative. The complex rank of X is either 1, 2 or ≥ 3. If it is 1 then X is in the real Segre variety, and in particular it is in the real rank two locus. If X has complex rank ≥ 3 then it is in the closure of the rank two tensors, but not rank two. Hence it must lie on a tangent line to the Segre variety. The tangent line can be chosen to be real, as follows. A real tensor on a tangent line can be decomposed as a sum

$$y^{(1)} \otimes x^{(2)} \otimes \cdots \otimes x^{(d)} + x^{(1)} \otimes y^{(2)} \otimes x^{(3)} \otimes \cdots \otimes x^{(d)} + \cdots + x^{(1)} \otimes \cdots \otimes x^{(d-1)} \otimes y^{(d)},$$

where the vectors $x^{(i)}$ and $y^{(j)}$ are real. This expression lies on the limit of secant lines spanned by $(x^{(1)} + \epsilon y^{(1)}) \otimes \cdots \otimes (x^{(d)} + \epsilon y^{(d)})$ and $x^{(1)} \otimes \cdots \otimes x^{(d)}$ as $\epsilon \to 0$. Hence a tensor of this form lies in the real rank two locus.

It remains to consider the case when X has complex rank two and real rank ≥ 3. The tensor X lies on a real secant line, spanned by a pair of complex conjugate points on the Segre variety. Consider any $2 \times 2 \times 2$ sub-tensor X' of X. We can write the entries x'_{ijk} of X' as

$$x'_{ijk} = (a_i + A_i\sqrt{-1})(b_j + B_j\sqrt{-1})(c_k + C_k\sqrt{-1}) + (a_i - A_i\sqrt{-1})(b_j - B_j\sqrt{-1})(c_k - C_k\sqrt{-1}),$$

where the parameters a, b, c, A, B, C are real. The hyperdeterminant of X' evaluates to

$$-(a_1A_2 - a_2A_1)^2 \cdot (b_1B_2 - b_2B_1)^2 \cdot (c_1C_2 - c_2C_1)^2 \cdot 4^3. \tag{2.2}$$

This expression is non-positive since all parameters are real. Our hypothesis that all $2\times2\times2$ sub-hyperdeterminants are non-negative means they must all be zero.

The rank two representation of X involves pairs of vectors $\{a, A\} \subset \mathbb{R}^{n_1}$, $\{b, B\} \subset \mathbb{R}^{n_2}$, $\{c, C\} \subset \mathbb{R}^{n_3}, \ldots$ Every $2 \times 2 \times 2$ sub-hyperdeterminant of X has the form in Equation (2.2) and equates to zero. From this we conclude that, for all but two of the pairs $\{a, A\}, \{b, B\}, \{c, C\}$, etc., the vectors in the pair are linearly dependent. If not, we could construct a non-vanishing sub-hyperdeterminant by choosing indices (i, j) from each vector pair for which the expression $a_iA_j - a_jA_i$ does not vanish. Hence X is the tensor product of a matrix with $d - 2$ vectors. This has real rank two, a contradiction. □

Example 2.3. *Consider a $2 \times 2 \times 2 \times 2$ tensor with complex rank two and real rank ≥ 3. Its entries x_{ijkl} have the parametric representation*

$$\begin{aligned} x_{ijkl} = \quad & (a_i + A_i\sqrt{-1})(b_j + B_j\sqrt{-1})(c_k + C_k\sqrt{-1})(d_l + D_l\sqrt{-1}) \\ + \quad & (a_i - A_i\sqrt{-1})(b_j - B_j\sqrt{-1})(c_k - C_k\sqrt{-1})(d_l - D_l\sqrt{-1}). \end{aligned}$$

I now describe the last part of the proof of Theorem 2.2 for this tensor. Suppose that the eight $2 \times 2 \times 2$ sub-hyperdeterminants of T are all non-negative. They are

$$\begin{array}{l} -(a_0^2 + A_0^2)^2(b_0B_1 - b_1B_0)^2(c_0C_1 - c_1C_0)^2(d_0D_1 - d_1D_0)^2 4^3, \\ -(a_1^2 + A_1^2)^2(b_0B_1 - b_1B_0)^2(c_0C_1 - c_1C_0)^2(d_0D_1 - d_1D_0)^2 4^3, \\ -(b_0^2 + B_0^2)^2(a_0A_1 - a_1A_0)^2(c_0C_1 - c_1C_0)^2(d_0D_1 - d_1D_0)^2 4^3, \\ -(b_1^2 + B_1^2)^2(a_0A_1 - a_1A_0)^2(c_0C_1 - c_1C_0)^2(d_0D_1 - d_1D_0)^2 4^3, \\ -(c_0^2 + C_0^2)^2(a_0A_1 - a_1A_0)^2(b_0B_1 - b_1B_0)^2(d_0D_1 - d_1D_0)^2 4^3, \\ -(c_1^2 + C_1^2)^2(a_0A_1 - a_1A_0)^2(b_0B_1 - b_1B_0)^2(d_0D_1 - d_1D_0)^2 4^3, \\ -(d_0^2 + D_0^2)^2(a_0A_1 - a_1A_0)^2(b_0B_1 - b_1B_0)^2(c_0C_1 - c_1C_0)^2 4^3, \\ -(d_1^2 + D_1^2)^2(a_0A_1 - a_1A_0)^2(b_0B_1 - b_1B_0)^2(c_0C_1 - c_1C_0)^2 4^3. \end{array} \tag{2.3}$$

Note that the first factor does not appear in Equation (2.2) because the fixed indices were subsumed into the expressions for one of the parameter pairs $\{a, A\}, \{b, B\}, \{c, C\}$.

It cannot be that a_0, A_0, a_1, A_1 are all zero, and similarly for the other letters. Hence

$$(a_0A_1 - a_1A_0)(b_0B_1 - b_1B_0)(c_0C_1 - c_1C_0) = (a_0A_1 - a_1A_0)(b_0B_1 - b_1B_0)(d_0D_1 - d_1D_0) =$$
$$(a_0A_1 - a_1A_0)(c_0C_1 - c_1C_0)(d_0D_1 - d_1D_0) = (b_0B_1 - b_1B_0)(c_0C_1 - c_1C_0)(d_0D_1 - d_1D_0) = 0.$$

Two of the four factors are zero. There are six cases. Up to relabeling, $a_0A_1 - a_1A_0 = b_0B_1 - b_1B_0 = 0$. This implies that $T = (a_0, a_1) \otimes (b_0, b_1) \otimes U$, where U is a 2×2-matrix. Clearly U has real rank ≤ 2. This shows that T has real rank ≤ 2, the necessary contradiction.

Example 2.4. *Consider symmetric tensors of format $2 \times 2 \times 2$. Under the correspondence from Example 1.7, they are binary cubics, which can be parametrized by*

$$f(s,t) = x_0 s^3 + 3x_1 s^2 t + 3x_2 s t^2 + x_3 t^3.$$

The set of symmetric rank one tensors, considered up to scale, is the twisted cubic curve in $\mathbb{P}^3$, the Veronese variety $\nu_3(\mathbb{P}^1)$. A membership test for the real rank two tenors is obtained by specializing the $2\times2\times2$ hyperdeterminant from Equation (1.15) to symmetric tensors. We obtain the condition $D \geq 0$, where

$$D = x_0^2 x_3^2 - 6x_0 x_1 x_2 x_3 - 3x_1^2 x_2^2 + 4x_1^3 x_3 + 4x_0 x_2^3 = \det \begin{pmatrix} x_0 & 2x_1 & x_2 & 0 \\ 0 & x_0 & 2x_1 & x_2 \\ x_1 & 2x_2 & x_3 & 0 \\ 0 & x_1 & 2x_2 & x_3 \end{pmatrix}. \tag{2.4}$$

The polynomial D is the discriminant of the binary cubic. We will see more about the connection between polynomials and symmetric tensors in Chapter 5.

We now have a membership test to test if a tensor is in the real rank two locus, but it relies on the evaluation of many polynomials. The number of sub-hyperdeterminants, see Equation (2.1), grows quickly in the format of the tensor. A tensor of format $n \times \cdots \times n$ has $\frac{1}{8}\binom{d}{3} n^d (n-1)^3$ sub-hyperdeterminants. There are several possibilities for minimizing the number of polynomials that must be tested to certify membership in the real rank two locus. If the tensor is symmetric then some sub-hyperdeterminants are the same polynomial, and do not need to be checked twice. The number of distinct sub-hyperdeterminants of a symmetric $n \times \cdots \times n$ (d times) tensor is

$$\binom{n+d-4}{n-1}\binom{\binom{n}{2}+2}{3},$$

because the fixed indices contribute a polynomial of degree $d-3$ in n variables and each hyperdeterminant is given by a polynomial of degree three in $\binom{n}{2}$ variables. Among these, it is enough to check the hyperdeterminants whose expansion as in Equation (2.3) is a sixth power like $(a_0A_1 - a_1A_0)^6$ times an extraneous factor $\prod_i (a_i^2 + A_i^2)^2$, which further reduces the number to

$$\binom{n+d-4}{n-1}\binom{n}{2}.$$

Each of these symmetric hyperdeterminants is a quartic polynomial, like Equation (2.4).

If a tensor X of format $n_1 \times \cdots \times n_d$ is in the real rank two locus, there exists a change of basis operation in $SL_{n_1} \times \cdots \times SL_{n_d}$, the product of special linear groups, or in $O_{n_1} \times \cdots \times O_{n_d}$, the product of orthogonal groups, such that the only non-zero entries of X with respect to the new basis are in a $2 \times \cdots \times 2$ block, denoted $\tilde{X}$. We can test whether X is real rank two by studying the smaller tensor $\tilde{X}$. Symmetry properties of X can be preserved in $\tilde{X}$. This construction is the subspace representation from Definition 1.10.

In conclusion, we can test if a tensor X has real rank two as follows. First, compute the third largest singular values of the flattenings, and see if they are within some threshold of zero. Next, compute the subspace representation and check the non-negativity of the hyperdeterminants of the resulting $2 \times \cdots \times 2$ tensor, to within some threshold. There are $\binom{d}{d-3} \cdot 2^{d-3}$ sub-hyperdeterminants to test.

2.2 Lower bounds on real rank

The previous section describes how to test if a tensor lies in the real rank two locus. But what about higher ranks? There are very few tools for finding the exact complex rank of a tensor. The main tool is the substitution method, which gives a lower bound on the rank of a tensor be reducing it to a tensor of smaller format. There are even fewer methods to compute the exact real rank of a tensor. In this section I generalize the substitution method to give lower bounds on the real rank of a tensor, using sums of squares certificates. The complex substitution method is the following.

Theorem 2.5 (The substitution method, e.g. [103, §5.3.1])**.** *Let $X \in \mathbb{C}^{n_1 \times n_2 \times n_3}$ be a tensor of rank r. We write $X = \sum_{i=1}^{n_1} e_i \otimes M_i$, where $\{e_i : 1 \leq i \leq n_1\}$ are the elementary basis vectors, and the M_i are $n_2 \times n_3$ matrices, known as the slices of the tensor. Reordering indices to ensure that $M_{n_1} \neq 0$, there exist constants $\lambda_1, \ldots, \lambda_{n_1-1}$ such that the following $(n_1 - 1) \times n_2 \times n_3$ tensor has rank at most $r - 1$:*

$$\sum_{i=1}^{n_1-1} e_i \otimes (M_i - \lambda_i M_{n_1}).$$

If M_{n_1} has rank one, then for this choice of λ_i the tensor above has rank exactly $r - 1$.

The following is the analogue for real ranks.

Theorem 2.6 (The substitution method over $\mathbb{R}$)**.** *Let $X \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ be a tensor of real rank r. We can write X in terms of its slices as $X = \sum_{i=1}^{n_1} e_i \otimes M_i$, where $\{e_i : 1 \leq i \leq n_1\}$ are the elementary basis vectors, and the M_i are $n_2 \times n_3$ real matrices. Reordering indices such that $M_{n_1} \neq 0$, there exist real constants $\lambda_1, \ldots, \lambda_{n_1-1}$ such that the following $(n_1-1) \times n_2 \times n_3$ real tensor has real rank at most $r - 1$:*

$$\sum_{i=1}^{n_1-1} e_i \otimes (M_i - \lambda_i M_{n_1}).$$

If M_{n_1} has rank one, then for this choice of λ_i the real tensor above has real rank exactly $r - 1$.

Proof. Assume X has real rank r, with real rank decomposition $X = X_1 + \cdots + X_r$. We can express each rank one tensor in the decomposition as $X_k = \sum_{i=1}^{n_1} \mu_{ki} e_i \otimes L_k$ where the

μ_{ki} are real scalars and L_k is a rank one real matrix. The slices of X can then be expressed as $M_i = \sum_{k=1}^{r} \mu_{ki} L_k$. By the assumption that M_{n_1} is non-zero, we can reorder the terms in the decomposition such that $\mu_{rn_1} \neq 0$. Setting $\lambda_i = \mu_{ri}$, the tensor $\sum_{i=1}^{n_1-1} e_i \otimes (M_i - \lambda_i M_{n_1})$ has all slices expressible as a linear combination of $L_1, \ldots, L_{r-1}$, and hence it has real rank at most $r-1$. The last sentence follows from the fact that if M_{n_1} has rank one, subtracting multiples of it can change the real rank by at most one. □

We saw in Example 1.7 that symmetric tensors are in correspondence with homogeneous polynomials. We illustrate Theorem 2.6 on the cubic surface (or symmetric $4 \times 4 \times 4$ tensor) $f = z_1(z_1^2 - z_2^2 - z_3^2 - z_4^2)$. Since the real and complex ranks differ [41], the usual substitution method in Theorem 2.5 does not give a tight lower bound on the real rank. We use the real substitution method to bound the real rank below by seven. We consider ranks of cubic surfaces further in Chapter 5.

Proposition 2.7. *The cubic surface $f = z_1(z_1^2 - z_2^2 - z_3^2 - z_4^2)$ has real rank at least seven.*

Proof. We use Theorem 2.6 to lower bound the rank of f. For computational convenience we scale the cubic, leaving the rank unchanged, to $z_1(z_1^2 - 3z_2^2 - 3z_3^2 - 3z_4^2)$ or, as a tensor,

$$\left[\begin{array}{cccc||cccc||cccc||cccc} 1 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & -1 \\ 0 & -1 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \end{array}\right].$$

We subtract off an arbitrary multiple of the slices of the tensor to give a $4 \times 4 \times 2$, $4 \times 2 \times 2$, and finally a $2 \times 2 \times 2$ tensor. We show that there do not exist *real* multiples that can be subtracted to give a tensor of zeros. If the pairs of slices we subtract are linearly independent, Theorem 2.6 then implies that the real non-symmetric rank of f is at least $1+2+2+2 = 7$.

Subtracting off three pairs of slices of f in multiples $s_i, t_i, u_i, v_i, w_i, x_i$, where $i = 1, 2$ denotes which of the two slices we subtract from, gives the $2 \times 2 \times 2$ tensor

$$\left[\begin{array}{cc||cc} s_1u_1 + t_1v_1 + t_1x_1 - v_1x_1 + 1 & s_2u_1 + t_2v_1 + t_2x_1 - w_1 & t_1x_2 - v_1x_2 + s_1 - u_1 & t_2x_2 + s_2 - w_2 \\ s_1u_2 + t_1v_2 - v_2x_1 - w_1 & s_2u_2 + t_2v_2 - 1 & -v_2x_2 - u_2 - w_2 & 0 \end{array}\right].$$

We show that the ideal generated by the eight entries does not contain any real points. Eliminating $w_1, w_2, x_1, x_2, t_1, t_2$ gives the hypersurface $(s_2u_1+s_1u_2)^2+(s_1-u_1)^2+(s_2+u_2)^2+(s_2v_1 + u_2v_1 + s_1v_2 - u_1v_2)^2 = 0$. Over the reals, this is zero if and only if the individual squares in the sum vanish, hence $s_1 = u_1$ and $s_2 = -u_2$. The ideal obtained by eliminating w_1, w_2, x_1, x_2, v_2 then has equation $(t_2u_1 + t_1u_2)^2 + (u_1u_2 - t_2v_1)^2 + t_1^2 + t_2^2 + u_1^2 + u_2^2 = -1$, which has no real solutions. This concludes the main case.

It remains to consider the case when some pairs of slices of the tensor are linearly dependent. The first and second pairs of slices we subtract are always linearly independent, taking us to a $4 \times 2 \times 2$ tensor whose real rank is four less than that of f. The third pair of slices are dependent only if $t_1 = v_1$ and $t_2 = -v_2$. The result then follows as above, by

choosing a different pair of slices to subtract, unless $s_1 = u_1$ and $s_2 = -u_2$. In this case, the $4 \times 2 \times 2$ tensor has four slices spanned by

$$M_1 = \begin{bmatrix} s_1u_1 + t_1v_1 + 1 & s_2u_1 + t_2v_1 \\ s_1u_2 + t_1v_2 & s_2u_2 + t_2v_2 - 1 \end{bmatrix} \quad \text{and} \quad M_2 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix},$$

with first slice a scalar multiple of M_1, and the remaining three slices scalar multiples of M_2. There does not exist a real multiple of M_2 that can be subtracted from M_1 to give a rank one matrix, because $\det(M_1 - w_1M_2) = -(u_2v_1 - u_1v_2)^2 - u_1^2 - u_2^2 - v_1^2 - v_2^2 - w_1^2 - 1 = 0$ has no solutions over the reals. By Theorem 2.6, the real rank of the $4 \times 2 \times 2$ tensor is at least three. Hence we obtain an overall lower bound of $3 + 2 + 2 = 7$ on the real rank. □

2.3 Alternative ranks

Why are low rank decompositions and approximations useful? The central idea is that rank one terms can be readily interpreted, and this allows us to think of a tensor of data as a mixture of signals. The rank one tensors are building blocks in which the variables have been de-coupled from one another. But un-coupling variables is not the only interpretable signal. There are other building blocks that give the right notion of interpretability for different contexts. This requires us to generalize rank to sums of terms of a different kind.

In this section I generalize the study of real ranks of tensors to decompositions involving other building blocks, via a geometric framework also studied in [31]. In the next section, I apply this geometric framework to give a description of the boundary of the real rank two tensors.

Let $\mathcal{X}$ be an irreducible variety in $\mathbb{C}^N$, a cone over a projective variety in $\mathbb{P}^{N-1}$, whose real points are Zariski dense in the complex variety. The *$\mathcal{X}$-rank* of a point $X \in \mathbb{C}^{N+1}$ is the smallest r such that X can be written as a sum of r terms $X = X_1 + \cdots + X_r$, where $X_1, \ldots, X_r \in \mathcal{X}$. If X is real then its *real $\mathcal{X}$-rank* is the smallest r such that there exists a decomposition in which all the X_i are real points on $\mathcal{X}$. The loci of $\mathcal{X}$-rank $\leq r$ and real $\mathcal{X}$-rank $\leq r$ may not be closed. We define the *(real) $\mathcal{X}$-border rank* of X to be the smallest r such that a vector can be written as a limit of vectors of (real) $\mathcal{X}$-rank r.

The secant variety $\sigma(\mathcal{X})$ is the set of points of $\mathcal{X}$-border rank ≤ 2. Geometrically, it is the closure of the set of points in $\mathbb{C}^N$ that lie on a line spanned by two points in $\mathcal{X}$. The *tangential variety* $\tau(\mathcal{X})$ is the closure of the set of points in $\mathbb{C}^N$ that lie on a tangent line to $\mathcal{X}$ at a smooth point. See Figure 2.1 for a picture of a point X on a secant line and a tangent line. Tangent lines can be obtained from secant lines by taking limits, hence the tangential variety is a subvariety of the secant varieties. For the secant and tangential varieties above, the Euclidean closure and Zariski closure coincide, by Theorem 1.21. If the inclusion $\tau(\mathcal{X}) \subset \sigma(\mathcal{X})$ is strict then, by [191, Theorem 1.4], both varieties have the expected dimensions:

$$\dim(\sigma(\mathcal{X})) = 2 \cdot \dim(\mathcal{X}) \qquad \text{and} \qquad \dim(\tau(\mathcal{X})) = 2 \cdot \dim(\mathcal{X}) - 1,$$

and the variety $\mathcal{X}$ is called *non-defective*. Conversely, if $\tau(\mathcal{X}) = \sigma(\mathcal{X})$ then $\mathcal{X}$ is called *defective*.

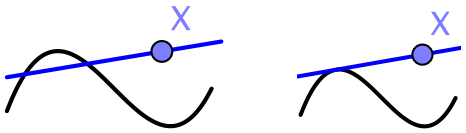


Figure 2.1: A point on a secant line (left) and a tangent line (right).

The main object of interest in this section is the *real rank two locus* of $\mathcal{X}$, denoted by $\rho(\mathcal{X})$, which consists of points that can be written as limits of sums $X_1 + X_2$ where the X_i are real points on $\mathcal{X}$. Geometrically, $\rho(\mathcal{X})$ is the Euclidean closure of the set of points that lie on a line spanned by two points real points in $\mathcal{X}$. The denseness of real points on $\mathcal{X}$ ensures that $\rho(\mathcal{X})$ is Zariski dense in $\sigma(\mathcal{X})$. The inclusion of the closed set $\rho(\mathcal{X})$ into the real points on the secant variety can be strict, as we saw for tensors in Section 2.1. The difference between the real points on $\sigma(\mathcal{X})$ and $\rho(\mathcal{X})$ consists of points of $\mathcal{X}$-rank two whose real $\mathcal{X}$-rank exceeds two. The picture to have in mind is Figure 2.2.

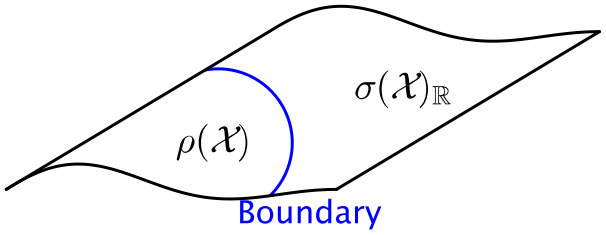


Figure 2.2: The real rank two locus $\rho(\mathcal{X})$ can be a strict subset of the real points on the secant variety $\sigma(\mathcal{X})_{\mathbb{R}}$.

The *real rank two boundary*, denoted by $\partial(\rho(\mathcal{X}))$ is the set $\rho(\mathcal{X})$ minus its relative interior, where relative refers to $\sigma(\mathcal{X})_{\mathbb{R}}$ being the ambient topological space. The Zariski closure of the set $\partial(\rho(\mathcal{X}))$ in $\mathbb{C}^N$, denoted $\partial_{\text{alg}}(\rho(\mathcal{X}))$, is called the *algebraic real rank two boundary* of $\mathcal{X}$.

We need one more definition. Let p and q be distinct smooth points on $\mathcal{X}$ whose corresponding tangent spaces $T_p(\mathcal{X})$ and $T_q(\mathcal{X})$ intersect non-trivially. The secant line spanned by such p and q is called an *edge* of $\mathcal{X}$, see Figure 2.3. The Euclidean closure of the union of all edges of $\mathcal{X}$ is a Zariski closed subset in $\mathbb{C}^N$ called the *edge variety* $\epsilon(\mathcal{X})$.

We can specialize this framework to study real ranks of tensors, by taking the projective variety $\mathcal{X}$ to be the Segre variety of rank one tensors, or the Veronese variety of symmetric rank one tensors. In this case the variety $\mathcal{X}$ is defined by the vanishing of the 2×2 minors of all flattenings of a tensor, and the variety $\sigma(\mathcal{X})$ is given by the vanishing of the 3×3 minors of all flattenings.

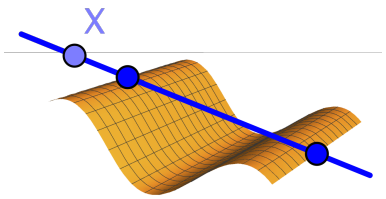


Figure 2.3: A point on an edge, a line shared by two tangent spaces.

In Section 2.1 we saw the importance of binary tensors, those of format $2 \times \cdots \times 2$, in the study of the real rank two locus. I give the defining equation of the above tangential and secant varieties for the case of symmetric binary tensors, or binary forms,

$$f(s,t) \quad = \quad \sum_{i=0}^{d} x_i \binom{d}{i} s^{d-i} t^i.$$

We construct a matrix of format $3 \times (d-1)$.

$$H \quad = \quad \begin{pmatrix} x_0 & x_1 & x_2 & \cdots & x_{d-2} \\ x_1 & x_2 & x_3 & \cdots & x_{d-1} \\ x_2 & x_3 & x_4 & \cdots & x_d \end{pmatrix}. \tag{2.5}$$

The matrix H is Hankel because the entries on each anti-diagonal are the same. The three varieties $\mathcal{X} \subset \tau(\mathcal{X}) \subset \sigma(\mathcal{X})$ are given by the conditions

- $\mathcal{X} = \{\text{rank}(H) \leq 1\} = \{\ell^d\}$ = the cone over the rational normal curve in $\mathbb{P}^d$;
- $\tau(\mathcal{X}) = \overline{\{\ell_1^{d-1}\ell_2\}}$ = closure of points on tangent lines of $\mathcal{X}$;
- $\sigma(\mathcal{X}) = \{\text{rank}(H) \leq 2\} = \overline{\{\ell_1^d + \ell_2^d\}}$ = closure of points on secant lines of $\mathcal{X}$.

These affine varieties have dimensions $2, 3$ and 4. Their defining equations are as follows.

Corollary 2.8. *The prime ideals of $\mathcal{X}$ and $\sigma(\mathcal{X})$ are generated, respectively, by the 2×2-minors and the 3×3-minors of the Hankel matrix H in Equation (2.5). The prime ideal of the tangential variety $\tau(\mathcal{X})$ is minimally generated by the quartic D from Equation (2.4) if $d = 3$, by the cubic $\det(H)$ and the quadric $Q = x_0x_4 - 4x_1x_3 + 3x_2^2$ if $d = 4$, and by $\binom{d-2}{2}$ linearly independent quadrics if $d \geq 5$.*

Proof. The equations for $\mathcal{X}$ and $\sigma(\mathcal{X})$ are classical, see e.g. [105]. The ideal of $\tau(\mathcal{X})$ is derived from [136, 146]. □

The real rank two locus $\rho(\mathcal{X})$ is a 3-dimensional semi-algebraic set. It consists of binary forms $\ell_1^d + \ell_2^d$ where ℓ_1, ℓ_2 are real. More generally, we have the following theorem.

Theorem 2.9. *Let $\mathcal{X}$ be a non-defective variety in $\mathbb{C}^N$, a cone over a projective variety in $\mathbb{P}^{N-1}$, whose real points are Zariski dense. If the algebraic real rank two boundary of $\mathcal{X}$ is non-empty then it is a variety of pure codimension one inside the secant variety $\sigma(\mathcal{X})$. Its irreducible components arise from the tangential variety and the edge variety. In symbols, we have the equi-dimensional inclusion*

$$\partial_{\mathrm{alg}}(\rho(\mathcal{X})) \subseteq \tau(\mathcal{X}) \cup \epsilon(\mathcal{X}).$$

The hypothesis that $\mathcal{X}$ is non-defective is essential for the theorem to hold. The Segre and Veronese varieties are non-defective provided that we are in the setting of tensors of order ≥ 3. The case of matrices is excluded from Theorem 2.9 because the varieties of rank one matrices are defective. A plane curve $\mathcal{X}$ is also defective. In [31, §3], the authors show that $\partial_{\mathrm{alg}}(\rho(\mathcal{X}))$ is a union of flex lines, provided it is non-empty. Such flex lines are not covered by Theorem 2.9.

Proof. The fact that $\partial_{\mathrm{alg}}(\rho(\mathcal{X}))$ is pure of dimension one can be derived from the general result in [172, Lemma 4.2]: if a semi-algebraic set $S \subset \mathbb{R}^k$ is nonempty and contained in the closure of its interior and the same is true for $\mathbb{R}^k \backslash S$, then the algebraic boundary of S is a variety of pure codimension one. Since the property is local, we can here replace $\mathbb{R}^k$ by $\mathcal{X}_{\mathbb{R}}$. The argument below will show that these hypotheses hold here.

For a general real point u on the secant variety $\sigma(\mathcal{X})$, there are only finitely many pairs $\{v_1, w_1\}, \ldots, \{v_k, w_k\}$ of points on $\mathcal{X}$ such that the line spanned by v_i and w_i contains u. The $2k$ non-singular points of $\mathcal{X}$ can be expressed locally as algebraic functions of u, by the Implicit Function Theorem. The point $u \in \sigma(\mathcal{X})_{\mathbb{R}}$ lies in $\rho(\mathcal{X})$ if at least one of these pairs $\{v_i, w_i\}$ consists of two real points, and it lies outside $\rho(\mathcal{X})$ if none of the pairs $\{v_i, w_i\}$ are real. By our assumption that the algebraic real rank two boundary is non-empty, both cases occur.

Consider a general real curve that passes through the boundary $\partial(\rho(\mathcal{X}))$ at a point u^*, and follow the k point pairs along that curve. This uses the Curve Selection Lemma in real algebraic geometry. Precisely one of two scenarios will happen at the transition point.

Case 1: A pair $\{v_i, w_i\}$ of real points merges into a single point on $\mathcal{X}$ and then transitions to a pair of conjugate complex points. As that transition occurs, the secant line degenerates to a tangent line. Hence the corresponding point u^* lies in the tangential variety $\tau(\mathcal{X})$.

Case 2: Two real pairs $\{v_i, w_i\}$ and $\{v_j, w_j\}$ come together, in the sense that v_i and v_j converge to a point $v \in \mathcal{X}$ while w_i and w_j converge to another point $w \in \mathcal{X}$. If this happens then the tangent spaces $T_v(\mathcal{X})$ and $T_w(\mathcal{X})$ meet non-transversally, by the following argument. The secant lines through u arising from the two pairs $\{v_i, w_i\}$ and $\{v_j, w_j\}$ span a plane that contains the line from v_i to v_j and the line from w_i to w_j. In the limit as $v_i, v_j \to v$, $w_i, w_j \to w$ and $u \to u^*$, a line in $T_v(\mathcal{X})$ will be co-planar to a line in $T_w(\mathcal{X})$. The meeting point of the two lines is their non-transverse intersection. Hence the secant line spanned by v and w must be an edge. We conclude that u^* lies in the edge variety $\epsilon(\mathcal{X})$.

The argument above shows that a generic path through $\partial(\rho(\mathcal{X}))$ meets the boundary at either $\tau(\mathcal{X})$ or $\epsilon(\mathcal{X})$. Since the set $\rho(\mathcal{X})$ does not have lower-dimensional components, the Zariski closure of such boundary points is the algebraic real rank two boundary $\partial_{\text{alg}}(\rho(\mathcal{X}))$. Since the two sets $\tau(\mathcal{X})$ and $\epsilon(\mathcal{X})$ are Zariski-closed, it follows that $\partial_{\text{alg}}(\rho(\mathcal{X}))$ is contained in their union $\tau(\mathcal{X}) \cup \epsilon(\mathcal{X})$. □

The motivation for understanding the geometry of the real rank two locus is to approximation a point $X \in \mathbb{R}^N$ by its best approximation of the form $X^* = X_1 + X_2$ where X_i are real points on the variety $\mathcal{X}$. However, since the set of real $\mathcal{X}$-rank two points is not closed, we must optimize over its closure, the real rank two locus $\rho(\mathcal{X})$. The above analysis of this set indicates that there are several possible forms for the best approximation X^* and, importantly, not all of them are of the form $X_1 + X_2$. A priori, there are five possible scenarios:

(a) X^* is the point in $\sigma(\mathcal{X})_{\mathbb{R}}$ that is closest to X, and it is a smooth point of $\sigma(\mathcal{X})$.

(b) X^* is the point in $\mathcal{X}_{\mathbb{R}}$ that is closest to X.

(c) X^* is the point in the singular locus of $\sigma(\mathcal{X})_{\mathbb{R}}$ that is closest to X, but it is not in $\mathcal{X}$.

(d) X^* is the point in $\tau(\mathcal{X})_{\mathbb{R}}$ that is closest to X.

(e) X^* is the point in $\epsilon(\mathcal{X})_{\mathbb{R}}$ that is closest to X.

The following theorem shows that case (b) cannot happen. This was proven for tensors in [175, Lemma 3.4]. The following theorem generalize the result to arbitrary varieties.

Theorem 2.10. *Suppose that $\mathcal{X} \subseteq \mathbb{R}^N$ is an affine variety, a cone over a projective variety in $\mathbb{P}^{n-1}$ which does not lie on a hyperplane. Let $X \in \mathbb{R}^N$ be a data point of real $\mathcal{X}$-border rank greater than r. Let X^* be its best approximation of real $\mathcal{X}$-border rank at most r. Then the real $\mathcal{X}$-border rank of X^* is exactly r, not smaller.*

The best approximation is taken with respect to a weighted Euclidean distance on $\mathbb{R}^N$ where all weights are strictly positive.

Proof. We begin with the case $r = 1$. Then $X \notin \mathcal{X}$ and we wish to show that its best border rank one approximation X^* is non-zero. By assumption, there exists a non-zero vector U in $\mathcal{X}$ that is not in the hyperplane perpendicular to X. This means that $\langle X, U\rangle \neq 0$, where the inner product comes from our choice of norm. The point $\frac{\langle X,U\rangle}{\langle U,U\rangle}U$ also lies in $\mathcal{X}$, and its squared distance to the given data point u is

$$\left\| \frac{\langle X,U\rangle}{\langle U,U\rangle}U - X \right\|^2 = \left\langle \frac{\langle X,U\rangle}{\langle U,U\rangle}U - X, \frac{\langle X,U\rangle}{\langle U,U\rangle}U - X \right\rangle$$

$$= \left(\frac{\langle X,U\rangle}{\langle U,U\rangle}\right)^2 \langle U,U\rangle - 2\frac{\langle X,U\rangle}{\langle U,U\rangle}\langle X,U\rangle + \langle X,X\rangle = \langle X,X\rangle - \frac{\langle X,U\rangle^2}{\langle U,U\rangle}.$$

This is strictly smaller than $\|X - 0\|^2 = \langle X, X \rangle$, so the closest point to X on $\mathcal{X}$ is non-zero.

We now suppose that $r \geq 2$ and let X^* be the best approximation to X among points of real $\mathcal{X}$-border rank at most r. We first suppose for contradiction that X^* has real $\mathcal{X}$-border rank at most $r-1$. We then construct a strictly better border rank r approximation of X by combining X^* with a best rank one approximation for $X - X^*$.

The point $V = X - X^*$ is non-zero. Its best real $\mathcal{X}$-rank one approximation V^* is also non-zero. When $V \notin \mathcal{X}$, we use the first paragraph of the proof to see this; otherwise $V^* = V \neq 0$. The point $X^* + V^*$ still has real $\mathcal{X}$-border rank at most r, and it is closer to X than X^*, since

$$\|X - (X^* + V^*)\| = \|V - V^*\| < \|V - 0\| = \|V\| = \|X - X^*\|.$$

Hence the best approximation to X cannot have real $\mathcal{X}$-border rank strictly less than r. □

The fact that case (b), above, cannot occur for best approximation by $\rho(\mathcal{X})$ is Theorem 2.10 for $r = 2$. The other four cases (a), (c), (d) and (e) are possible in general. Case (a) is the usual best real rank two approximation. Cases (d) and (e) are especially important to understand because the solutions X^* are not critical for the distance function on $\sigma(\mathcal{X})$, so different optimization methods must be used to find such points. We will see how these cases impact real rank two tensor approximation in the next section.

2.4 The real rank two boundary

In practice, a tensor of two real signals arising in an application will not be exactly real rank two due to the presence of noise. In order to recover the two real rank one signals, we seek its best real rank two approximation. This is a projection to the boundary of the real rank two locus. The real rank two locus is denoted $\rho(\mathcal{X})$ and its boundary is $\partial(\rho(\mathcal{X}))$. Here $\mathcal{X}$ is the cone over the Segre variety, the set of rank one tensors.

At first, we might think that our description of the real rank two locus from Theorem 2.2, in terms of equations and inequalities, immediately gives a description of the boundary, by setting the inequalities to zero. However points on the strict interior of the set can also have all sub-hyperdeterminants vanishing, as the following proposition shows.

Proposition 2.11. *When $d \geq 4$ there exist tensors with all sub-hyperdeterminants vanishing on the interior of $\rho(\mathcal{X})$.*

Proof. Let $\mathcal{X}$ be the affine cone over $\text{Seg}(\mathbb{P}^1 \times \mathbb{P}^1 \times \mathbb{P}^1 \times \mathbb{P}^1)$. Let $\{e_1, e_2\}$ be the standard basis of $\mathbb{R}^2$. The rank two tensor

$$X = e_1 \otimes e_1 \otimes e_1 \otimes e_1 + e_2 \otimes e_2 \otimes e_2 \otimes e_2$$

is in the relative interior of the real rank two locus $\rho(\mathcal{X})$. All eight $2 \times 2 \times 2$ sub-tensors have rank one, so the eight hyperdeterminants vanish. This tensor can now be embedded into all larger formats, and we get the conclusion for $d \geq 4$. □

We have the following characterization of the algebraic real rank two boundary for tensors.

Theorem 2.12. *Let $\mathcal{X}$ be the Segre variety (resp. the Veronese variety) whose points are d-dimensional tensors (resp. symmetric tensors) up to scale, and of rank one, where $d \geq 3$. The algebraic real rank two boundary of $\mathcal{X}$ is equal to the tangential variety of $\mathcal{X}$. In symbols,*

$$\partial_{\text{alg}}(\rho(\mathcal{X})) = \tau(\mathcal{X}).$$

Proof. The secant variety $\sigma(\mathcal{X})$ is identifiable, since Kruskal's Theorem [93] holds generically for rank two tensors. Therefore $\epsilon(\mathcal{X})$ does not exist, since points on $\epsilon(\mathcal{X})$ are limits of tensors lying on at least two distinct secant lines. To prove the theorem, we must exclude the possibility $\partial_{\text{alg}}(\rho(\mathcal{X}))$ is empty. By taking sums of complex conjugate pairs of points in $\mathcal{X}$, we can create many tensors that lie in $\sigma(\mathcal{X})_{\mathbb{R}}$ but not in $\rho(\mathcal{X})$. Hence the rank two locus $\rho(\mathcal{X})$ has a non-empty boundary inside $\sigma(\mathcal{X})_{\mathbb{R}}$, and the algebraic boundary $\partial_{\text{alg}}(\rho(\mathcal{X}))$ is a non-empty hypersurface in $\sigma(\mathcal{X})$. That hypersurface is contained in the irreducible hypersurface $\tau(\mathcal{X})$, by Theorem 2.9. This implies that they are equal. □

Theorem 2.12 shows that, for real rank two tensor approximation, we do not need to consider the possibility that the best approximation lies on the edge variety. However, we do need to consider the case where it lies on the tangential variety. The presence of such points gives the geometric reason behind numerical issues in low rank tensor approximation such as [174].

The best real rank two approximation can also lie in the singular locus of the variety $\sigma(\mathcal{X})$. This indicates that it has special structure, as seen in the following example.

Example 2.13. *Consider the space of $3 \times 3 \times 3$ tensors. The variety of rank one tensors $\mathcal{X}$ is the affine cone over $\text{Seg}(\mathbb{P}^2 \times \mathbb{P}^2 \times \mathbb{P}^2)$. The singular locus of $\sigma(\mathcal{X})$ has three irreducible components, given by $\mathbb{P}^2 \times \sigma(\text{Seg}(\mathbb{P}^2 \times \mathbb{P}^2))$ and its permutations, see [123, Cor. 7.17]. These parametrize tensors $v \otimes M$, where $v \in \mathbb{R}^3$ and M is a 3×3-matrix of rank two. Consider the tensor $X = v \otimes M'$ where M' is a general real 3×3-matrix. Let M be the best rank two approximation of M'. The entries of $v \otimes M$ are three copies of M, multiplied by coefficients v_1, v_2 and v_3. The tensor $X^* = v \otimes M$ gives the unique best approximation to X in all three slices, hence X^* is the best approximation to X in $\rho(\mathcal{X})$.*

2.5 Space curve rank

In this section I examine the geometry of computing rank with respect to a curve in real projective space $\mathbb{P}^3$. We seek to express a point in the form $X_1 + X_2$, where the X_i are real points on a fixed curve $\mathcal{X}$. This section builds on [31, §3], in which the authors characterize the real rank two locus $\rho(\mathcal{X})$ for a curve $\mathcal{X}$ in the plane $\mathbb{P}^2$. We assume that the curve $\mathcal{X}$ does not lie in a plane and that its real points are Zariski dense in $\mathcal{X}$.

The boundary of the real rank two locus of a variety $\mathcal{X}$, can consist of the tangential variety $\tau(\mathcal{X})$ or the edge variety $\epsilon(\mathcal{X})$, see Theorem 2.9. The following result shows that all possible combinations of the two varieties can occur.

Proposition 2.14. *There exist rational varieties $\mathcal{X}_1, \mathcal{X}_2, \mathcal{X}_3$ and $\mathcal{X}_4$, cones over curves in $\mathbb{P}^3$, such that*

$$\partial_{\text{alg}}(\mathcal{X}_1) = \tau(\mathcal{X}_1), \quad \partial_{\text{alg}}(\mathcal{X}_2) = \epsilon(\mathcal{X}_2) \cup \tau(\mathcal{X}_2), \quad \partial_{\text{alg}}(\mathcal{X}_3) = \emptyset, \quad \text{and} \quad \partial_{\text{alg}}(\mathcal{X}_4) = \epsilon(\mathcal{X}_4).$$

Proof. By Theorem 2.12, the twisted cubic curve in Example 2.4 is an example for $\mathcal{X}_1$. The quartic curve in [164, Example 3.4] is an example of $\mathcal{X}_2$. For $\mathcal{X}_3$ we take the Morton curve from [147, Example 4.4]. This is rational of degree six and forms a trefoil knot, see [147, Figure 3].

Rational curves $\mathcal{X}_4$ in $\mathbb{P}^3_{\mathbb{R}}$ with $\partial_{\text{alg}}(\mathcal{X}_4) = \epsilon(\mathcal{X}_4)$ are a bit harder to find. A piecewise-linear connected example, resembling a 3D Peano curve, can be constructed in two steps. First, we make a curve from six edges of the unit cube. Starting from $(0, 0, 0)$, the curve travels to $(1, 1, 1)$ via intermediate vertices $(1, 0, 0)$ and $(1, 1, 0)$, and then loops back to $(0, 0, 0)$ via intermediate vertices $(0, 1, 1)$ and $(0, 0, 1)$. In the middle third of each line segment we insert a piecewise linear detour of height $\frac{1}{2}$ in the direction of the next segment. Four views of this space curve are shown in Figure 2.4. There are relatively few viewpoints from which the curve has no crossings. From such positions, crossings are always gained in pairs, via transitions along the edge variety.

The existence of a rational algebraic curve $\mathcal{X}_4$ with the same property can be concluded from the Weierstrass Approximation Theorem. To exclude the possibility that the algebraic boundary is strictly contained in the edge variety, it suffices to show the existence of an approximating curve whose edge variety is irreducible. This can be ensured using [148, Equation (3.6)], since the rational curve $\mathcal{X}_4$ can be parametrized by sufficiently generic polynomials. □

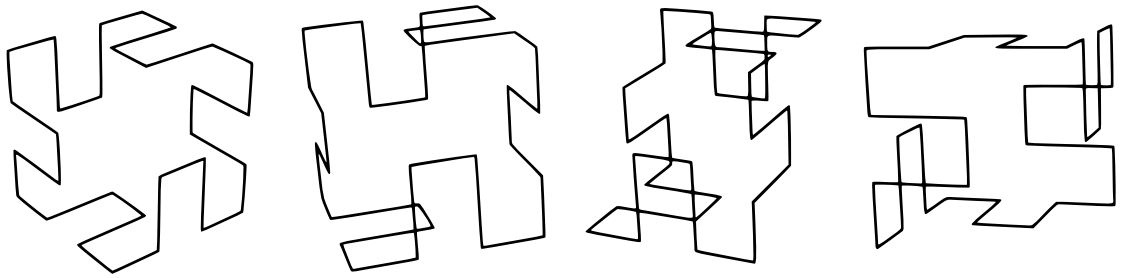


Figure 2.4: A space curve whose algebraic real rank two boundary consists only of the edge variety, seen from four angles.

Real space curve rank admits a visual interpretation. Consider a point $X \in \mathbb{P}^3_{\mathbb{R}}$, and the plane curve in $\mathbb{P}^2$ obtained by projecting $\mathcal{X}$ from the center X. If the projected curve has a node, then the point X is real rank two, because the node is the projection of a line spanned

by two real points on $\mathcal{X}$ that passes through X. The existence of such a line is exactly the geometric condition of being real rank two. Conversely, if the projection does not have any crossings, the point X has real $\mathcal{X}$-border rank at least three. If the curve is a knot or link, then every planar projection has a crossing and all points are in the real rank two locus.

We can also visualize the transition between real ranks two and three. As X moves through space and transitions from real $\mathcal{X}$-rank two to real $\mathcal{X}$-rank three, all ordinary real double points disappear from the projected curve. If the transition occurs via $\tau(\mathcal{X})$ then the intermediate singularity of the projected curve is a cusp. If it occurs via $\epsilon(\mathcal{X})$ then that singularity is known classically as a *tacnode*. Figure 2.5 shows the two possible transitions. The transitions are two of the three classical *Reidemeister moves* from knot theory. Transitions via the third Reidemeister move do not cause a change in real X-rank.

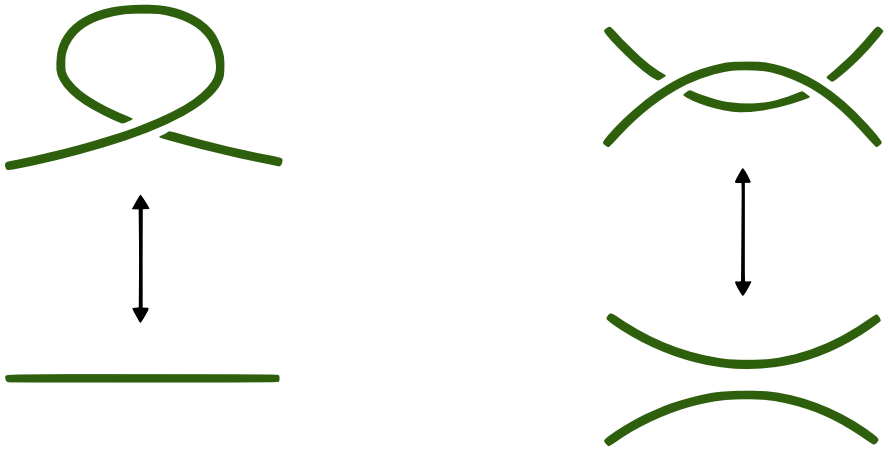


Figure 2.5: The transitions between real space curve ranks two and three via the tangential surface (left) and the edge surface (right). The arrows indicate the direction of change in viewpoint.

Chapter 3

Singular vectors

Singular vectors of matrices are central to low rank approximation and visualization of matrix data, as we saw in Section 1.2. A singular vector pair (u, v) is characterized by the property that Mv is parallel to u, and $M^{\mathsf{T}}u$ is parallel to v. Singular vectors can also be defined for a tensor. A tensor has a singular vector tuple, one vector for each index of the tensor. Multiplying an order d tensor by $(d-1)$ vectors contracts all but one of the indices to give a vector. In the matrix (order two tensor) case, multiplying the matrix by one vector gives a vector. Extending the singular vector condition from order two to general order d, we have the following definition, see Definition 1.14.

A singular vector tuple $(v^{(1)}, \ldots, v^{(d)})$ of an order d tensor X is a tuple of d vectors, such that every subset of $(d-1)$ vectors in the tuple gives a contracted tensor parallel to the remaining vector. That is, the singular vector tuple of an order d tensor must satisfy d conditions:

$$
\begin{array}{lllllllllllll}
[\![X; & v^{(1)} & , & v^{(2)} & , & \ldots & , \ldots , & v^{(d-2)} & , & v^{(d-1)} & , & \cdot &]\!] \text{ is parallel to } v^{(d)} \\
[\![X; & v^{(1)} & , & v^{(2)} & , & \ldots & , \ldots , & v^{(d-2)} & , & \cdot & , & v^{(d)} &]\!] \text{ is parallel to } v^{(d-1)} \\
 & & & & & & & & & & & & \vdots \\
[\![X; & v^{(1)} & , \ldots , & v^{(j-1)} & , & \cdot & , & v^{(j+1)} & , & \ldots & , & v^{(d)} &]\!] \text{ is parallel to } v^{(j)} \\
 & & & & & & & & & & & & \vdots \\
[\![X; & v^{(1)} & , & \cdot & , & v^{(3)} & , \ldots , & \ldots & , & v^{(d-1)} & , & v^{(d)} &]\!] \text{ is parallel to } v^{(2)} \\
[\![X; & \cdot & , & v^{(2)} & , & v^{(3)} & , \ldots , & \ldots & , & v^{(d-1)} & , & v^{(d)} &]\!] \text{ is parallel to } v^{(1)}.
\end{array}
$$

A general tensor has finitely many singular vector tuples. The number of singular vector tuples is the coefficient of a particular polynomial, given in [70]. Eigenvectors of tensors can be defined similarly. A vector v is an eigenvector of a tensor of format $n \times \cdots \times n$ (d times) if it satisfies the above equations, with $v = v^{(j)}$ for all $j \in \{1, \ldots, d\}$. For a study of the configurations of vectors that can arise as the eigenvectors of a tensor, see [2].

Once we fix a scaling of the vectors, the singular vector tuples of a tensor can also be defined using a variational approach, see [109]. If we impose that all the vectors in the tuple have norm one, then the singular vector tuples of a tensor are the critical points of the

optimization problem

$$\begin{aligned} \text{maximize} \quad & [\![X; v^{(1)}, \ldots, v^{(d)}]\!] \\ \text{subject to} \quad & \|v^{(1)}\| = \cdots = \|v^{(d)}\| = 1. \end{aligned}$$

Remark 3.1 (Nomenclature warning)**.** *Several different names can be given to singular vectors and singular values of tensors, see [44, 145]. The names depend on if the singular value is real, or if the equation is homogenized (cf. the discussion of singular values on Page 20). In this thesis, I only distinguish between two types of singular vectors and values. I consider singular vector tuples of a tensor, as above, with a singular value as in Definition 1.15. I also study the higher-order singular values, obtained from flattenings of a tensor, see Section 1.3 and Chapter 4.*

The singular value decomposition expresses a matrix as a sum of rank one matrices of orthogonal vectors. Not all tensors possess a decomposition into rank one tensors of orthogonal vectors, for the following reason. There are at most n_j orthonormal vectors in $\mathbb{R}^{n_j}$. Hence a decomposition of a tensor of format $n_1 \times \cdots \times n_d$ into rank one tensors of orthogonal vectors has rank at most $\min(n_1, \ldots, n_d)$. However, the rank of a general tensor of this format is higher. The number of parameters in the secant variety $\sigma_r(\mathrm{Seg}(\mathbb{P}^{n_1-1} \times \cdots \times \mathbb{P}^{n_d-1}))$ upper bounds the dimension of the set of rank r tensors, and is equal to $r(\sum_{i=1}^{d}(n_i - 1)) + (r-1)$. The generic rank is the smallest r such that the secant variety fills the space of tensors up to scale, which has dimension $n_1 \cdots n_d - 1$.

The tensors that have a decomposition into rank one terms of orthogonal vectors are called *orthogonally decomposable* tensors. Orthogonally decomposable tensors were first introduced in [192]. They have been applied to parameter estimation in latent variable models [8] and image reconstruction [139]. Finding the decomposition of a general tensor is NP-hard [85], however finding the decomposition of an orthogonally decomposable tensor can be done efficiently via iterative rank one updates [192]. Orthogonally decomposable tensors have also been explored in [92]. The variety of orthogonally decomposable tensors was studied in [33], and the eigenvectors of symmetric orthogonally decomposable tensors were studied in [151].

The best rank one approximation of a tensor is given by the singular vector tuple corresponding to the largest singular value, as we saw in Theorem 1.16. For higher ranks, low rank approximations does not generally involve singular vectors. However, for orthogonally decomposable tensors, low-rank approximation is given by a truncation of singular vector tuples. This always holds for matrices, which are always orthogonally decomposable.

In this chapter, I characterize the singular vector tuples of orthogonally decomposable tensors as a variety in a product of projective spaces. Orthogonally decomposable tensors have infinitely many singular vector tuples. For some examples, I describe how the positive-dimensional spaces of singular vectors adopt generic behavior under a small perturbation. This chapter is based on joint work with Elina Robeva, published in Linear and Multilinear Algebra [153].

3.1 Orthogonally decomposable tensors

Definition 3.2. *A tensor $X \in \mathbb{R}^{n_1 \times \cdots \times n_d}$ is* orthogonally decomposable *if it can be written*

$$X = \sum_{i=1}^{n} \sigma_i v_i^{(1)} \otimes \cdots \otimes v_i^{(d)}, \tag{3.1}$$

where $n = \min(n_1, \ldots, n_d)$, the scalars σ_i are real, and the vectors $v_1^{(j)}, \ldots, v_n^{(j)} \in \mathbb{R}^{n_j}$ are orthonormal for every fixed $j \in \{1, \ldots, d\}$.

The decomposition in Equation (3.1) will in general be unique up to re-ordering the summands. Some singular vector tuples of an orthogonally decomposable tensor can be seen directly from the decomposition. The tuples $(v_i^{(1)}, \ldots, v_i^{(d)})$ occurring as rank one terms in the decomposition are singular vector tuples. This is because, when we multiply the tensor X by the vector $v_i^{(j)}$, the orthogonality of the vectors $v_1^{(j)}, \ldots, v_n^{(j)}$ means we are left with a single rank one term. For generic matrices $M \in \mathbb{R}^{n_1 \times n_2}$, the rank-one terms in the singular value decomposition constitute all of the singular vector pairs. In contrast, orthogonally decomposable tensors have additional singular vector tuples that do not appear as terms in the decomposition.

The definition of a singular vector tuple depends on vectors being parallel, so it is unchanged by scaling the vectors. Hence we can consider the singular vector tuple to lie in the product of projective spaces $\mathbb{P}^{n_1-1} \times \cdots \times \mathbb{P}^{n_d-1}$.

Remark 3.3. *I make a distinction between the cases*

$$[\![X; v^{(1)}, \ldots, v^{(d)}]\!] \neq 0 \qquad \textit{and} \qquad [\![X; v^{(1)}, \ldots, v^{(d)}]\!] = 0.$$

In the former case, the singular vector tuple is a fixed point *of the map of projective space $\mathbb{P}^{n_1-1} \times \cdots \times \mathbb{P}^{n_{j-1}-1} \times \mathbb{P}^{n_{j+1}-1} \times \cdots \times \mathbb{P}^{n_d-1} \to \mathbb{P}^{n_j-1}$ induced by X. In the latter case the singular vector tuple is a* base point *of the map.*

The main theorem of this chapter is the following description of the singular vector tuples of an orthogonally decomposable tensor.

Theorem 3.4. *The projective variety of singular vector tuples of an orthogonally decomposable tensor $X \in \mathbb{R}^{n_1 \times \cdots \times n_d}$, with $d \geq 3$, consists of*

$$\frac{(2^{d-1}(d-2)+1)^n - 1}{2^{d-1}(d-2)}$$

fixed points, of which $\frac{(2^{d-1}+1)^n - 1}{2^{d-1}}$ are real, and $\binom{d}{2}^n - c(d-1)^n + \binom{c}{2}$ positive-dimensional components of base points, each a product of linear spaces of dimension $\sum_{j=1}^{d}(n_j - 1) - 2n$. Here, $n = min(n_1, \ldots, n_d)$ and $c = \#\{j : n_j = n\}$.

Note that taking the limit of the fixed point count as $d \to 2$ using e.g. L'Hôpital's rule recovers n, the number of singular vectors of a matrix. The formula for the number of positive-dimensional components equals zero in this case, which is true for a generic matrix. Theorem 3.4 implies that for all but a few small cases the singular vector tuples of an orthogonally decomposable tensor comprise a positive-dimensional variety. This is in contrast to the variety of eigenvectors of a symmetric orthogonally decomposable tensor, which is zero-dimensional [151].

The rest of this chapter is organized as follows. In Section 3.2, I use the theory of binomial ideals [65] to describe the singular vector tuples of an orthogonally decomposable tensor. I describe the positive dimensional components of the variety of singular vector tuples, and then give the proof of Theorem 3.4. In Section 4, I explore the combinatorial structure of the positive dimensional components.

3.2 Singular vector tuples

In this section I give a formula for the singular vector tuples of an orthogonally decomposable tensor. I start by considering a diagonal orthogonally decomposable tensor, in which the orthogonal vectors are the elementary basis vectors. A general orthogonally decomposable tensor can be obtained from a diagonal one by applying an orthogonal change of coordinates.

Lemma 3.5. *Let $S \in \mathbb{R}^{n_1 \times \cdots \times n_d}$ be the tensor*

$$S = \sum_{i=1}^{n} \sigma_i e_i^{(1)} \otimes \cdots \otimes e_i^{(d)},$$

where $d \geq 3$, the scalars $\sigma_1, \ldots, \sigma_n \neq 0$, the vector $e_i^{(j)}$ is the ith basis vector in $\mathbb{R}^{n_j}$, and $n = \min\{n_1, \ldots, n_d\}$. The singular vector tuples of S are as follows.

Type I:

$$\sigma_{\tau(1)}^{-\frac{1}{d-2}} \left(e_{\tau(1)}^{(1)}, e_{\tau(1)}^{(2)}, \ldots, e_{\tau(1)}^{(d)} \right) + \sum_{i=1}^{m} \eta_i \sigma_{\tau(i)}^{-\frac{1}{d-2}} \left(e_{\tau(i)}^{(1)}, \chi_i^{(2)} e_{\tau(i)}^{(2)}, \ldots, \chi_i^{(d)} e_{\tau(i)}^{(d)} \right) \tag{3.2}$$

where $1 \leq m \leq n$, the scalars $\chi_i^{(j)} \in \{\pm 1\}$ are such that $\prod_{j=2}^{d} \chi_i^{(j)} = 1$ for every $i = 1, ..., m$, τ is any permutation on $\{1, \ldots, n\}$, and each scalar η_i is a $(2d-4)$-th root of unity. The tuples with real coordinates are those for which $\eta_i \in \{\pm 1\}$.

Type II: All tuples $(v^{(1)}, \ldots, v^{(d)})$ such that the $n \times d$ matrix $V = (v_i^{(j)})$ has at least two zeros in each row and no column equal to zero.

Before proving Lemma 3.5, I illustrate it with the following example.

Example 3.6. *The orthogonally decomposable tensor* $S = e_1^{(1)} \otimes e_1^{(2)} \otimes e_1^{(3)} + e_2^{(1)} \otimes e_2^{(2)} \otimes e_2^{(3)} \in \mathbb{R}^{2\times 3\times 3}$ *has six Type I singular vector tuples*

$$\left(e_1^{(1)}, e_1^{(2)}, e_1^{(3)}\right), \left(e_2^{(1)}, e_2^{(2)}, e_2^{(3)}\right)$$

$$\left(e_1^{(1)} + e_2^{(1)}, e_1^{(2)} + e_2^{(2)}, e_1^{(3)} + e_2^{(3)}\right), \left(e_1^{(1)} + e_2^{(1)}, e_1^{(2)} - e_2^{(2)}, e_1^{(3)} - e_2^{(3)}\right),$$
$$\left(e_1^{(1)} - e_2^{(1)}, e_1^{(2)} + e_2^{(2)}, e_1^{(3)} - e_2^{(3)}\right), \left(e_1^{(1)} - e_2^{(1)}, e_1^{(2)} - e_2^{(2)}, e_1^{(3)} + e_2^{(3)}\right).$$

The Type II singular vectors are five copies of $\mathbb{P}^1$

$$\left(\square e_1^{(1)} + \square e_2^{(1)}, e_3^{(2)}, e_3^{(3)}\right), \left(e_1^{(1)}, \square e_2^{(2)} + \square e_3^{(2)}, e_3^{(3)}\right), \left(e_1^{(1)}, e_3^{(2)}, \square e_2^{(3)} + \square e_3^{(3)}\right),$$

$$\left(e_1^{(1)}, \square e_1^{(2)} + \square e_3^{(2)}, e_3^{(3)}\right), \left(e_2^{(1)}, e_3^{(2)}, \square e_1^{(3)} + \square e_3^{(3)}\right),$$

where two $\square$*'s in a vector indicate a copy of* $\mathbb{P}^1$ *on those two coordinates. The five copies of* $\mathbb{P}^1$ *intersect in two triple intersections, see Figure 3.1.*

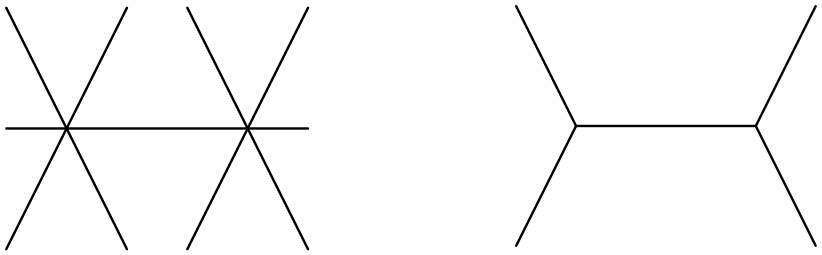


Figure 3.1: An orthogonally decomposable $2 \times 3 \times 3$ tensor has singular vectors that are five copies of $\mathbb{P}^1$ meeting at two triple intersection points (left), depicted as a polyhedral complex (right).

The generic number of singular vector tuples of a $2 \times 3 \times 3$ *tensor is 15, by [70], so the five copies of* $\mathbb{P}^1$ *degenerate from nine points. For example, consider the family of perturbed tensors*

$$S_\epsilon = S + \epsilon X,$$

where X *is the* $2 \times 3 \times 3$ *tensor*

$$\left[\begin{array}{ccc||ccc} 0 & 40 & 10 & 7 & 1 & 1 \\ 100 & 3 & 3 & 8 & 0 & 2 \\ 3 & 2 & 6 & 2 & 2 & 3 \end{array}\right].$$

For ϵ *on the order of* 10^{-6} *we obtain nine points: one point near each copy of* $\mathbb{P}^1$*, and two points of multiplicity 2 near each triple intersection.*

I will return this example in Section 3.3.

Proof of Lemma 3.5. By definition, $(v^{(1)}, \dots, v^{(d)})$ is a singular vector tuple of S if and only if for each $j = 1, \dots, d$ the following matrix has rank at most one:

$$M_{S,j} = \left[[\![S; v^{(1)}, \dots, v^{(j-1)}, \cdot, v^{(j+1)}, \dots, v^{(d)}]\!] \quad v^{(j)} \right] = \begin{bmatrix} \sigma_1 v_1^{(1)} \cdots \widehat{v_1^{(j)}} \cdots v_1^{(d)} & v_1^{(j)} \\ \vdots & \vdots \\ \sigma_n v_n^{(1)} \cdots \widehat{v_n^{(j)}} \cdots v_n^{(d)} & v_n^{(j)} \end{bmatrix}$$

where $v_i^{(j)}$ is the ith entry of the vector $v^{(j)}$, and $\widehat{v_i^{(j)}}$ denotes the omission of $v_i^{(j)}$ from the product. There are three cases to consider.

Case 1: exactly one of the entries $v_i^{(1)}, \dots, v_i^{(d)}$ vanishes, for some $i \in \{1, \dots, n\}$.

Suppose that $v_i^{(j)}$ is the vanishing entry. Then the ith row of the matrix $M_{S,j}$ has first entry non-zero and second entry zero. In order for the matrix to have rank one, we need the whole second column to be zero. Since the contraction $[\![S; v^{(1)}, \dots, v^{(j-1)}, \cdot, v^{(j+1)}, \dots, v^{(d)}]\!]$ lies in the span of $e_1^{(j)}, \dots, e_n^{(j)}$, in order for it to be parallel to $v^{(j)}$ it has to be 0. In particular, its i-th entry $\sigma_i v_i^{(1)} \dots \widehat{v_i^{(j)}} \dots v_i^{(d)}$ has to be zero, a contradiction. Hence it cannot be the case that exactly one of the entries $v_i^{(1)}, \dots, v_i^{(d)}$ is zero for some i.

Case 2: at least two of $v_i^{(1)}, \dots, v_i^{(d)}$ (but not all of them) vanish, for some $i \in \{1, \dots, n\}$.

Assume $v_i^{(j)} \neq 0$. The entry in the ith row and the first column of $M_{S,j}$ is zero, but the entry in the second column is non-zero. In order for the matrix to be rank one, the whole first column must be zero. Therefore, for every k, at least one of the entries $v_k^{(1)}, \dots, v_k^{(j-1)}, v_k^{(j+1)}, \dots, v_k^{(d)}$ is zero. Then by Case 1 at least two $v_k^{(1)}, \dots, v_k^{(d)}$ are zero for every k. Conversely, if for every k at least two of the entries $v_k^{(1)}, \dots, v_k^{(d)}$ are equal to 0, in such a way that $v^{(j)} \in \mathbb{P}^{n_j - 1}$, then $(v^{(1)}, \dots, v^{(d)})$ is a singular vector tuple of S. This gives the singular vector tuples of Type II.

Case 3: for every $i \in \{1, \dots, n\}$, either $v_i^{(1)} = \dots = v_i^{(d)} = 0$ or $\prod_{j=1}^{d} v_i^{(j)} \neq 0$.

After reordering the indices i, we can assume that $v_i^{(1)}, \dots, v_i^{(d)} \neq 0$ for $i \in \{1, \dots, m\}$ and $v_i^{(1)} = \dots = v_i^{(d)} = 0$ for $i \in \{m+1, \dots, n\}$, for some m. Then $(v^{(1)}, \dots, v^{(d)})$ is a singular vector tuple if and only if it satisfies the equations

$$\sigma_i v_i^{(1)} \dots \widehat{v_i^{(j)}} \dots v_i^{(d)} v_l^{(j)} = \sigma_l v_l^{(1)} \dots \widehat{v_l^{(j)}} \dots v_l^{(d)} v_i^{(j)} \quad \text{for all} \quad \begin{matrix} j \in \{1, \dots d\} \\ i, l \in \{1, \dots, m\}. \end{matrix} \tag{3.3}$$

We solve this system of equations by viewing it in the Laurent polynomial ring generated by the entries $v_i^{(j)}$ and their inverses, using the theory of binomial ideal decomposition [65].

We first construct the lattice corresponding to the binomial ideal. Let $L_\rho \subseteq \mathbb{Z}^{d\times m}$ be the lattice generated by

$$\sum_{k=1}^{d}(e_i^{(k)} - e_l^{(k)}) - 2(e_i^{(j)} - e_l^{(j)}) \quad \text{for all} \quad \begin{matrix} j \in \{1,\dots d\} \\ i,l \in \{1,\dots,m\} \end{matrix}$$

where $e_b^{(a)}$ is the elementary basis vector in $\mathbb{Z}^{d\times m}$ with a 1 in coordinate (a,b). Let $\rho : L_\rho \to \mathbb{C}^*$ be the partial character

$$\rho\left(\sum_{k=1}^{d}(e_i^{(k)} - e_l^{(k)}) - 2(e_i^{(j)} - e_l^{(j)})\right) = \frac{\sigma_l}{\sigma_i} \quad \text{for all} \quad \begin{matrix} j \in \{1,\dots d\} \\ i,l \in \{1,\dots,m\} \end{matrix} \tag{3.4}$$

Then the lattice ideal $I(\rho) = \langle v^x - \rho(x) : x \in L_\rho\rangle$ is the system of equations from Equation (3.3), where v^x denotes taking the variables $v_i^{(j)}$ in the ring to the powers indicated by the lattice element x.

We have the inclusion $L_\rho \subseteq L = \langle e_i^{(j)} - e_l^{(j)} : 1 \le j \le d, 1 \le i,l \le m\rangle$. Therefore by [65, Theorem 2.1], the ideal of $I(\rho)$ is the intersection of ideals of partial characters that extend ρ to L, i.e.

$$I(\rho) = \bigcap_{\rho' \text{ extends } \rho \text{ to } L} I(\rho').$$

We find the ideal $I(\rho)$ from the $I(\rho')$, as follows. Summing Equation (3.4) over $1 \le j \le d$ gives

$$\rho\left(\sum_{j=1}^{d}\left(\sum_{k=1}^{d}(e_i^{(k)} - e_l^{(k)}) - 2(e_i^{(j)} - e_l^{(j)})\right)\right) = \rho\left((d-2)\sum_{k=1}^{d}(e_i^{(k)} - e_l^{(k)})\right) = \left(\frac{\sigma_l}{\sigma_i}\right)^d$$

Therefore, any ρ' extending ρ satisfies $\rho'\left(\sum_{k=1}^{d}(e_i^{(k)} - e_l^{(k)})\right) = \phi_{il}\left(\frac{\sigma_l}{\sigma_i}\right)^{\frac{d}{d-2}}$, where ϕ_{il} is a $(d-2)$-th root of unity. By rearranging Equation (3.4), we furthermore see that any such ρ' must satisfy $\rho'\left(2(e_i^{(j)} - e_l^{(j)})\right) = \rho'\left(\sum_{k=1}^{d}(e_i^{(k)} - e_l^{(k)})\right)\left(\frac{\sigma_i}{\sigma_l}\right)$. Combining these yields

$$\rho'\left(e_i^{(j)} - e_l^{(j)}\right) = \phi_{il}^{(j)}\left(\frac{\sigma_l}{\sigma_i}\right)^{\frac{1}{d-2}} \tag{3.5}$$

where the $\phi_{il}^{(j)}$ are $2(d-2)$-th roots of unity such that $(\phi_{il}^{(j)})^2 = \phi_{il}$ for all $j \in \{1,\dots,d\}$. It remains to find the relations that hold for the $\phi_{il}^{(j)}$ as the indices i,l,j vary, so that Equation (3.4) holds. Substituting our expression for ρ' from Equation (3.5) into Equation (3.4) yields

$$\prod_{k=1}^{d}\left(\phi_{il}^{(k)}\left(\frac{\sigma_l}{\sigma_i}\right)^{\frac{1}{d-2}}\right)\phi_{il}^{-1}\left(\frac{\sigma_l}{\sigma_i}\right)^{-\frac{2}{d-2}} = \frac{\sigma_l}{\sigma_i} \quad \Longrightarrow \quad \prod_{k=1}^{d}\phi_{il}^{(k)} = \phi_{il}.$$

We can satisfy these conditions by expressing the roots of unity in the following way. Denote the $(2d-4)$-th root of unity $\phi_{il}^{(1)}$ by η_{il}. Since $\eta_{il}^2 = \phi_{il} = (\phi_{il}^{(j)})^2$ for all $j \in \{1,\dots,d\}$, each $\phi_{il}^{(j)}$ can be written in terms of η_{il} as $\phi_{il}^{(j)} = \eta_{il}\chi_{il}^{(j)}$, where $\chi_{il}^{(j)} \in \{\pm 1\}$. We now have the condition

$$\eta_{il}^d \prod_{j=2}^{d} \chi_{il}^{(j)} = \phi_{il} = \eta_{il}^2 \quad \Longrightarrow \quad \eta_{il}^{d-2} = \prod_{j=2}^{d} \chi_{il}^{(j)}.$$

Since η_{il} is a $(2d-4)$-th root of unity η_{il}^{d-2} is in $\{\pm 1\}$. Finally, since $(e_i^{(j)} - e_l^{(j)}) + (e_l^{(j)} - e_h^{(j)}) + (e_h^{(j)} - e_i^{(j)}) = 0$, applying ρ gives $\chi_{il}^{(j)}\eta_{il}\left(\frac{\sigma_l}{\sigma_i}\right)^{\frac{1}{d-2}} \chi_{lh}^{(j)}\eta_{lh}\left(\frac{\sigma_h}{\sigma_l}\right)^{\frac{1}{d-2}} \chi_{hi}^{(j)}\eta_{hi}\left(\frac{\sigma_i}{\sigma_h}\right)^{\frac{1}{d-2}} = 1$. We now have all the relations required to find the ideals $I(\rho')$:

$$I(\rho') = \left\langle v_i^{(j)} - \chi_{i1}^{(j)}\eta_{i1}\left(\frac{\sigma_1}{\sigma_i}\right)^{\frac{1}{d-2}} v_1^{(j)} : 1 \le i \le m, 1 \le j \le d \right\rangle$$

where $\chi_{i1}^{(j)} \in \{\pm 1\}$ are such that $\chi_{i1}^{(1)} = 1$, $\prod_{j=2}^{d} \chi_{i1}^{(j)} = 1$ and the η_{i1} are $(2d-4)$-th roots of unity. Setting $\chi_i^{(j)} = \chi_{i1}^{(j)}$ and $\eta_i = \eta_{i1}$, and taking I to be the intersection of the $I(\rho')$, we obtain the required form of the singular vector tuples:

$$I = \bigcap_{\eta,\chi} \left\langle v_i^{(j)} - \chi_i^{(j)}\eta_i\left(\frac{\sigma_1}{\sigma_i}\right)^{\frac{1}{d-2}} v_1^{(j)} : 1 \le i \le m, 1 \le j \le d \right\rangle.$$

The zeros of this ideal are the singular vector tuples of Type I. □

The description of the singular vector tuples of a general orthogonally decomposable tensor is as follows.

Proposition 3.7. *Let $X = \sum_{i=1}^{n} \sigma_i v_i^{(1)} \otimes \cdots \otimes v_i^{(d)} \in \mathbb{R}^{n_1 \times \cdots \times n_d}$ be an orthogonally decomposable tensor, with $d \ge 3$ and the $v_i^{(j)} \in \mathbb{R}^{n_j}$ orthonormal vectors. Let $V^{(j)} \in \mathbb{R}^{n_j \times n_j}$ be any orthogonal matrix whose first n columns are $v_1^{(j)}, \dots, v_n^{(j)}$. Then, the singular vector tuples of X are as follows.*

Type I: Tuples $\left(V^{(1)}x^{(1)}, \dots, V^{(d)}x^{(d)}\right)$ where $(x^{(1)}, \dots, x^{d)})$ is a Type I singular vector of the diagonal orthogonally decomposable tensor in Equation (3.2).

Type II: Tuples $\left(V^{(1)}x^{(1)}, \dots, V^{(d)}x^{(d)}\right)$, where the matrix with (i,j) entry $x_i^{(j)}$ has at least two zeros in each row and no vector $x^{(j)}$ is identically zero.

Proof. Assume that $(y^{(1)}, \dots, y^{(d)})$ is a singular vector tuple of X. This means for all $j \in \{1,\dots,d\}$ the vector $[\![X; y^{(1)}, \dots, y^{(j-1)}, \cdot, y^{(j+1)}, \dots, y^{(d)}]\!]$ is parallel to $y^{(j)}$. Applying this definition to the decomposition of X, we obtain

$$[\![X; y^{(1)}, \dots, y^{(j-1)}, \cdot, y^{(j+1)}, \dots, y^{(d)}]\!] = V^{(j)}[\![S; x^{(1)}, \dots, x^{(j-1)}, \cdot, x^{(j+1)}, \dots, x^{(d)}]\!],$$

where $S = \sum_{i=1}^{n} \sigma_i e_i^{(1)} \otimes \cdots \otimes e_i^{(d)}$ is the diagonal orthogonally decomposable tensor and $y^{(j)} = V^{(j)} x^{(j)}$. Since $V^{(j)}$ is orthogonal, $[\![X; y^{(1)}, \ldots, y^{(j-1)}, \cdot, y^{(j+1)}, \ldots, y^{(d)}]\!]$ is parallel to $y^{(j)}$ if and only if $[\![S; x^{(1)}, \ldots, x^{(j-1)}, \cdot, x^{(j+1)}, \ldots, x^{(d)}]\!]$ is parallel to $x^{(j)}$. Therefore, $(x^{(1)}, \ldots, x^{(d)})$ is a singular vector tuple of S, and the solutions for all such $(x^{(1)}, \ldots, x^{(d)})$ are given in Lemma 3.5. □

We can now prove the main result, which enumerates the singular vector tuples.

Proof of Theorem 3.4. The Type I singular vectors are enumerated from the description in Proposition 3.7. We have one singular vector tuple for every choice of $m \in \{1, \ldots, n\}$, a subset of $\{1, \ldots, n\}$ of size m, scalars η_i which are $(2d-4)$-th roots of unity (where $i \in \{2, \ldots, m\}$), and $\chi_i^{(j)} \in \{\pm 1\}$ such that $\prod_{j=2}^{d} \chi_i^{(j)} = 1$ (where $i \in \{2, \ldots, m\}$ and $j \in \{2, \ldots, d\}$). Therefore, the total number of singular vector tuples of Type I is

$$\sum_{m=1}^{n} \binom{n}{m} (2d-4)^{m-1} 2^{(m-1)(d-2)} = \frac{(2^{d-1}(d-2)+1)^n - 1}{2^{d-1}(d-2)}.$$

If we impose that the singular vector tuples be real, we have only two values for the choice of each η_i rather than $(2d-4)$ which yields the real count of $\frac{(2^{d-1}+1)^n - 1}{2^{d-1}}$.

It remains to count the Type II singular vector tuples. We first study the case in which all dimensions are equal, $n_1 = \cdots = n_d = n$. Here, the tuple $(x^{(1)}, \ldots, x^{(d)})$ is a Type II singular vector tuple if and only if the matrix with (i,j) entry $x_i^{(j)}$ has at least two zeros in every row and none of the vectors $x^{(j)}$ is identically zero. This configuration is a subvariety of $\mathbb{P}^{n-1} \times \cdots \times \mathbb{P}^{n-1}$. Its ideal is given by

$$\sum_{i=1}^{n} \langle x_i^{(1)} \cdots \widehat{x_i^{(j)}} \cdots x_i^{(d)} : j = 1, \ldots, d \rangle \quad = \quad \sum_{i=1}^{n} \bigcap_{1 \le j < k \le d} \langle x_i^{(j)}, x_i^{(k)} \rangle. \tag{3.6}$$

We count the number of components in this subvariety by looking at the Chow ring of $\mathbb{P}^{n-1} \times \cdots \times \mathbb{P}^{n-1}$, which is $\mathbb{Z}[t_1, \ldots, t_d]/\langle t_1^n, \ldots, t_d^n \rangle$. Each t_j represents the class of a hyperplane in $\mathbb{P}^{n_j - 1}$, the jth projective space in the product. The equivalence class of the variety $\mathcal{V}\left(\langle x_i^{(j)}, x_i^{(k)} \rangle\right)$ is given by $t_j t_k$. We consider the variety

$$\mathcal{V}\left(\bigcap_{1 \le j < k \le d} \langle x_i^{(j)}, x_i^{(k)} \rangle \right) = \bigcup_{1 \le j < k \le d} \mathcal{V}\left(\langle x_i^{(j)}, x_i^{(k)} \rangle \right)$$

which yields our variety of interest when we intersect over i. Its equivalence class is given by $\sum_{1 \le j < k \le d} t_j t_k$. From this, we see that the equivalence class in the Chow ring of the total configuration is given by

$$p(t_1, \ldots, t_d) = \left(\sum_{1 \le j < k \le d} t_j t_k \right)^n. \tag{3.7}$$

To count the number of linear spaces in the configuration of Type II singular vector tuples, we count the number of monomials of the polynomial in Equation (3.7) as an element of the Chow ring. Equivalently we count the terms in the expansion of Equation (3.7) that are not divisible by t_j^d for any j. A monomial is produced by multiplying one of the $\binom{d}{2}$ terms in each of the n factors. This produces the first term, $\binom{d}{2}^n$, in the expression for the number of components in the base locus. We must now subtract those terms that are divisible by t_j^d for some fixed j. These are formed by selecting the terms $t_j t_{k_1}, \ldots, t_j t_{k_n}$ from consecutive factors. There are $d-1$ choices for each k_s, and d choices for the fixed j, yielding $d(d-1)^n$ terms of this format. However, we have double-counted those terms of the form $t_j^n t_k^n$ for fixed j and k, of which there are $\binom{d}{2}$. Combining these terms gives

$$\binom{d}{2}^n - d(d-1)^n + \binom{d}{2}. \tag{3.8}$$

The codimension of the ideal in Equation (3.6) is $2n$, so the linear spaces have dimension $d(n-1)-2n$.

The case of non-equal dimensions follows similarly: to count the number of maximal-dimensional linear spaces, we consider the same polynomial from Equation (3.7), and we count the terms that are not divisible by $t_j^{n_j}$ for any $j = 1, \ldots, d$. A term cannot be divisible by $t_j^{n_j}$ for any $n_j > n$. Equation (3.8) therefore generalizes to

$$\binom{d}{2}^n - c(d-1)^n + \binom{c}{2},$$

where $c = \#\{j : n_j = n\}$, and the dimension of each components is $\sum_{j=1}^d (n_j - 1) - 2n$. □

3.3 Visualizing the singular vectors

In this section I study the Type II singular vector tuples of the diagonal orthogonally decomposable tensor. The singular vector tuples of a general orthogonally decomposable tensor can be obtained by applying an orthogonal transformation.

We can identify each projective space $\mathbb{P}^{n_j-1}$ with the simplex Δ_{n_j-1} and consider the linear spaces as polyhedral complexes. They are prodsimplicial complexes, in the boundary of the product of simplices $\Delta_{n_1-1} \times \ldots \times \Delta_{n_d-1}$. The number of components in the variety of Type II singular vector tuples is the number of facets in this complex.

In the case of $2 \times 3 \times 3$ tensors, from Example 3.6, we have six Type I singular vector tuples, and the Type II singular vector tuples give five copies of $\mathbb{P}^1$. The polyhedral complex of Type II singular vector tuples, in $\Delta_1 \times \Delta_2 \times \Delta_2$, is given in Figure 3.1 (right). Motivated by this example, I investigate the shape of the Type II singular vector tuples of other orthogonally decomposable tensors.

We can stratify orthogonally decomposable tensors according to the dimension of their Type II singular vectors.

Proposition 3.8. *For each dimension k, the orthogonally decomposable tensors whose Type II singular vector tuples have dimension k have a finite list of possible formats $n_1 \times \cdots \times n_d$, where $d \geq 3$.*

Proof. By Theorem 3.4, an orthogonally decomposable tensor of format $n_1 \times \cdots \times n_d$ has Type II singular vector tuples consisting of a product of linear spaces of dimension k, where

$$\sum_{j=1}^{d}(n_j - 1) - 2n = k \tag{3.9}$$

and $n = \min(n_1, \ldots, n_d)$. Without loss of generality, we can assume that $n = n_1 \leq \ldots \leq n_d$. Let the constant α be such that $n_2 = n+\alpha$. Rearranging Equation (3.9) gives $\sum_{j=3}^{d}(n_j-1) = k + 2 - \alpha$. This has finitely many solutions, since the non-negativity of the left hand side means there will be solutions for only finitely many values of α, and each summand on the left hand side has strictly positive integer size. □

The orthogonally decomposable tensors with finitely many singular vector tuples are those with format $2 \times 2 \times 2$, $3 \times 3 \times 3$, or $2 \times 2 \times 2 \times 2$. In this case, the count agrees with the number of singular vector tuples of a generic tensor, given in [70, Theorem 1].

Tensor format	Type I count	Type II count	Generic count
$2 \times 2 \times 2$	6	0	6
$3 \times 3 \times 3$	31	6	37
$2 \times 2 \times 2 \times 2$	18	6	24

Table 3.1: Orthogonally decomposable tensors with finitely many singular vectors attain the generic count.

For larger tensor formats, an orthogonally decomposable tensor has infinitely many singular vector tuples. The tensor formats whose singular vector tuples make a one-dimensional variety are given in Table 3.2. For such tensors, each intersection of copies of $\mathbb{P}^1$ is a triple intersection. Under a small perturbation, each copy of $\mathbb{P}^1$ contributes one singular vector tuple, and two arise from each triple intersection. We observe in Table 3.2 that summing the Type I count, the number of copies of $\mathbb{P}^1$, and twice the number of triple intersections yields the generic count. The $2 \times 3 \times 3$ case is Example 3.6. In the $3 \times 3 \times 4$ and $2 \times 2 \times 2 \times 3$ cases the simplicial complexes of the Type II singular vector tuples are the same shape: 12 copies of $\mathbb{P}^1$ meeting at six triple intersections. In format $2 \times 2 \times 2 \times 2 \times 2$ we have 30 copies of $\mathbb{P}^1$ that meet at 20 triple intersection points. In format $4 \times 4 \times 4$ there are 36 copies of $\mathbb{P}^1$ meeting at 24 triple intersection. See Figure 3.2.

I conclude this chapter with an orthogonally decomposable tensor whose singular vector tuples make a two-dimensional projective variety, format $2 \times 2 \times 3 \times 3$. The number of components in the Type II configuration is 19. There are four copies of the projective

Tensor format	Type I count	#$\mathbb{P}^1$s	#Triple intersections	Generic count
$2 \times 3 \times 3$	6	5	2	15
$2 \times 2 \times 4$	6	2	0	8
$3 \times 3 \times 4$	31	12	6	55
$4 \times 4 \times 4$	156	36	24	240
$2 \times 2 \times 2 \times 3$	18	12	6	42
$2 \times 2 \times 2 \times 2 \times 2$	50	30	20	120

Table 3.2: Orthogonally decomposable tensors with a one-dimensional locus of singular vectors.

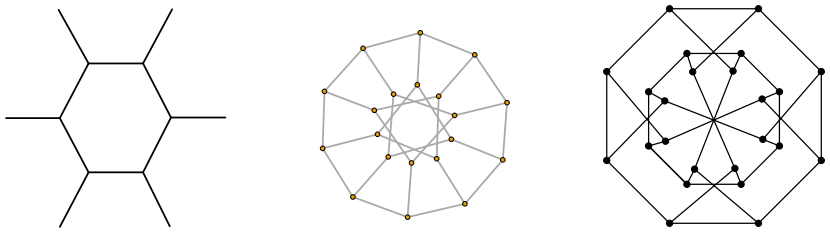


Figure 3.2: The singular vectors of an orthogonally decomposable $3 \times 3 \times 4$ or $2 \times 2 \times 2 \times 3$ tensor (left), $2 \times 2 \times 2 \times 2 \times 2$ tensor (middle), and $4 \times 4 \times 4$ tensor (right).

plane $\mathbb{P}^2$ and 15 copies of $\mathbb{P}^1 \times \mathbb{P}^1$, shown in Figure 3.3. They are arranged around a central orange square. Each edge is a copy of $\mathbb{P}^1$ and its colour represents the factor in which it occurs, in the order: red, yellow, blue, green. For example, green edges refer to copies of $\mathbb{P}^1$ of the form $\mathbb{P}^0 \times \mathbb{P}^0 \times \mathbb{P}^0 \times \mathbb{P}^1$. The configuration is not realizable in three-dimensional space: in this depiction the diagonally opposite blue and green triangles intersect. The generic count for the number of singular vector tuples of a tensor of this format is 98, by [70], and the Type I singular vector tuples contribute 18 points. Therefore, the surface accounts for 80 points, which re-appear under a general perturbation of an orthogonally decomposable tensor of this format.

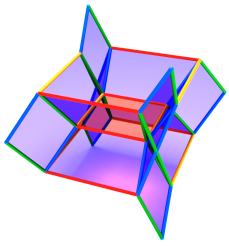


Figure 3.3: The singular vector tuples of an orthogonally decomposable $2 \times 2 \times 3 \times 3$ tensor.

Chapter 4

Singular values

The singular values of a matrix characterize it up to orthogonal changes of coordinates. The singular values quantify the importance of the principal directions of the matrix. The number of strictly positive singular values is the rank of the matrix, and the number of 'large' singular values gives the number of 'large' rank one signals. There are various ways to define a singular values of a tensor: there does not exist a single notion of singular value that generalizes all of these properties to the multi-dimensional setting. What properties should a good definition of a singular value of a tensor have? A singular value should summarize some of the structure in a tensor, and it should be unchanged under transformations of a tensor that do not affect this structure.

Recall from the previous chapter that a singular vector tuple of a tensor is a tuple of vectors $(v^{(1)}, \ldots, v^{(d)})$ such that contracting the tensor X by all choices of $(d-1)$ vectors from the tuple gives a vector parallel to the remaining vector. We can define the singular value of the singular vector tuple as the scalar σ such that

$$[\![X; v^{(1)}, \ldots, v^{(d-1)}, \cdot]\!] = \sigma v^{(d)}. \tag{4.1}$$

However, the singular value depends on the choice of scaling of the tuple, as we saw in the discussion after Definition 1.14. If we apply the scaling $(v^{(1)}, \ldots, v^{(d)}) \mapsto (\lambda v^{(1)}, \ldots, \lambda v^{(d)})$ to the condition in Equation (4.1), the left hand side is multiplied by λ^{d-1} and the right hand side is multiplied by λ. Hence the singular value changes, $\sigma \mapsto \lambda^{d-2}\sigma$ and this leaves the singular value unchanged only if $d = 2$, i.e. in the setting of matrices. The lack of invariance under scaling a direction is not a good property for a singular value to have. Each direction should have a well-defined value that measures its importance. There are two ways to get around this problem. The first is to impose a norm condition on the vectors $v^{(i)}$. Once the norm is fixed, the singular value is also fixed, but the choice of scaling may be considered unnatural in practice, especially when it comes to the comparison of singular values of multiple tensors. The alternative is to homogenize Equation (4.1) in the entries of the vectors $v^{(i)}$, by replacing the vector $v^{(d)}$ on the right hand side by a new vector whose entries are the $(d-1)$th entrywise powers of the entries of $v^{(d)}$. This makes the singular value unchanged under a re-scaling of the vector tuple, but it is no longer an orthogonal invariant,

since applying an orthogonal transformation to a general $v^{(d)}$ results in a non-orthogonal transformation of its $(d-1)$th entrywise power.

Another possible definition for the singular values of a tensor are the higher-order singular values from Definition 1.11. Recall that a higher-order singular value of a tensor is a singular value of one of its principal flattenings, the ways to reshape its $n_1 \cdots n_d$ entries into a matrix of format $n_i \times \prod_{j \neq i} n_j$. A singular value of a flattening is an orthogonal invariant of the tensor, since an orthogonal change of coordinates on the original tensor results in an orthogonal transformation of its flattenings. The higher-order singular values also have practical advantages. Each flattening is a matrix, hence the higher-order singular values can be obtained using linear algebra methods. This makes them computationally feasible to compute for large tensors, and they can be used to compress large multi-dimensional datasets [11].

However, it is difficult to ascribe meaning to the higher-order singular values. Each higher-order singular value is associated to a principal direction of a matrix, not of the original tensor. Capturing the structure of the tensor necessitates understanding how the singular values of the different flattenings relate to one another, see Problem 1.13. The singular values of a matrix can take any non-negative values. In contrast, a tensor has relations between the higher-order singular values, i.e. relations between its orthogonal invariants.

Moreover, while the higher-order singular values are *some* of the orthogonal invariants of a tensor, they are not a basis of orthogonal invariants. The following is an open problem.

Problem 4.1. *Give a basis of invariants for a tensor of format $n_1 \times \cdots \times n_d$ up to orthogonal changes of coordinates.*

The first fundamental theorem of invariant theory is, given a group and a space on which it acts, the task of finding generators for the ring of invariants. The second fundamental theorem is to give relations between the generators. Problem 4.1 is the first fundamental theorem for a product of orthogonal groups acting on a space of real tensors. For the case $d = 2$, i.e. the action of the product of orthogonal groups $O_{n_1} \times O_{n_2}$ on matrices of format $n_1 \times n_2$, the solution was described in the discussion of matrix structure via polynomials on Page 13. Further aspects of Problem 4.1 have been studied before. In [187], the author finds the invariants that are of degree at most $\min(n_1, \ldots, n_d)$. In [75], the first fundamental theorem for the action of a single orthogonal group O_n acting on the space of tensors of format $n \times \cdots \times n$ is given. In [158, 62], the authors characterize the tensors of format $n \times \cdots \times n$ that are invariant under certain subgroups of O_n. Applications of orthogonal invariants of tensors include analysis of magnetic resonance imaging (MRI) data and cryogenic electron microscopy (cryo-EM) data [66, 19].

In this chapter, I study the higher-order singular values of a tensor via polynomial orthogonal invariants, the determinants of flattenings. I focus on the case of $2 \times \cdots \times 2$ tensors. By finding relations between the determinants, I answer a question raised in [78] concerning

the tensors whose higher-order singular values take extremal values. We see that it is possible for a tensor to have extremal higher-order singular values without this property being visible from looking at a single flattening of the tensor. I also give, for a first tensor format, a solution to Problem 4.1. This chapter is based on my paper [160], published in Linear Algebra and its Applications.

There have been a number of other papers studying feasibility of higher-order singular values. The question of studying the feasibility was originally posed in [78]. In [61], the authors find necessary inequalities that hold between the top higher-order singular values of the different flattenings. Other flattenings may also be considered. In [96], the author studies relations between singular values of the flattenings of a tensor that arise in the tensor train format. There are close connections to the Horn conjecture on eigenvalues of sums of Hermitian matrices [90] and the quantum marginal problem. For more on different choices of flattenings of a tensor, see Chapter 6.

4.1 Orthogonal invariants of tensors

A tensor $X \in \mathbb{R}^{n_1 \times \cdots \times n_d}$ has many possible flattenings, or ways to reshape its entries into a matrix, as we saw in the subsection on flattenings on Page 15. For each flattening M we can form the Gram matrix MM^{T}. The coefficients of the characteristic polynomial of MM^{T} are *some* of the orthogonal invariants of X.

A *binary* tensor X is a tensor of format $2 \times \cdots \times 2$. There are d ways it can be flattened into a matrix of format $2 \times 2^{d-1}$. Each choice of flattening is given by a choice of index to label the rows. Denoting the rows by vectors v and w, the Gram matrix of a flattening is

$$\begin{bmatrix} \leftarrow & v & \rightarrow \\ \leftarrow & w & \rightarrow \end{bmatrix} \cdot \begin{bmatrix} \uparrow & \uparrow \\ v & w \\ \downarrow & \downarrow \end{bmatrix} = \begin{bmatrix} \|v\|^2 & \langle v, w \rangle \\ \langle v, w \rangle & \|w\|^2 \end{bmatrix}.$$

The characteristic polynomial of the Gram matrix has two coefficients: the trace and the determinant. The trace is $\|v\|^2 + \|w\|^2$, the squared Frobenius norm $\|X\|^2$ of the original tensor X. This invariant does not depend on the choice of flattening. The determinant $\det(MM^{\mathsf{T}})$ is given by the Cauchy-Schwarz expression $\|v\|^2\|w\|^2 - \langle v, w \rangle^2$. I call the determinant of the Gram matrix of the ith flattening the ith *Gram determinant* of X, and denote it by g_i. By the Cauchy-Binet formula, the Gram determinant g_i is the sum of squares of the 2×2 minors of the ith flattening matrix.

The Gram determinants give the higher-order singular values of a binary tensor, as follows. The higher-order singular values of X are the non-negative square roots of the eigenvalues of the n Gram matrices. Thus the higher-order singular values from the ith flattening are the non-negative solutions to the univariate polynomial in t

$$t^4 - \alpha t^2 + g_i,$$

where $\alpha = \|X\|^2$. Therefore, the map that sends a binary tensor to its higher-order singular values is obtained from the Gram determinants by

$$g_i \mapsto \left(\sqrt{\frac{\alpha + \sqrt{\alpha^2 - 4g_i}}{2}}, \sqrt{\frac{\alpha - \sqrt{\alpha^2 - 4g_i}}{2}} \right).$$

I use algebraic relations between the Gram determinants to find relations between the higher-order singular values, giving a partial solution to Problem 1.13.

The main result in this section is the following relations between the Gram determinants of a binary tensor.

Theorem 4.2. *Consider a tensor of format $2 \times \cdots \times 2$ (d times). Then each Gram determinant is bounded above by the sum of the others,*

$$g_i \leq \sum_{j \neq i} g_j, \qquad 1 \leq i \leq d. \tag{4.2}$$

I will prove Theorem 4.2 by constructing a sum of squares certificate for the difference $(\sum_{j \neq i} g_j) - g_i$. Observe that scaling a tensor by a number λ scales each of the Gram determinants by λ^4. Hence, since Equation (4.2) is homogeneous in the g_i, it suffices to prove that it holds for tensors of norm one. Note that Theorem 4.2 is also true in the case of matrices, when $d = 2$. In this case the inequalities simplify to $g_1 = g_2$, and it is well known that the two Gram determinants are equal, $\det(MM^\mathsf{T}) = \det(M^\mathsf{T}M)$.

In the discussion of flattenings of $2 \times 2 \times 2$ tensors on Page 26, I describe the Gram determinants of tensors of format $2 \times 2 \times 2$. We can show that the inequality $g_1 \leq g_2 + g_3$ holds between the three Gram determinants, by finding a sum of squares certificate for $g_2 + g_3 - g_1$. I now describe the construction of the sum of squares certificate in a way that will enable generalization to larger tensors. Each determinant g_i is given by a sum of squares expression, and I describe how to absorb the subtraction of g_1 into the expressions for g_2 and g_3.

Example 4.3 ($2 \times 2 \times 2$ Gram determinants)**.** *Define the quantity $g_2 + g_3 - g_1$ to be*

$$D^{(3)} = D_2^{(3)} + D_3^{(3)},$$

where $D_m^{(3)}$ consists of minors whose monomials differ in m indices. I find a sum of squares certificate for the two pieces $D_2^{(3)}$ and $D_3^{(3)}$ individually. All terms in $D_2^{(3)}$ that appear in g_1 also appear in either g_2 or g_3, and hence they cancel out in $D_2^{(3)}$, see Figure 1.9. Therefore $D_2^{(3)}$ is a sum of squares polynomial consisting of all squared minors in g_2 or g_3 but not in g_1:

$$D_2^{(3)} = 2(x_{000}x_{011} - x_{010}x_{001})^2 + 2(x_{100}x_{111} - x_{110}x_{101})^2.$$

The remaining terms can be expressed as a perfect square:

$$D_3^{(3)} = (x_{010}x_{101} + x_{001}x_{110} - x_{011}x_{100} - x_{000}x_{111})^2.$$

Summing these gives a sum of squares expression for $D^{(3)}$,

$$\begin{aligned} g_2 + g_3 - g_1 &= 2(x_{000}x_{011} - x_{010}x_{001})^2 + 2(x_{100}x_{111} - x_{110}x_{101})^2 \\ &+ (x_{010}x_{101} + x_{001}x_{110} - x_{011}x_{100} - x_{000}x_{111})^2. \end{aligned} \tag{4.3}$$

We can depict the information from the polynomial $D^{(3)}$ *as follows. The monomial containing the term* x_{0ij} *is determined by* i *and* j*, hence it can be labeled by* a_{ij}*. So it is represented by a vertex in Figure 4.1, and the edges connect monomials that appear in the same minor.*

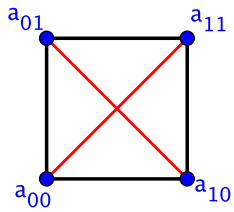


Figure 4.1: Edges represent minors unique to one flattening. The vertical and horizontal edges are from flattenings two or three, the red diagonal edges are from flattening one.

Equation (4.3) gives the proof of Theorem 4.2 in the case of $2 \times 2 \times 2$ tensors. Before proving the theorem in general, I work through the next case too, that of $2 \times 2 \times 2 \times 2$ tensors.

Example 4.4 ($2 \times 2 \times 2 \times 2$ Gram determinants)**.** *Define* $D^{(4)} := g_2 + g_3 + g_4 - g_1$*. We have*

$$D^{(4)} = D_2^{(4)} + D_3^{(4)} + D_4^{(4)}$$

where, as above, $D_m^{(4)}$ *consists of those minors whose monomials* $x_{\mathbf{i}}x_{\mathbf{j}}$ *have* $\mathbf{i}$ *and* $\mathbf{j}$ *differing in* m *indices.* $D_2^{(4)}$ *is already in sum of squares form:*

$$\begin{array}{lll} 2(x_{0000}x_{0011} - x_{0001}x_{0010})^2 + & 2(x_{0000}x_{0101} - x_{0100}x_{0001})^2 + & 2(x_{0000}x_{0110} - x_{0010}x_{0100})^2 + \\ 2(x_{1000}x_{1011} - x_{1001}x_{1010})^2 + & 2(x_{1000}x_{1101} - x_{1100}x_{1001})^2 + & 2(x_{1000}x_{1110} - x_{1010}x_{1100})^2 + \\ 2(x_{0100}x_{0111} - x_{0101}x_{0110})^2 + & 2(x_{0010}x_{0111} - x_{0110}x_{0011})^2 + & 2(x_{0001}x_{0111} - x_{0011}x_{0101})^2 + \\ 2(x_{1100}x_{1111} - x_{1101}x_{1110})^2 + & 2(x_{1010}x_{1111} - x_{1110}x_{1011})^2 + & 2(x_{1001}x_{1111} - x_{1011}x_{1101})^2. \end{array}$$

The piece $D_3^{(4)}$ *has sum of squares certificate*

$$\begin{array}{lll} (x_{0100}x_{1010} + x_{0010}x_{1100} - x_{0110}x_{1000} - x_{0000}x_{1110})^2 + & (x_{0101}x_{1011} + x_{0011}x_{1101} - x_{0111}x_{1001} - x_{0001}x_{1111})^2 + \\ (x_{0010}x_{1001} + x_{0001}x_{1010} - x_{0011}x_{1000} - x_{0000}x_{1011})^2 + & (x_{0110}x_{1101} + x_{0101}x_{1110} - x_{0111}x_{1100} - x_{0100}x_{1111})^2 + \\ (x_{0100}x_{1001} + x_{0001}x_{1100} - x_{0101}x_{1000} - x_{0000}x_{1101})^2 + & (x_{0110}x_{1011} + x_{0011}x_{1110} - x_{0111}x_{1010} - x_{0010}x_{1111})^2 + \end{array}$$

$$\begin{array}{lll} (x_{0000}x_{0111} - x_{0001}x_{0110})^2 + & (x_{0000}x_{0111} - x_{0010}x_{0101})^2 + & (x_{0000}x_{0111} - x_{0100}x_{0011})^2 + \\ (x_{1000}x_{1111} - x_{1001}x_{1110})^2 + & (x_{1000}x_{1111} - x_{1010}x_{1101})^2 + & (x_{1000}x_{1111} - x_{1100}x_{1011})^2 + \\ (x_{0001}x_{0110} - x_{0011}x_{0100})^2 + & (x_{0001}x_{0110} - x_{0101}x_{0010})^2 + & (x_{0010}x_{0101} - x_{0100}x_{0011})^2 + \\ (x_{1001}x_{1110} - x_{1011}x_{1100})^2 + & (x_{1001}x_{1110} - x_{1101}x_{1010})^2 + & (x_{1010}x_{1101} - x_{1100}x_{1011})^2. \end{array}$$

The final piece $D_4^{(4)}$ has sum of squares certificate

$$\begin{array}{lll} (x_{0010}x_{1101}+x_{0111}x_{1000}-x_{0011}x_{1100}-x_{0110}x_{1001})^2 & + & (x_{0000}x_{1111}-x_{0001}x_{1110}-x_{0100}x_{1011}+x_{0101}x_{1010})^2 \quad + \\ (x_{0000}x_{1111}+x_{0111}x_{1000}-x_{0010}x_{1101}-x_{0101}x_{1010})^2 & + & (x_{0100}x_{1011}+x_{0011}x_{1100}-x_{0001}x_{1110}-x_{0110}x_{1001})^2, \end{array}$$

which we obtain as follows. The monomials in $D_4^{(4)}$ are of the form $x_{\mathbf{i}}x_{\mathbf{j}}$ where $\mathbf{i}$ and $\mathbf{j}$ differ in all four indices. We can label the monomial containing the factor x_{0ijk} by a_{ijk}, since the other term in the monomial is fixed by the first one. In the cube a_{ijk}, we draw an edge between two vertices if those two monomials appear in a minor. The minors coming from g_1 are the red diagonal edges, and the minors from g_2 and g_3 are the black edges of the cube on the left in Figure 4.2. The polynomial represented by this picture has a sum of squares certificate. We can write it as the sum of four pieces, on the right in Figure 4.2, where a black edge is present with weight one and a red edge with weight -1. Each piece is a relabeled copy of $D_3^{(3)}$ from Figure 4.1, hence the summands are perfect squares.

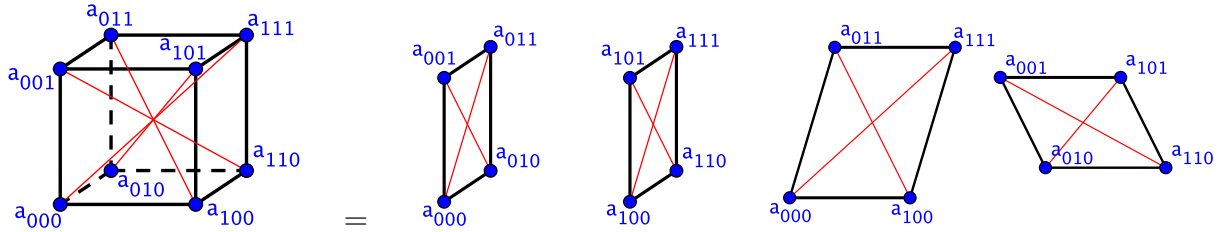


Figure 4.2: The sum of squares certificate for the difference of Gram determinants of a $2 \times 2 \times 2 \times 2$ tensor.

Proof of Theorem 4.2. We aim to show that the Gram determinants g_i of a real binary tensor satisfy the inequality

$$D^{(d)} := g_2 + g_3 + \cdots + g_d - g_1 \geq 0.$$

Each Gram determinant g_i of a tensor X is the sum of squares of the 2×2 minors of the ith flattening of X. That is, each g_i is given by a sum of squares expression. The polynomial $D^{(d)}$ is degree four in the entries of the original tensor, and we seek a sum of squares certificate for it. The set-up is symmetric in the different g_i, so this certificate can be re-labeled to give the other inequalities.

We first split up the polynomial $D^{(d)}$ into manageable pieces, and find a sum of squares certificate for each piece. The first Gram determinant can be written

$$g_1 = \sum (x_{0\mathbf{i}}x_{1\mathbf{j}} - x_{1\mathbf{i}}x_{0\mathbf{j}})^2$$

where the sum is taken over all $\mathbf{i}, \mathbf{j} \subset \{0,1\}^{d-1}$ with $\mathbf{i} \neq \mathbf{j}$. Similarly, the kth determinant is expressible in this form, where the kth index instead of the first index is swapped in each term. The polynomial $D^{(d)}$ can thus be written in terms of degree two monomials $x_{\mathbf{i}}x_{\mathbf{j}}$, where

$\mathbf{i}, \mathbf{j} \in \{0,1\}^d$, and the multi-indices $\mathbf{i}$ and $\mathbf{j}$ differ in at least 2 locations. For a monomial $x_{\mathbf{i}}x_{\mathbf{j}}$, let m count the number of locations where $\mathbf{i}$ and $\mathbf{j}$ differ (so $2 \leq m \leq d$). Let $D_m^{(d)}$ denote the terms of $D^{(d)}$ with fixed value of m. We seek a sum of squares certificate for each summand $D_m^{(d)}$.

The rest of the proof proceeds as follows. We first find the sum of squares certificate for the piece $D_2^{(d)}$. Then we show that the polynomial $D_m^{(d)}$, with $m < d$, is equal to a sum of polynomials, each equal to $D_m^{(m)}$ up to relabeling indices. Finally we relate the structure of $D_m^{(m)}$ to $D_{m-1}^{(m-1)}$, and hence we can conclude the proof by induction.

Terms in $D_2^{(d)}$ that come from g_1 are of the form $(x_{0\mathbf{i}}x_{1\mathbf{j}} - x_{1\mathbf{i}}x_{0\mathbf{j}})^2$, where $\mathbf{i}$ and $\mathbf{j}$ differ in exactly one location. Without loss of generality, we can assume they differ in their first location, and that $\mathbf{i} = (0, \ldots)$. We can therefore re-write the term as

$$(x_{00\mathbf{k}}x_{11\mathbf{k}} - x_{10\mathbf{k}}x_{01\mathbf{k}})^2, \qquad \mathbf{k} \in \{0,1\}^{d-2}.$$

We observe that this term also appears in d_2. Relabeling the above example, we see that all $D_2^{(d)}$ terms in g_1 also appear in some other g_k, and hence they do not appear in $D_2^{(d)}$. Therefore $D_2^{(d)}$ is a sum of squares polynomial: it consists of all squared minors that appear in some g_k, $2 \leq k \leq d$, but not in g_1.

Next we relate $D_m^{(d)}$ to $D_m^{(m)}$. Consider some term in $D_m^{(d)}$ coming from g_1. It is of the form

$$(x_{0\mathbf{i}}x_{1\mathbf{j}} - x_{1\mathbf{i}}x_{0\mathbf{j}})^2, \qquad \mathbf{i}, \mathbf{j} \in \{0,1\}^{d-1},$$

where $\mathbf{i}$ and $\mathbf{j}$ differ in exactly $m-1$ locations. Without loss of generality, we can assume that $\mathbf{i}$ and $\mathbf{j}$ differ in their first $m-1$ locations. Forgetting the remaining $d-m$ indices gives a projection onto $D_m^{(m)}$. Repeating for all subsets of m indices gives $\binom{d}{m}$ copies of $D_m^{(m)}$. We can obtain a sum of squares certificate for $D_m^{(d)}$ from one for $D_m^{(m)}$ by re-labeling $\binom{d}{m}$ times and summing.

The rest of the proof is by induction, with the base case $D_3^{(3)}$ from Example 4.3. For the induction step, we relate $D_m^{(m)}$, where $m \geq 3$, to $D_{m-1}^{(m-1)}$ and $D_3^{(3)}$. We saw above that the polynomial $D^{(m)}$ consists of monomials $x_{\mathbf{i}}x_{\mathbf{j}}$ with multi-indices $\mathbf{i}, \mathbf{j} \in \{0,1\}^m$. Those in $D_m^{(m)}$ have $\mathbf{i}$ different from $\mathbf{j}$ in all m locations. For example, the monomial $x_{00\ldots0}x_{11\ldots1}$ appears in $D_m^{(m)}$. For such monomials, the second variable is uniquely determined by the first.

We label the monomials in $D_m^{(m)}$ by $\{0,1\}^{m-1}$ according to the $m-1$ indices that appear after the 0 in the term that starts with a 0. These are the 2^{m-1} vertices of the following graph. We build an edge between two vertices labeled by $\mathbf{i}$ and $\mathbf{j}$ if $x_{0\mathbf{i}}$ and $x_{0\mathbf{j}}$ appear in the same term in some g_k, $1 \leq k \leq d$. Thus each edge of the graph is a summand in $D_m^{(m)}$. The edges are weighted by the coefficient with which the term appears in $D_m^{(m)}$. Those coming from g_1 have weight -1, while all others have weight $+1$. The positively-weighted edges make the $(m-1)$-dimensional cube. The negatively-weighted edges are the diagonals of this cube. For example, the summand $(x_{000\ldots0}x_{111\ldots1} - x_{100\ldots0}x_{011\ldots1})^2$ contains both $x_{000\ldots0}$ and $x_{011\ldots1}$, hence corresponds to the edge between $(0,0,\ldots,0)$ and $(1,1,\ldots,1)$.

There are 2^{m-2} diagonals in the $(m-1)$-dimensional cube. We group them into 2^{m-3} pairs, where the two diagonals in a pair differ in their first index. We extract 2^{m-3} sub-graphs by considering the edges contained in the four vertices of the two diagonals. We build part of the sum of squares certificate from each of the sub-graphs, and then a certificate from the remaining edges. Each sub-graph looks like Figure 4.3.

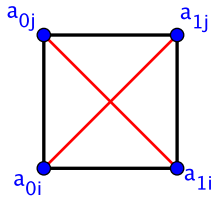


Figure 4.3: The inductive step to construct the sum of squares certificate for the difference of Gram determinants of a general binary tensor.

The vertical edges in Figure 4.3 are positively-weighted in the original graph. The red edges are negatively-weighted in the original graph. The horizontal edges were not in the original graph, but we include them in each sub-graph, at the expense of including them with negative weight among the remaining edges (this ensures they are present with overall weight zero). This graph is the 2-cube with negatively-weighted diagonals. Hence it encodes $D_3^{(3)}$, and thus the polynomial obtained from these sub-graphs has a sum of squares certificate.

It remains to consider the structure of the remaining positively and negatively-weighted edges. We have disconnected vertices according to the value of their first index. So we have two cubes of dimension $m-2$. The new negatively-weighted edges are the diagonals of these two smaller cubes. Hence we have two copies of $D_{m-1}^{(m-1)}$. By our induction hypothesis, these both have a sum of squares certificate. This concludes the proof. □

The following example shows that the above inequalities in the Gram determinants do not always hold for tensors of format $n_1 \times \cdots \times n_d$ with some $n_i > 2$.

Example 4.5. *Consider the* $2 \times 2 \times 3$ *tensor* X *with entries*

$$x_{111} = \frac{1}{\sqrt{2}}, \quad x_{213} = \frac{1}{\sqrt{2}}, \quad x_{ijk} = 0 \text{ otherwise.}$$

A computation shows that $g_1 = \frac{1}{4}$ *while* $g_2 = g_3 = 0$.

4.2 Extremal singular values

We seek tensors whose higher-order singular values are extremal, in that a small perturbation of the singular values results in a collection of numbers that are not the higher-order singular values of a tensor. These tensors lie on the boundary of the feasible set from Problem 1.13.

In [78, §1], the authors conjecture that a tensor whose singular values lie on the boundary of the feasible set must have two singular values the same in some flattening. That is, the authors conjecture that tensors with strictly decreasing and positive singular values in each flattening lie in the interior of the feasible set. The main result in this section is Example 4.11, a counter-example to this conjecture. It is obtained by giving an exact description of the feasible higher-order singular values in the case of $2 \times 2 \times 2$ tensors. The fact that the conjecture is false means that a tensor can have extremal singular values without this property being visible in a single flattening matrix.

Recall that one orthogonal invariant of a tensor is its norm. Once the norm is fixed, the feasible combinations of the Gram determinants g_i give a closed bounded semi-algebraic set. In the following definition, I focus on the set of feasible Gram determinants for tensors of norm at most one. The relations between Gram determinants of tensors of any norm can be obtained by re-scaling, since the Gram determinants are homogeneous polynomials. Scaling a tensor $X \mapsto \lambda X$ changes the Gram determinants by $g_i \mapsto \lambda^4 g_i$.

Definition 4.6 (The Gram locus)**.** *Consider real binary tensors of format $2 \times \cdots \times 2$ (d times). The map $\mathscr{G}$ sends a real binary tensor X to its tuple of Gram determinants:*

$$\begin{array}{rccc} \mathscr{G}: & \mathbb{R}^{2\times\cdots\times 2} & \to & \mathbb{R}^d \\ & X & \mapsto & (g_1, \ldots, g_d). \end{array}$$

The Gram locus *is the image $\mathscr{G}(\mathcal{B})$, where $\mathcal{B}$ is the unit ball of tensors,*

$$\mathcal{B} = \left\{ X \in \mathbb{R}^{2\times\cdots\times 2} : \|X\|^2 = \sum_{ij\ldots k} x_{ij\ldots k}^2 \leq 1 \right\}.$$

Points in the Gram locus are Gram determinants of some tensor of norm at most one. The Gram locus is not convex. A natural outer approximation is its convex hull.

Theorem 4.7. *Consider a tensor of format $2 \times \cdots \times 2$ (d times). The boundary of the convex hull of the Gram locus is described by the following linear inequalities:*

$$g_i \leq \sum_{j \neq i} g_j, \qquad 0 \leq g_i \leq \frac{1}{4}, \qquad 1 \leq i \leq d.$$

This is a convex polytope with $2^d - d$ vertices: the point $(0, \ldots, 0)$ and all points $(\frac{1}{4}, \ldots, \frac{1}{4}, 0, \ldots, 0)$ consisting of any $i \geq 2$ coordinates $\frac{1}{4}$, and the remaining coordinates zero.

Proof. Theorem 4.2 shows that the inequalities $g_i \leq \sum_{j\neq i} g_j$ are necessary. We write a flattening as

$$\begin{bmatrix} \leftarrow & v & \rightarrow \\ \leftarrow & w & \rightarrow \end{bmatrix}.$$

The trace of the Gram matrix is $\|v\|^2 + \|w\|^2$ and the determinant is $\|v\|^2\|w\|^2 - \langle v, w\rangle^2$. The Cauchy-Schwarz inequality shows that the lower bound for the determinant is 0. The

upper bound is $\frac{1}{4}$, since this is the maximum value taken by the product of two numbers that sum to one. Thus the image is contained in the cube $[0, \frac{1}{4}]^d$. This shows that the true convex hull is contained in the one in the statement of the theorem.

To conclude, I show that the vertices of the polytope in the statement of the theorem are in the Gram locus. The determinant tuple $(0, 0, \ldots, 0)$ is obtained from any rank one tensor. Consider the tensor X with entries

$$x_{00\ldots0} = \frac{1}{\sqrt{2}}, \quad x_{110\ldots0} = \frac{1}{\sqrt{2}}, \quad x_{ij\ldots k} = 0, \text{ otherwise.}$$

The first two flattenings have one non-zero entry in each of v and w, with the two vectors v and w orthogonal. Hence the determinants of the corresponding Gram matrices both evaluate to $\frac{1}{4}$. For all other flattenings, w is the zero vector and the Gram determinant is zero. Permuting indices, we see that all points with two coordinates $\frac{1}{4}$, and all others equal to zero, are in the image. Modifying the above example, so that the second non-vanishing entry is at $x_{1,1,\ldots,1,0,0,\ldots,0}$, with i indices equal to 1, shows similarly that vertices with $i > 2$ coordinates at $\frac{1}{4}$ are in the image $\mathscr{G}(\mathcal{B})$. □

The true Gram locus is a semi-algebraic subset of the convex hull. I now take steps towards its description, giving an exact formula in the case $d = 3$ and a conjecture for $d \geq 4$. The exact description of the Gram locus requires two polynomials. The first is the product of the linear conditions above:

$$Q_1 = \prod_{i=1}^{d} \left(\sum_{j \neq i} g_j - g_i \right).$$

Inside the positive orthant, the non-negativity of Q_1 is equivalent to the non-negativity of each of its linear factors. The second polynomial is given by the following product of linear factors in the $\sqrt{g_i}$:

$$Q_2 = \frac{1}{2} \times \prod_{i,j,\ldots,k \in \{\pm 1\}} (i\sqrt{g_1} + j\sqrt{g_2} + \cdots + k\sqrt{g_d}).$$

This is a product of 2^d terms, yielding a polynomial of degree 2^{d-1} in the g_i. Each term appears twice in the product, up to global sign change. Hence Q_2 is a perfect square.

Theorem 4.8. *Let $d = 3$. The Gram locus is described, inside the cube $0 \leq g_i \leq \frac{1}{4}$, by the union of the following two semi-algebraic sets:*

1. *The region $Q_1 \geq Q_2$*
2. *The region $Q_1 \leq Q_2$ and $(g_i - g_j)^2 + \frac{1}{2}(g_i + g_j) \leq \frac{3}{16}$ for all $\{i, j\} \subset \{1, 2, 3\}$.*

Conjecture 4.9. *Let $d \geq 4$. The Gram locus is given by $Q_1 \geq Q_2$ and $0 \leq g_i \leq \frac{1}{4}$ for $i = 1, \ldots, d$.*

Theorem 4.2 says the convex hull of $\mathscr{G}(\mathcal{B})$ is given, inside the cube $[0, \frac{1}{4}]^d$, by $Q_1 \geq 0$. The region $Q_1 \geq Q_2$ is contained in the convex hull, since the polynomial Q_2 is a square. The reason for the discrepancy between Conjecture 4.2 and the $d = 3$ case can be understood by evaluating $Q_1 - Q_2$ when all $g_i = \frac{1}{4}$. We obtain

$$Q_1\left(\frac{1}{4}, \ldots, \frac{1}{4}\right) = \left(\frac{d-2}{4}\right)^d, \quad Q_2\left(\frac{1}{4}, \ldots, \frac{1}{4}\right) = \frac{1}{2^{2^d+1}} \prod_{k=0}^{d} (d-2k)^{\binom{d}{k}}.$$

The value of Q_2 is 0 for all even d. Among odd d, the difference $Q_1 - Q_2$ grows in d, and is positive for $d \geq 5$. Hence the connected component of the complement of $\mathcal{V}(Q_1 - Q_2)$ containing the point $(\frac{1}{4} - \epsilon, \frac{1}{4} - \epsilon, \frac{1}{4} - \epsilon)$, for some small $\epsilon > 0$, is the same as the piece $Q_1 - Q_2 \geq 0$ for all $d \geq 4$. The condition $Q_1 \geq Q_2$ has been tested for one million randomly generated tensors with $d \in \{4, \ldots, 7\}$. Random tensors were generated with entries uniformly distribution on the interval $[0, 1]$, as well as normally distributed entries with mean zero and standard deviation one.

Proof of Theorem 4.8. We first find the Zariski closure of the boundary of the Gram locus. Following the approach in [100], this is contained in the branch locus of the map $\mathscr{G}$ and that of its restriction to the boundary $\partial\mathcal{B} = \{X \in \mathbb{R}^{2\times 2\times 2} : \|X\| = 1\}$. These branch loci are p and q respectively, obtained by direct computation (using the code on Page 74):

$$\begin{array}{c} p = g_1 g_2 g_3 (g_1 - g_2)(g_1 - g_3)(g_2 - g_3), \\ q = \prod_{i<j}(g_i - g_j) \times \prod_i \left(g_i - \frac{1}{4}\right) \times Q, \end{array}$$

where

$$\begin{array}{rcl} Q & = & \prod_{i=1}^3 \left(\sum_{j\neq i} g_j - g_i\right) - \frac{1}{2} \times \prod_{(i,j,k)\in\{\pm 1\}^3} (i\sqrt{g_1} + j\sqrt{g_2} + k\sqrt{g_3}) \\ & = & Q_1 - Q_2 \\ & = & (g_1 + g_2 - g_3)(g_1 - g_2 + g_3)(-g_1 + g_2 + g_3) - \frac{1}{2}(g_1^2 + g_2^2 + g_3^2 - 2(g_1 g_2 + g_1 g_3 + g_2 g_3))^2. \end{array}$$

The polynomial Q is the non-linear part of the boundary of the Gram locus, depicted in Figure 4.4. The Zariski closure of the boundary of $\mathscr{G}(\mathcal{B})$ is contained in $\mathcal{V}(pq)$, the vanishing locus of the polynomial pq.

The Gram locus is the closure of the union of some connected components in $\mathbb{R}^3 \backslash \mathcal{V}(pq)$: each connected component is either contained in the image, or disjoint from it. It suffices to consider components contained inside the convex hull of the Gram locus. Figure 4.4 shows that $[0, \frac{1}{4}]^3 \backslash \mathcal{V}(Q)$ has five connected components. The connected component containing $(\frac{1}{4} - \epsilon, \epsilon, \epsilon)$, for $\epsilon > 0$ sufficiently small, intersects the set $g_1 > g_2 + g_3$, hence by Theorem 4.2 it is not contained in the image. There are three such components by symmetry. The interior of the surface $\mathcal{V}(Q)$ is contained in the convex hull of the image. Likewise for the component containing the point $(\frac{1}{4} - \epsilon, \frac{1}{4} - \epsilon, \frac{1}{4} - \epsilon)$, for $\epsilon > 0$ sufficiently small. A direct computation finds tensors that map to each connected component of $[0, \frac{1}{4}]^3 \backslash \mathcal{V}(pq)$ in these two last pieces, hence they are the Gram locus. It remains to find the semi-algebraic description. The

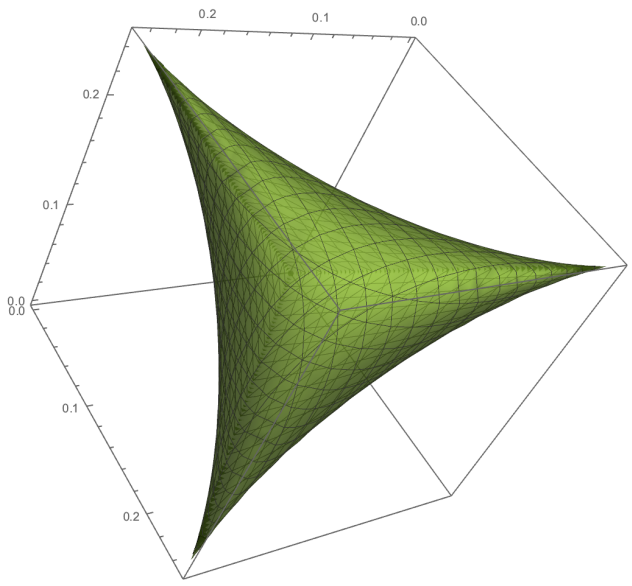


Figure 4.4: The relations between the Gram determinants of a $2 \times 2 \times 2$ tensor.

interior of the surface $\mathcal{V}(Q)$ is given by $Q \geq 0$. The surface $Q = Q_1 - Q_2$ meets the plane $g_1 = \frac{1}{4}$ along the planar curve $(g_2 - g_3)^2 + \frac{1}{2}(g_2 + g_3) - \frac{3}{16}$ with multiplicity two. Imposing that all three such polynomials, obtained by relabeling, be positive yields the component of $[0, \frac{1}{4}]^3 \backslash \mathcal{V}(Q)$ containing the point $(\frac{1}{4} - \epsilon, \frac{1}{4} - \epsilon, \frac{1}{4} - \epsilon)$. □

Polynomials p and q from the proof of Theorem 4.8 are computed in Macaulay2 [76] as follows. Computational speed-ups are obtained by changing coordinates from the x_{ijk}, the eight entries of the array, to coordinates y_{ijk} that are invariant under the orthogonal group $O_2 \times O_2 \times O_2$. The variables `gi` refer to the determinants, while `a` is the trace of any flattening. First, make two ideals using the y_{ijk} coordinates

```
C1 = minors(3,jacobian(ideal(g1,g2,g3)));
C2 = minors(4,jacobian(ideal(g1,g2,g3,a)))+ideal(1-a);
```

Saturate with respect to the known ramification locus:

```
c = ideal((g1 - g2)*(g1 - g3)*(g2 - g3));
C1 = C1:c; C2 = C2:c;
```

Project C_1 and C_2 to the ring $\mathbb{Q}[g_1, g_2, g_3]$ to obtain p and q respectively. The computation takes 5 minutes, on a computer with a CPU clock speed of 2.6GHz.

A sufficient condition for a tensor to have extremal higher-order singular values is for it to lie on the boundary of the convex hull of the Gram locus. I describe such tensors in the following result.

Corollary 4.10. *The Gram determinants of a real binary tensor satisfy $g_1 = g_2 + \cdots + g_d$ if and only if they have at most two determinants non-zero: g_1 and one other. Such tensors are given by the tensor product of a 2×2 matrix, M, with $d-2$ vectors, $v^{(j)}$, according to the formula:*

$$x_{i_1 \ldots i_d} = M_{i_1 i_j} v^{(2)}_{i_2} \cdots \widehat{v^{(j)}_{i_j}} \cdots v^{(d)}_{i_d},$$

where $\widehat{v^{(j)}_{i_j}}$ denotes the omission of the jth term from the product.

Proof. The hypothesis that $D^{(d)} = g_2 + \cdots + g_d - g_1 = 0$ means all terms in the sum of squares certificate for $D^{(d)}$ vanish. Without loss of generality, assume that the first and second determinants, g_1 and g_2, are non-zero. It suffices to show that the third determinant vanishes.

Write out the second flattening of the tensor, denoted $X^{(2)}$, arranging the columns in two blocks according to the value of the first index

$$X^{(2)} = \begin{bmatrix} \leftarrow & x_{00*} & \rightarrow & \leftarrow & x_{10*} & \rightarrow \\ \leftarrow & x_{01*} & \rightarrow & \leftarrow & x_{11*} & \rightarrow \end{bmatrix}.$$

All 2×2 minors for which the first index is constant appear as terms in the sum of squares certificate for $D^{(d)}$ (see the proof of Theorem 4.2). Therefore the left and right hand halves of $X^{(2)}$ are two rank one matrices. Say they are given by multiples of vectors x and y respectively, of length 2^{d-2}. We write

$$X^{(2)} = \begin{bmatrix} t_0 x & s_0 y \\ t_1 x & s_1 y \end{bmatrix}.$$

We now write the third flattening in terms of vectors x and y. We write $x = \begin{bmatrix} x^{(0)} & x^{(1)} \end{bmatrix}$, where the entries of x are arranged according to the value of the third index: $x^{(0)}$ are those entries of the tensor with a 0 in their third index, and $x^{(1)}$ are those with a 1 in their third index. Similarly for y. We can then write the third flattening as

$$X^{(3)} = \begin{bmatrix} t_0 x^{(0)} & t_1 x^{(0)} & s_0 y^{(0)} & s_1 y^{(0)} \\ t_0 x^{(1)} & t_1 x^{(1)} & s_0 y^{(1)} & s_1 y^{(1)} \end{bmatrix}.$$

Just as for the second flattening, we have organized the columns of the third flattening according to the value of the first index. So the matrix is formed of two rank one matrices concatenated side-by-side. This implies that there exists vectors x' and y' such that

$$X^{(3)} = \begin{bmatrix} \alpha_0 t_0 x' & \alpha_0 t_1 x' & \beta_0 s_0 y' & \beta_0 s_1 y' \\ \alpha_1 t_0 x' & \alpha_1 t_1 x' & \beta_1 s_0 y' & \beta_1 s_1 y' \end{bmatrix}. \tag{4.4}$$

The term

$$\left(x_{01\mathbf{j}} x_{10\mathbf{i}} + x_{00\mathbf{i}} x_{11\mathbf{j}} - x_{01\mathbf{i}} x_{10\mathbf{j}} - x_{00\mathbf{j}} x_{11\mathbf{i}}\right)^2$$

appears in a sum of squares certificate for $D^{(d)}$, for all $\mathbf{i}$ and $\mathbf{j}$, as follows. Let m be such that $\mathbf{i}$ and $\mathbf{j}$ differ in $m-2$ indices. Projecting to the m indices consisting of these and the first two, we obtain one of the combinations of six minors from $D_m^{(m)}$ depicted in Figure 4.3. Hence it must be zero. Substituting in our expression in Equation (4.4) for the entries of the tensor yields the equation

$$(\alpha_1\beta_2 s_1 t_2 + \alpha_2\beta_1 t_1 s_2 - \alpha_2\beta_1 s_1 t_2 - \alpha_1\beta_2 s_2 t_1) x'_{\mathbf{i}} y'_{\mathbf{j}} = 0, \quad \text{for all } \mathbf{i} \text{ and } \mathbf{j},$$

where the entry of x' corresponding to multi-index $\mathbf{i}$ is denoted $x'_{\mathbf{i}}$, and likewise for y'. Hence one of x' and y' must be zero, which contradicts $X^{(2)}$ being full rank, or $(\alpha_2\beta_1 - \alpha_1\beta_2)(s_2 t_1 - s_1 t_2) = 0$ which shows that $X^{(3)}$ is rank one, and hence $g_3 = 0$, as required. □

The true boundary of the Gram locus contains parts of all the hyperplanes $g_i = \frac{1}{4}$. If a tensor of norm one lies on the hyperplane $g_i = \frac{1}{4}$, its singular values in the ith flattening are both $\frac{1}{\sqrt{2}}$ and, in particular, the two singular values are the same. However, the following example shows that not all tensors on the boundary of the Gram locus have two singular values the same in some flattening. This disproves the conjecture from [78, §1].

Example 4.11. *Consider the tensor*

$$x_{000} = \frac{1}{\sqrt[4]{2}}, \qquad x_{101} = x_{011} = \sqrt{\frac{1}{2} - \frac{1}{2\sqrt{2}}}, \quad x_{ijk} = 0, \ \textit{otherwise}\ .$$

Its tuple of Gram determinants,

$$(g_1, g_2, g_3) = \left(\frac{1}{8}, \frac{1}{8}, \frac{\sqrt{2}-1}{2}\right),$$

lies on the part of $\mathcal{V}(Q)$ that contributes to the boundary of $\mathscr{G}(\mathcal{B})$. The higher-order singular values are: the square roots of $\frac{1+\sqrt{2}}{2\sqrt{2}}$ and $\frac{1}{2} - \frac{1}{2\sqrt{2}}$, for flattenings one and two, and the square roots of $\frac{1}{\sqrt{2}}$ and $1 - \frac{1}{\sqrt{2}}$ in the third flattening. It is labeled in Figure 4.5 by a black dot on the boundary hypersurface.

A non-linear change of coordinates converts the Gram determinants into the higher-order singular values, see Section 4.1. In [78], the authors work in the three-dimensional space of the highest singular values from each flattening. The image of the nonlinear part of the boundary of the Gram locus in these coordinates is depicted in Figure 4.5. The point of the star near $(1, 1, 1)$ is the true algebraic description for the experiments with random tensors in [78, Figure 3.1].

4.3 Orthogonal equivalence of tensors

Fixing the singular values of a matrix determines it up to orthogonal changes of coordinates. The same is not true for tensors. There exist tensors with the same higher-order singular

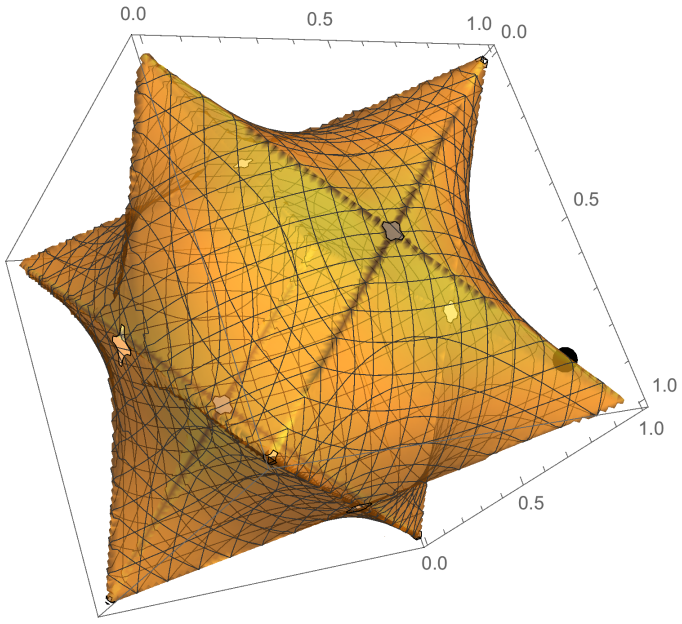


Figure 4.5: The feasible higher-order singular values of a $2 \times 2 \times 2$ tensor are the tuples inside this surface.

values that are not related by an orthogonal change of coordinates. In general, a complete list of orthogonal invariants of a tensor is not known, see Problem 4.1. In this section I give a solution to the first case of this problem.

After fixing the norm, the tensors with the same higher-order singular values are given by the fibers of the Gram determinant map, $\mathscr{G}^{-1}(g_1, \ldots, g_d) \subseteq \mathbb{R}^{2\times\cdots\times 2}$. Each fiber is defined by d non-homogeneous quartics in the space of binary tensors. It is a union of orbits under the orthogonal equivalence action for binary tensors, $O_2 \times \cdots \times O_2$ [78, Proposition 2.2]. Dimension counting reveals that the quotient

$$\mathscr{G}^{-1}(g_1, \ldots, g_d) / (O_2 \times \cdots \times O_2)$$

has dimension exponential in d: there are many tensors with the same higher-order singular values that do not differ by an orthogonal change of coordinates. To characterize tensors up to orthogonal changes of coordinates, we need to distinguish between distinct orbits inside the fiber. A computation proves the following for the case $d = 3$.

Theorem 4.12. *A $2\times2\times2$ tensor is defined up to orthogonal equivalence by its higher-order singular values and its hyperdeterminant.*

The hyperdeterminant, given in Equation (1.15), is the unique invariant up to scaling under the product of special linear groups $SL_2 \times SL_2 \times SL_2$. Hence Theorem 4.12 says that if two tensors are related by a change of basis and have the same higher-order singular values,

they are related by an *orthogonal* change of basis. That is, the fibers of the map

$$(g_1, g_2, g_3, \alpha, \text{hyperdet}) : \mathbb{R}^{2\times 2\times 2} \to \mathbb{R}^5,$$

are the equivalence classes of tensors under orthogonal change of basis, where α is the norm of a tensor. The result extends to tensors of multilinear rank $(2, 2, 2)$ by applying the subspace representation from Definition 1.10.

Consider the map that sends a tensor X of general format $n_1 \times \cdots \times n_d$ to its higher-order singular values. The tensors Y in the same fiber as X are those whose ith principal flattening is orthogonally equivalent to the ith principal flattening of X, for all $1 \leq i \leq d$. In particular, the first flattening is orthogonally equivalent to the first flattening of X, hence Y is in the orbit $(O_{n_1} \times O_{n_2 \cdots n_d}) \cdot X$. Repeating for all $1 \leq i \leq d$, we obtain that the fiber is exactly those tensors in the intersection of orbits

$$\bigcap_i (O_{n_i} \times O_{n_1 \cdots \widehat{n_i} \cdots n_d} \cdot X).$$

However, the element of the group $O_{n_i} \times O_{n_1 \cdots \widehat{n_i} \cdots n_d}$ will be different for each i. Tensors in the same fiber that differ by the same matrices in each flattening are orthogonally equivalent, since

$$\bigcap_i (O_{n_i} \times O_{n_1 \ldots \widehat{n_i} \ldots n_d}) = O_{n_1} \times \cdots \times O_{n_d}.$$

We can only express the fiber as the above intersection of d orbits, not as a single orbit. It is an open problem, see Problem 4.1 to extend Theorem 4.12 to larger tensor formats: to give the invariants describing tensors up to orthogonal equivalence.

Chapter 5

Rank vs. symmetric rank

Recall that a symmetric matrix is one that is unchanged under taking the transpose. A decomposition of a symmetric matrix into rank one terms can be chosen to be symmetric, by the eigen-decomposition. That is, the rank one terms in a decomposition can be chosen to be of the form $\lambda v \otimes v$, where v is a vector and λ is a scalar. This is useful for interpreting the rank one signals present in symmetric matrix data: each rank one term is given by a direction, the vector v, and a magnitude, the scalar λ, which measures the importance of that direction. For example, in principal component analysis the principal components of a data matrix M are the rank one matrices with largest coefficients λ in a decomposition of MM^{T}. See the discussion of the eigen-decomposition and PCA starting on Page 12. A non-symmetric decomposition of MM^{T} would have principal components of the form $\lambda u \otimes v$, with direction represented by a pair of vectors (u, v). The component would correspond to a pair of directions in the data matrix M, rather than a single direction, and this is harder to interpret. As a general principle, if a data set has some symmetry, we would like the symmetry to also be present in its decomposition.

Symmetric tensors X have entries $x_{i_1,\ldots,i_d}$ that only depend on the multi-set of indices $\{i_1, \ldots, i_d\}$, not on the order of the indices, see Definition 1.6. An order three tensor X with entries x_{ijk} is symmetric if and only if it satisfies $x_{ijk} = x_{ikj} = x_{jik} = x_{jki} = x_{kij} = x_{kji}$, see Equation (1.4). A symmetric decomposition of a symmetric tensor $X \in \mathbb{K}^{n \times \cdots \times n}$ (d times) into rank one terms has the form

$$X = \sum_{i=1}^{r} \lambda_i v_i^{\otimes d},$$

for vectors $v_i \in \mathbb{K}^n$ and scalars λ_i. As in the matrix case, each rank one term is given by a vector direction v_i and a scalar λ_i measuring the importance of that direction. The rank and symmetric rank of a tensor may not agree, by a result of Shitov [168], although it was previously conjectured that they were the same, see Conjecture 1.18. Studying the rank and symmetric rank of tensors in general is a topic of ongoing study. It is an open problem to characterize which tensors have the same rank and symmetric rank, for different notions of rank.

The study of symmetric tensors connects to questions from classical algebraic geometry. Symmetric tensors of format $n \times \cdots \times n$ (d times) are in bijection with homogeneous polynomials of degree d in n variables, as we saw in Example 1.7. A symmetric decomposition of a symmetric tensor translates to a decomposition of the corresponding polynomial into a sum of powers of linear forms, see Equation (1.12).

In this chapter, I study the smallest tensor format for which the agreement of rank and symmetric rank was not known: symmetric $4 \times 4 \times 4$ tensors, or cubic surfaces (homogeneous polynomials of degree 3 in four variables). I show that rank and symmetric rank agree for cubic surfaces. A generic matrix has the highest possible rank, but the same is not true for tensors, see Example 1.25. I introduce a test for certain high rank tensors via discriminant loci. I use it to prove that cubic surfaces with finitely many singular points are a sum of at most six cubic powers of linear forms, generalizing a classical result from [159]. These results extend to order three tensors of all formats, implying the equality of rank and symmetric rank of all tensors whose symmetric rank is at most seven.

The corresponding algebraic problem concerns border ranks. The non-equality of border rank and symmetric border rank is not yet known, see the discussion after Conjecture 1.18. I show that the non-symmetric border rank coincides with the symmetric border rank for cubic surfaces. As part of my analysis, I obtain minimal ideal generators for the symmetric analogue to the secant variety from the salmon conjecture [23, 69]. This analysis implies the equality of border rank and symmetric border rank when the symmetric border rank is at most five. This chapter is based on my preprint [161].

5.1 Ranks of cubic surfaces

A cubic surface is the zero set in projective space $\mathbb{P}^3$ of a homogeneous cubic polynomial in four variables,

$$f = c_{3000}z_1^3 + c_{2100}z_1^2z_2 + c_{1200}z_1z_2^2 + c_{0300}z_2^3 + c_{2010}z_1^2z_3 + \cdots + c_{0003}z_4^3. \tag{5.1}$$

Such a polynomial has 20 coefficients, so the space of cubic surfaces is 19-dimensional. Cubic surfaces are a central topic of study in classical algebraic geometry, and a motivating example for more modern topics in algebraic geometry too. Most prominently, the discovery of the 27 lines on the cubic surface in 1849 is celebrated as the beginning of modern algebraic geometry [159, 181]. A modern project to understand the computational algebraic properties of cubic surfaces can be found in the online collaboration [86]. Cubic surfaces correspond to symmetric $4 \times 4 \times 4$ tensors, via the correspondence

$$f(z_1, z_2, z_3, z_4) = \sum_{i,j,k=1}^{4} x_{ijk}z_iz_jz_k,$$

a special case of Equation (1.5). A symmetric tensor of format $4 \times 4 \times 4$ has entries x_{ijk} for $1 \leq i, j, k \leq 4$ that satisfy the symmetry relations $x_{ijk} = x_{ikj} = x_{jik} = x_{jki} = x_{kij} = x_{kji}$, hence there are 20 distinct entries. The main results in this chapter are the following.

Theorem 5.1. *The rank and symmetric rank agree for cubic surfaces.*

The conclusion extends to symmetric tensors of format $n \times n \times n$, by giving a range of ranks among which all tensors have agreement of rank and symmetric rank.

Corollary 5.2. *The rank and symmetric rank of a cubic polynomial in n variables (order three symmetric tensor) are the same, whenever the symmetric rank is at most seven.*

I also consider ranks over the real numbers, and show that real rank and symmetric real rank agree for generic cubic surfaces. I make the following contributions for border ranks over the complex numbers.

Theorem 5.3. *The border rank and symmetric border rank agree for cubic surfaces.*

Corollary 5.4. *The border rank and symmetric border rank of a cubic polynomial are the same whenever the symmetric border rank is at most five.*

The example of a tensor whose rank and symmetric rank disagree, the counter'example to Conjecture 1.18 given in [168], is a large tensor, both in format and rank. In contrast, the above results imply the agreement of rank and symmetric rank for small tensors and those of low rank. This suggests the following two open problems.

Problem 5.5. *(a) Find a symmetric tensor of format $n \times n \times n$, with n minimal, whose rank and symmetric rank differ.*

(b) Find a tensor of symmetric rank r, with r minimal, whose rank and symmetric rank differ.

This problem is relevant in determining whether rank and symmetric rank agree for the formats and ranks of tensors occurring in a particular application. The results in this chapter show that the n in Problem 5.5(a) satisfies $n \geq 5$ while [168] shows that $n \leq 800$, and further reductions can be made via flattening ranks. For Problem 5.5(b), the results in this chapter imply that $r \geq 8$ while [168] implies that $r \leq 906$.

In the rest of this section, I prove that the rank and symmetric rank coincide for a cubic surface. Each subsection proves the equality of rank and symmetric rank for the family of cubic surfaces in the title. Together the subsections prove Theorem 5.1.

Cones over cubic curves

The flattenings of a tensor are the reshapings of its entries into a matrix, see Page 15. A $4 \times 4 \times 4$ symmetric tensor has only one distinct flattening, a matrix of format 4×16. Cones over cubic curves have a natural characterization in terms of tensors: they are the symmetric tensors whose flattening matrix has rank ≤ 3. Such tensors parametrize the *subspace variety* [105, §7.1] defined by the vanishing of the 4×4 minors of the flattening. We change coordinates by an element M of the general linear group GL_4, with entries m_{ij}, to obtain a tensor X' with non-zero entries only in its upper-left $3 \times 3 \times 3$ block. Its entries are expressed in terms of X and M as

$$x'_{ijk} = \sum_{a,b,c=1}^{4} x_{abc} m_{ai} m_{bj} m_{ck}.$$

Rank is invariant under general linear group action, hence X' has the same rank as X. Given an expression for X' as a sum of rank one tensors, setting the fourth entry of all vectors that appear in the decomposition to zero gives a valid expression with the same number of terms.

Hence, to study ranks of cones over cubic curves it suffices to study ranks of plane cubic curves or symmetric $3 \times 3 \times 3$ tensors. It is known that the rank and symmetric rank agree for cubic curves of sub-generic flattening rank ($2 \times 2 \times 2$ tensors and rank one tensors) e.g. via their normal forms. Note that there does not exist a finite list of normal forms in the case of cubic surfaces, because the dimension of the projective general linear group PGL_4 is 15, whereas the space of cubic surfaces is 19-dimensional. For the $3 \times 3 \times 3$ case, I use the following result from [68].

Theorem 5.6 ([68, Theorem 1.1])**.** *Let $\mathbb{K}$ be a field with at least three elements. Consider a symmetric tensor $T \in (\mathbb{K}^n)^{\otimes d}$ whose rank is bounded above by its flattening rank plus one. Then the rank and symmetric rank of T defined over $\mathbb{K}$ coincide.*

It follows from Theorem 5.6 that, if the symmetric rank is bounded above by the flattening rank plus two, then the rank and symmetric rank coincide: the alternative is that the rank is strictly less than the symmetric rank, which means it satisfies the hypothesis of the theorem. Cubic curves have a generic flattening rank of three. Therefore Theorem 5.6 says that the rank and symmetric rank coincide provided the symmetric rank is at most five. The classification of cubic curves in [159, §96] shows that five is the maximum possible symmetric rank. This concludes the proof of Theorem 5.1 for cones over cubic curves.

Non-singular cubic surfaces

Based on the previous subsection, it remains to consider cubic surfaces with flattening rank four. When the rank is at most five, Theorem 5.6 implies that the rank and symmetric rank coincide. This leaves the cubic surfaces that are not expressible as a sum of five linear

powers, those for which Theorem 1.19 fails to give a decomposition. There are two such families of non-singular cubic surfaces, see [159, §94], with equations

$$\begin{array}{c}(z_1^3+z_2^3+z_3^3)+z_4^2(\lambda_1 z_1+\lambda_2 z_2+\lambda_3 x_3+\lambda_4 z_4),\\ \mu_1 z_1^3+z_2^3+z_3^3-3z_1(\mu_2 z_1 z_2+z_1 z_3+z_4^2).\end{array} \tag{5.2}$$

The parameters λ_i, μ_j are arbitrary subject to maintaining non-singularity. The failure of Sylvester's Pentahedral Theorem for these surfaces is due to the non-genericity of their Hessian quartic surface, which has fewer than 10 distinct singular points. These cubics have symmetric rank six [159, §97]. The non-symmetric rank cannot be five or less by Theorem 5.6.

Cubic surfaces with infinitely many singular points

I begin with the reducible cubic surfaces, followed by the irreducible cubic surfaces with infinitely many singular points. The three normal forms of reducible cubic surfaces are given in [41]. They are $z_1(z_1^2+z_2^2+z_3^2+z_4^2)$, $z_1(z_2^2+z_3^2+z_4^2)$, and $z_1(z_1z_2+z_3^2+z_4^2)$. The first two have symmetric rank six [41], hence by Theorem 5.6 they also have rank six. The third has symmetric rank seven [159]. I show that the rank of this normal form is seven, and hence that its rank and symmetric rank agree.

Proposition 5.7. *The cubic surface $f = z_1(z_1z_2+z_3^2+z_4^2)$ has non-symmetric rank seven.*

Proof. The polynomial f can be written up to scale as the symmetric $4\times4\times4$ tensor

$$\left[\begin{array}{cccc||cccc||cccc||cccc}
0&1&0&0&1&0&0&0&0&0&1&0&0&0&0&1\\
1&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
0&0&1&0&0&0&0&0&1&0&0&0&0&0&0&0\\
0&0&0&1&0&0&0&0&0&0&0&0&1&0&0&0
\end{array}\right].$$

Apply the substitution method, from Theorem 2.5, iteratively to the second, third, and fourth slices of f. The slices are linearly independent 4×4 matrices. No linear combination of them can be subtracted from the first slice to give a vanishing determinant. These two observations imply that the rank of f is bounded from below by $4+3=7$. Since the symmetric rank is seven, the non-symmetric rank cannot exceed seven. □

There are two normal forms of irreducible cubic surfaces with infinitely many singular points [159, §97], with representatives $z_1z_2^2+z_3z_4^2$, which has symmetric rank six, and $z_1^2z_2+z_1z_3z_4+z_3^3$ with symmetric rank at most seven. In the former case the non-symmetric rank is also six, using Theorem 5.6. In the latter case we follow an approach as in Proposition 5.7.

Proposition 5.8. *The cubic surface $f = z_1^2z_2+z_1z_3z_4+z_3^3$ has non-symmetric rank seven.*

Proof. The polynomial f is the symmetric $4 \times 4 \times 4$ tensor

$$\left[\begin{array}{cccc||cccc||cccc||cccc} 0 & \frac{1}{3} & 0 & 0 & \frac{1}{3} & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{6} & 0 & 0 & \frac{1}{6} & 0 \\ \frac{1}{3} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{6} & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & \frac{1}{6} & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{6} & 0 & 0 & 0 & 0 & 0 & \frac{1}{6} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right].$$

The second, third, and fourth slices are linearly independent. No linear combination of them can be subtracted from the first slice to give a vanishing determinant. Hence the rank is at least $4 + 3 = 7$. The symmetric rank is at most seven hence both ranks are seven. □

Cubic surfaces with finitely many singular points

I introduce a test to show that a cubic surface f has symmetric rank at most five. The test checks that f does not lie on two discriminant loci that contain the tensors of higher rank. I use it to prove Theorem 5.1 for cubic surfaces with finitely many singular points.

The singular cubic surfaces lie on the discriminant hypersurface [74, Introduction II]. Non-singular cubic surfaces, on the complement of the hypersurface, have symmetric rank at most five unless they are of the form in Equation (5.2). The surfaces in Equation (5.2) are contained in a second discriminant locus. The test is the following: if neither discriminant vanishes at f, it has symmetric rank at most five.

I now explain how to construct the second discriminant. The determinant of a 4×4 symmetric matrix of indeterminates defines a hypersurface in $\mathbb{P}^9$ with a degree 10 and codimension three locus of singular points, where the 3×3 minors of the matrix vanish. Setting the entries of the 4×4 matrix to be linear forms in four variables gives a codimension six linear space in the space of symmetric matrices of indeterminates. The determinant is now a hypersurface in $\mathbb{P}^3$ and, for a generic choice of linear forms, the singular locus consists of 10 points. The Hurwitz form of a variety is a hypersurface in the Grassmannian consisting of linear spaces that intersect the variety in a sub-generic number of points [178]. The linear forms whose determinant hypersurfaces have fewer than 10 singular points are the Hurwitz form of the variety of rank two 4×4 symmetric matrices, a hypersurface in the Grassmannian of codimension six linear spaces. Applying [178, Theorem 1.1] shows that the Hurwitz form in this setting is an irreducible hypersurface of degree 30 in the Plücker coordinates of the Grassmannian, since the sectional genus is six. The Hurwitz form has degree 120 in the coordinates of the indeterminates, since each Plücker coordinate has degree four.

The Hessian matrix of a cubic surface is a 4×4 symmetric matrix of linear forms in four indeterminates, the second order partial derivatives. The determinant of the matrix is the defining equation of the Hessian surface, which generically has 10 singular points at which the 3×3 minors of the matrix vanish. The cubic surfaces in Equation (5.2) are special in that their Hessian surfaces have fewer than 10 distinct singular points. Hence they lie on the specialization of the Hurwitz form above to Hessian matrices of cubic surfaces. This is a discriminant hypersurface in the space of cubic surfaces, which I call the *Hessian*

discriminant. It divides the specialization of the Hurwitz form. The above paragraph implies the following.

Proposition 5.9. *The Hessian discriminant is a hypersurface of degree at most 120 in the 20 coefficients of the cubic surfaces.*

We obtain a test for a cubic surface having symmetric rank at most six as follows. If there exists a linear form l such that $f + l^3$ has symmetric rank at most five, then f has symmetric rank at most six. To check that f has symmetric rank at most six, it suffices to check that neither discriminant vanishes identically on the set of cubic surfaces of the form $f + l^3$, as l ranges over the linear forms. I first prove this for the discriminant of singular cubics via the following result, which is stated without proof in [159, §97].

Lemma 5.10. *Let $f \in \mathbb{C}[z_1, z_2, z_3, z_4]$ be a cubic surface with finitely many singular points. Then for a generic linear form $l \in \mathbb{C}[z_1, z_2, z_3, z_4]$ the cubic surface $f + l^3$ is non-singular.*

Proof. A generic l satisfies $l(p) \neq 0$ at all singular points p of f, since the plane perpendicular to the coefficients of l needs to avoid finitely many points. A singular point of $g = f + l^3$ at which $l(p) \neq 0$ must satisfy $g(p) = 0$, and $\left(\frac{\partial f}{\partial z_1}|_p : \frac{\partial f}{\partial z_2}|_p : \frac{\partial f}{\partial z_3}|_p : \frac{\partial f}{\partial z_4}|_p\right) = (l_1 : l_2 : l_3 : l_4)$. The partial derivatives of f as p varies over $g = 0$ parametrize a subset of $\mathbb{P}^3$ of dimension at most two. Hence for generic l this equation will not hold at any p on the surface g. □

Remark 5.11. *Lemma 5.10 can fail for surfaces with infinitely many singular points, such as $z_1(z_1z_2 + z_3^2 + z_4^2)$ from Proposition 5.7. It is singular at $(z_1 : z_2 : z_3 : z_4) = (0 : t_1 : t_2 : \pm it_2)$ for $(t_1 : t_2) \in \mathbb{P}^1$. Every linear form l vanishes at a non-zero singular point of f and at that point $f + l^3$ is also singular.*

I now prove the following result concerning the Hessian discriminant, which uses computations in the computer algebra systems Macaulay2, Magma and Maple [76, 35, 117].

Lemma 5.12. *For all cubic surfaces with finitely many singular points, except those of singularity type E_6, there exists a linear form l such that $f + l^3$ does not lie on the Hessian discriminant.*

Proof. I refer to the classification of cubic surfaces with finitely many singular points in [39, 156, 171]. There are infinitely many normal forms, which fall into 20 classes according to the structure of the singularities. Thirteen classes have a single normal form representative. For these, we compute in Macaulay2 the ideal of singular points of the Hessian of $f + l^3$ for random linear form l. For 12 classes, all except for the singularity type named E_6 in [39], this computation gives an ideal of degree 10 and $f + l^3$ does not lie on the Hessian discriminant.

It remains to consider the seven classes from [156, Theorem 2] which are given in terms of parameters, $f = f(\rho)$. We sample linear forms l_i and compute the discriminant of $f(\rho) + l_i^3$. This gives a polynomial condition in the parameters which vanishes when $f(\rho)+l_i^3$ lies on the Hessian discriminant. We consider sufficiently many linear forms, in order that there does

not exist a choice of parameters such that the Hessian discriminant vanishes at $f(\rho)+l_i^3$ for all linear forms in the sample. We choose linear forms for which the computation to form the discriminant is not prohibitively slow. We construct the discriminant using Macaulay2 or Maple, and check that no parameters satisfy all discriminants using Macaulay2 or Magma.

Often a good choice of linear form is $l = 0$; if the Hessian discriminant does not vanish at f then it also does not vanish at $f + l^3$ when l has sufficiently small coefficients. In some cases we consider enough linear forms such that the Hessian discriminant only vanishes at all $f(\rho)+l_i^3$ for a finite number of parameters ρ, and then we check the remaining parameter values one by one in the same way as for the single normal form representatives. □

It remains to consider the singularity type E_6, with normal form $z_1^2z_4 + z_1z_3^2 + z_2^3$. Here we can see that the symmetric rank is at most six directly, since the normal form can be re-written as a sum of six linear powers,

$$\frac{1}{6}z_4^3 + \frac{1}{6}(2z_1 + z_4)^3 - \frac{1}{3}(z_1 + z_4)^3 + z_2^3 - \frac{1}{2}(z_1 + \frac{i}{\sqrt{3}}z_3)^3 - \frac{1}{2}(z_1 - \frac{i}{\sqrt{3}}z_3)^3.$$

Hence we have proved the following.

Theorem 5.13. *Cubic surfaces with finitely many singular points have symmetric rank at most six.*

When the symmetric rank is at most six, the equality of rank and symmetric rank follows from Theorem 5.6, hence this concludes the proof of Theorem 5.1. To conclude the section I prove Corollary 5.2.

Proof of Corollary 5.2. By Theorem 5.1, it remains to consider tensors of flattening rank five or more. By Theorem 5.6, the rank and symmetric rank agree when the rank is at most the flattening rank plus one. Hence they agree up to rank six, and symmetric rank seven. □

5.2 Border rank vs. symmetric border rank

The set of rank one $n \times n \times n$ tensors and the set of rank one $n \times n \times n$ symmetric tensors, up to scale, are respectively the Segre and Veronese varieties in complex projective space, see Definitions 1.20 and 1.22. In this section I denote them by

$$\mathcal{S}_n := \mathrm{Seg}(\mathbb{P}^{n-1} \times \mathbb{P}^{n-1} \times \mathbb{P}^{n-1}) \qquad \text{and} \qquad \mathcal{V}_n := \nu_3(\mathbb{P}^{n-1}).$$

Recall that the rth secant variety $\sigma_r(\mathcal{S}_n)$ consists of all tensors of non-symmetric border rank at most r. Likewise $\sigma_r(\mathcal{V}_n)$ consists of all tensors of symmetric border rank at most r. I denote the linear subspace of symmetric tensors up to scale inside $\mathbb{P}^{n^3-1}$ by $\mathcal{L}_n$.

The statement that border rank and symmetric border rank agree for cubic surfaces, in Theorem 5.3, is equivalently the statement that $\sigma_r(\mathcal{V}_4)$ and $\sigma_r(\mathcal{S}_4) \cap \mathcal{L}_4$ are equal for all r. Corollary 5.4, that the border rank and symmetric border rank agree up to symmetric border rank five, is the statement that $\sigma_r(\mathcal{V}_n) = \sigma_r(\mathcal{S}_n) \cap \mathcal{L}_n$ for all n, whenever $r \leq 4$.

Border ranks of cones over cubic curves

As for the rank result, we can apply a symmetric change of basis to such cubic surfaces to ensure that only the top-left $3 \times 3 \times 3$ block contains non-zero entries. The tensors in any approximating sequence can always be chosen to have this property, hence it suffices to consider cubic curves. The space of cubic curves is 10-dimensional. The secant varieties of the Veronese variety $\mathcal{V}_3 = \nu_3(\mathbb{P}^2)$ are not defective, by the Alexander-Hirschowitz Theorem [4]. The dimensions are

$$\dim(\mathcal{V}_3) = 2, \quad \dim(\sigma_2(\mathcal{V}_3)) = 5 \quad \dim(\sigma_3(\mathcal{V}_3)) = 8, \quad \dim(\sigma_4(\mathcal{V}_3)) = 10.$$

Since the fourth secant variety fills the space $S^3(\mathbb{C}^3)$, cubic curves have border rank ≤ 4.

Lemma 5.14. *The border rank and symmetric border rank of cubic curves coincide.*

Proof. We compare the equations defining the secant variety $\sigma_r(\mathcal{V}_3)$ with the symmetric restriction of the equations defining the non-symmetric secant $\sigma_r(\mathcal{S}_3)$, for $1 \leq r \leq 4$. The equations defining the Segre variety $\mathcal{S}_3$ are the 2×2 minors of all flattenings. Restricting these equations to symmetric tensors gives the equations defining $\mathcal{V}_3$, the 2×2 minors of the most symmetric catalecticant. Similarly $\sigma_2(\mathcal{S}_3)$ is given by the vanishing of the 3×3 minors of the flattenings. Restricting to symmetric tensors, we get the equations for $\sigma_2(\mathcal{V}_3)$, the 3×3 minors of the most symmetric catalecticant. The equations defining $\sigma_3(\mathcal{S}_3)$ are Strassen's commuting conditions. Restricting these to symmetric tensors recovers the Aronhold invariant which defines $\sigma_3(\mathcal{V}_3)$, see [105, Exercise 3.10.1.2].

Cubic curves outside $\sigma_3(\mathcal{V}_3)$ have non-symmetric border rank at least four, as they do not lie in the symmetric restriction of $\sigma_3(\mathcal{S}_3)$. Their non-symmetric border rank cannot exceed their symmetric border rank, so the non-symmetric border rank must be exactly four. □

Symmetric salmon equations

Finding ideal generators for the secant variety $\sigma_4(\mathcal{S}_4)$ is the salmon conjecture, posed by Allman in 2007. In [23, 69], set-theoretic equations for the variety are found, although ideal-theoretic equations are not known. Here we obtain the prime ideal for $\sigma_4(\mathcal{V}_4)$, a 'symmetric salmon' result.

The description for the set $\sigma_4(\mathcal{S}_4)$ consists of equations in degrees five, six and nine. There are 1728 degree five equations. Restricting the equations to symmetric tensors yields 36 linearly independent quintics that vanish on the set $\sigma_4(\mathcal{V}_4)$. Here is one of the quintics, in the coefficients of the cubic surface from Equation (5.1).

$$16c_{1002}^2c_{0201}c_{0120}^2 - 8c_{1002}^2c_{0210}c_{0120}c_{0111} - 12c_{1011}c_{1002}c_{0201}c_{0120}c_{0111} + 4c_{1011}c_{1002}c_{0210}c_{0111}^2 + c_{1011}^2c_{0201}c_{0111}^2$$
$$+4c_{1020}c_{1002}c_{0201}c_{0111}^2 + 4c_{1101}c_{1002}c_{0120}c_{0111}^2 - c_{1101}c_{1011}c_{0111}^3 - 2c_{1110}c_{1002}c_{0111}^3 + c_{2001}c_{0111}^4$$
$$+8c_{1011}c_{1002}c_{0210}c_{0120}c_{0102} + 4c_{1011}^2c_{0201}c_{0120}c_{0102} - 16c_{1101}c_{1002}c_{0120}^2c_{0102} - 4c_{1011}^2c_{0210}c_{0111}c_{0102}$$
$$-8c_{1020}c_{1002}c_{0210}c_{0111}c_{0102} - 4c_{1020}c_{1011}c_{0201}c_{0111}c_{0102} + 4c_{1101}c_{1011}c_{0120}c_{0111}c_{0102} + 8c_{1110}c_{1002}c_{0120}c_{0111}c_{0102}$$
$$+2c_{1110}c_{1011}c_{0111}^2c_{0102} - 8c_{2001}c_{0120}c_{0111}^2c_{0102} + 8c_{1020}c_{1011}c_{0210}c_{0102}^2 - 8c_{1110}c_{1011}c_{0120}c_{0102}^2 + 16c_{2001}c_{0120}^2c_{0102}$$
$$+16c_{1002}^2c_{0210}^2c_{0021} + 8c_{1011}c_{1002}c_{0210}c_{0201}c_{0021} - 4c_{1011}^2c_{0201}^2c_{0021} - 16c_{1020}c_{1002}c_{0201}^2c_{0021} + 8c_{1101}c_{1002}c_{0201}c_{0120}c_{0021}$$
$$-12c_{1101}c_{1002}c_{0210}c_{0111}c_{0021} + 4c_{1101}c_{1011}c_{0201}c_{0111}c_{0021} + 8c_{1110}c_{1002}c_{0201}c_{0111}c_{0021} + c_{1101}^2c_{0111}^2c_{0021}$$
$$+4c_{1200}c_{1002}c_{0111}^2c_{0021} - 8c_{2001}c_{0201}c_{0111}^2c_{0021} - 4c_{1101}c_{1011}c_{0210}c_{0102}c_{0021} - 16c_{1110}c_{1002}c_{0210}c_{0102}c_{0021}$$
$$+8c_{1101}c_{1020}c_{0201}c_{0102}c_{0021} + 16c_{1200}c_{1002}c_{0120}c_{0102}c_{0021} - 16c_{2001}c_{0201}c_{0120}c_{0102}c_{0021}$$
$$-4c_{1110}c_{1101}c_{0111}c_{0102}c_{0021} - 4c_{1200}c_{1011}c_{0111}c_{0102}c_{0021} + 24c_{2001}c_{0210}c_{0111}c_{0102}c_{0021}$$
$$+4c_{1110}^2c_{0102}^2c_{0021} - 16c_{1200}c_{1020}c_{0102}^2c_{0021} - 4c_{1101}^2c_{0201}c_{0021}^2 - 16c_{1200}c_{1002}c_{0201}c_{0021}^2 + 16c_{2001}c_{0201}^2c_{0021}^2$$
$$+8c_{1200}c_{1101}c_{0102}c_{0021}^2 - 16c_{1011}c_{1002}c_{0210}^2c_{0012} + 16c_{1020}c_{1002}c_{0210}c_{0201}c_{0012} + 8c_{1020}c_{1011}c_{0201}^2c_{0012}$$
$$+8c_{1101}c_{1002}c_{0210}c_{0120}c_{0012} - 4c_{1101}c_{1011}c_{0201}c_{0120}c_{0012} - 16c_{1110}c_{1002}c_{0201}c_{0120}c_{0012} + 4c_{1101}c_{1011}c_{0210}c_{0111}c_{0012}$$
$$+8c_{1110}c_{1002}c_{0210}c_{0111}c_{0012} - 4c_{1101}c_{1020}c_{0201}c_{0111}c_{0012} - 4c_{1110}c_{1011}c_{0201}c_{0111}c_{0012} - 4c_{1101}^2c_{0120}c_{0111}c_{0012}$$
$$-8c_{1200}c_{1002}c_{0120}c_{0111}c_{0012} + 24c_{2001}c_{0201}c_{0120}c_{0111}c_{0012} + 2c_{1110}c_{1101}c_{0111}^2c_{0012} - 8c_{2001}c_{0210}c_{0111}^2c_{0012}$$
$$-8c_{1101}c_{1020}c_{0210}c_{0102}c_{0012} + 8c_{1110}c_{1011}c_{0210}c_{0102}c_{0012} + 8c_{1110}c_{1101}c_{0120}c_{0102}c_{0012} - 8c_{1200}c_{1011}c_{0120}c_{0102}c_{0012}$$
$$-16c_{2001}c_{0210}c_{0120}c_{0102}c_{0012} - 4c_{1110}^2c_{0111}c_{0102}c_{0012} + 16c_{1200}c_{1020}c_{0111}c_{0102}c_{0012} + 4c_{1101}^2c_{0210}c_{0021}c_{0012}$$
$$+8c_{1200}c_{1011}c_{0201}c_{0021}c_{0012} - 16c_{2001}c_{0210}c_{0201}c_{0021}c_{0012} - 4c_{1200}c_{1101}c_{0111}c_{0021}c_{0012}$$
$$-8c_{1110}c_{1101}c_{0210}c_{0012}^2 + 16c_{2001}c_{0210}^2c_{0012}^2 + 4c_{1110}^2c_{0201}c_{0012}^2 - 16c_{1200}c_{1020}c_{0201}c_{0012}^2 + 8c_{1200}c_{1101}c_{0120}c_{0012}^2.$$

The above polynomial is one of the quintics in the following result.

Proposition 5.15. *The prime ideal of $\sigma_4(\mathcal{V}_4)$ is generated by 36 quintics.*

Proof. The 36 quintics are obtained by restricting the degree five salmon equations to symmetric tensors. Using symbolic computations in Macaulay2, they are shown to generate an ideal of degree at most 105 and codimension 4. Their ideal is Gorenstein, with symmetric minimal free resolution

$$R^1 \leftarrow R^{36} \leftarrow R^{70} \leftarrow R^{36} \leftarrow R^1 \leftarrow 0,$$

where $R = \mathbb{C}[c_{3000}, \ldots, c_{0003}]$. Using the numerical algebraic geometry methods of Bertini [24], the highest dimensional component of the variety defined by the 36 quintics is shown to be irreducible, and to have degree 105. The Gorenstein property means the unmixedness theorem [64, Corollary 18.14] applies: there cannot be lower-dimensional components. The zero set of the 36 quintics contains the codimension four set $\sigma_4(\mathcal{V}_4)$ of symmetric border rank four tensors, and hence since the codimensions agree, and the former set is irreducible, they are equal as sets. Furthermore, the ideal generated by the 36 quintics is prime, hence they generate the ideal of $\sigma_4(\mathcal{V}_4)$. □

Proposition 5.16. *The 36 quintics defining $\sigma_4(\mathcal{V}_4)$ are the irreducible module $S_{5,4,4,2}(\mathbb{C}^4)$.*

Proof. Proposition 5.15 shows that $\sigma_4(\mathcal{V}_4)$ is generated by 36 quintics. Since $\sigma_4(\mathcal{V}_4)$ is invariant under GL_4 action, the quintics are a GL_4 module in the 42504-dimensional space of quintic polynomials in the coefficients of cubic surfaces, $S^5(S^3\mathbb{C}^4)$. The GL_4 modules in $S^5(S^3\mathbb{C}^4)$ are a subset of those from $(\mathbb{C}^4)^{\otimes 15}$. The irreducible modules of the latter are indexed by Young diagrams with 15 boxes and no more than four rows [105]. We compute in SAGE [154] which GL_4-modules from $(\mathbb{C}^4)^{\otimes 15}$ occur in the decomposition of $S^5(S^3\mathbb{C}^4)$, by evaluating

```
s = SymmetricFunctions(QQ).schur(); s[5].plethysm(s[3])
```

and then selecting modules whose diagrams have at most four parts. We obtain

$$S_{5,4,4,2} \oplus S_{6,4,4,1} \oplus S_{6,5,2,2} \oplus S_{6,6,3} \oplus S_{7,4,2,2} \oplus S_{7,4,3,1} \oplus S_{7,4,4} \oplus S_{7,5,2,1}$$
$$\oplus S_{7,6,2} \oplus S_{8,3,2,2} \oplus S_{8,4,2,1} \oplus S_{8,4,3} \oplus S_{8,5,2} \oplus S_{8,6,1} \oplus S_{9,2,2,2} \oplus 2S_{9,4,2}$$
$$\oplus S_{9,6} \oplus S_{10,3,2} \oplus S_{10,4,1} \oplus S_{10,5} \oplus S_{11,2,2} \oplus S_{11,4} \oplus S_{12,3} \oplus S_{13,2} \oplus S_{15}.$$

The numbers labeling each module are the length of the rows of the Young diagram. A highest weight vector analysis shows that the quintics are the 36-dimensional module $S_{5,4,4,2}\mathbb{C}^4$. Alternatively, this is the only combination of irreducible modules of dimension 36. □

Proof of border rank results

Proposition 5.17. *If the border rank and symmetric border rank agree for $r \times r \times r$ tensors of border rank r, then they agree for $n \times n \times n$ tensors of border rank r, for all $n \geq r$.*

Proof. The containment $\sigma_r(\mathcal{V}_n) \subseteq \sigma_r(\mathcal{S}_n) \cap \mathcal{L}_n$ always holds. It remains to prove the reverse containment. We can use the technique of *inheritance* (see [105, Example 5.7.3.8 and §7.4]). Equations for $\sigma_r(\mathcal{S}_n)$ consist of $(r+1) \times (r+1)$ minors of flattenings, and copies of equations for $\sigma_r(\mathcal{S}_r)$ obtained by choosing a basis of size r in each factor $\mathbb{C}^n$. The $(r+1) \times (r+1)$ minors intersect with $\mathcal{L}_n$ to give the minors of the symmetric flattenings, while the equations for $\sigma_r(\mathcal{S}_r)$ intersect with $\mathcal{L}_n$ to give $\sigma_r(\mathcal{V}_r)$ by the hypothesis of the proposition. We can then compare with the equations for $\sigma_r(\mathcal{V}_n)$ given in [105, Corollary 7.4.2.3]. The equations are the $(r+1) \times (r+1)$ minors of the symmetric flattenings, as well as copies of equations for $\sigma_r(\mathcal{V}_r)$ given by choosing the same basis of size r in each factor $\mathbb{C}^n$. All such choices of basis are covered by the non-symmetric choices in the equations for $\sigma_r(\mathcal{S}_n)$, hence this proves the reverse containment. □

Proof of Theorem 5.3. By the Alexander-Hirschowitz theorem [4], the secant variety $\sigma_5(\mathcal{V}_4)$ fills the space of symmetric $4 \times 4 \times 4$ tensors. As in Lemma 5.14, we compare the equations defining the secant variety $\sigma_r(\mathcal{V}_4)$ with the symmetric restriction of the equations defining the non-symmetric secant $\sigma_r(\mathcal{S}_4)$, for $1 \leq r \leq 5$. The result for $r = 1, 2, 3$ follows from Lemma 5.14 combined with Proposition 5.17. When $r = 4$ the result follows from Proposition 5.15. Finally, all tensors outside of $\sigma_4(\mathcal{V}_4)$ have symmetric complex border rank five. Proposition 5.15 implies that they must also have non-symmetric complex border rank five. □

Proof of Corollary 5.4. Theorem 5.3 combined with Proposition 5.17 shows that all tensors of border rank r also have symmetric border rank r, for $1 \leq r \leq 4$. Consider a tensor of symmetric border rank five. Its border rank cannot be four by Theorem 5.3. Hence the border rank is also five. □

5.3 Real rank vs. symmetric real rank

I first prove the following, by combining results from the literature.

Proposition 5.18. *Real rank and real symmetric rank coincide for generic real cubic surfaces.*

Proof. A generic real cubic surface has complex rank five, $f = l_1^3 + l_2^3 + l_3^3 + l_4^3 + l_5^3$. The linear forms l_i define five planes in $\mathbb{P}^3$ that comprise Sylvester's pentahedron. The triple intersections of the planes are the singular points of the Hessian surface. Since f has real coefficients, so does its Hessian surface, and the singular points of the Hessian occur in complex conjugate pairs. Hence the complex linear forms appearing in the decomposition of f also occur in complex conjugate pairs. There can be zero, one, or two complex conjugate pairs in the decomposition. A cubic $l^3 + \overline{l}^3$, where l is complex and $\overline{l}$ its complex conjugate, has real symmetric rank three. Hence in the first two cases the real symmetric rank is bounded above by six. In [28], the authors show that the symmetric rank of the third case is also at most six, and therefore that a generic real cubic surface has real symmetric rank five or six. Generic cubic surfaces have flattening rank four, hence we can apply Theorem 5.6, which also holds over the field $\mathbb{R}$, to conclude that the real symmetric and non-symmetric ranks coincide up to rank five, and hence up to symmetric rank six. □

We consider special cubic surfaces in more detail, starting with cones over cubic curves.

Proposition 5.19. *Real rank and real symmetric rank coincide for cones over cubic curves.*

Proof. Such surfaces have flattening rank at most three. We apply a general linear group transformation to obtain a real symmetric tensor with non-zero entries only in its top-left $3 \times 3 \times 3$ block, and we study the cone as a cubic curve. Using Theorem 5.6, equality of real rank and real symmetric rank holds whenever the real symmetric rank is at most two more than the flattening rank. To conclude the proof, we check that all real cubic curves have this property [18, Table 1]. □

It remains to consider cubic surfaces of maximal flattening rank and, by Theorem 5.6, those of real symmetric rank at least seven. One example is given in Proposition 2.7. We saw above that every real cubic surface is arbitrarily close to one of real symmetric rank five or six. The real rank five locus is separated from the real rank six locus by a degree 40 hypersurface [122]. We are interested in the real analogue of Theorem 5.3: to show that the real border rank and real symmetric border rank agree. Generic tensors have the same real rank as real border rank, hence their real border rank and real symmetric border rank agree by Proposition 5.18. To conclude the chapter, I prove the following result.

Proposition 5.20. *Real border rank and real symmetric border rank coincide for all cubic surfaces of sub-generic real symmetric border rank.*

Proof. The set of real rank one tensors is closed, so I begin by considering a cubic surface of real border rank two. Such cubic surfaces lie in the *real rank two locus*, which is defined by the non-negativity of the hyperdeterminant of all $2 \times 2 \times 2$ blocks, by Theorem 2.2. The locus of real symmetric border rank two tensors is contained in this set, being described by the non-negativity of the diagonal (symmetric) $2 \times 2 \times 2$ blocks [164]. All diagonal combinations occur among the non-symmetric inequalities, hence the two sets are equal.

I now consider real border rank three cubic surfaces. Since the flattening rank is bounded above by the border rank, the flattening rank is at most three and we can change coordinates, as in the previous sections, to consider f as a plane cubic curve. From Theorem 5.3, it suffices to consider the orbits in [18, Table 1-2] of cubic curves whose complex (symmetric) border rank is strictly less than their real symmetric border rank. This applies to only one orbit, which has border rank two, hence it is covered by the first paragraph. Finally, assume f is a cubic surface with real symmetric border rank four. The real non-symmetric border rank cannot be strictly less than four by the above cases. □

The results in this section constitute progress towards the real rank analogues of Theorem 5.1 and Theorem 5.3. Completing the real rank version of Theorem 5.1 requires proving the equality of real rank and real symmetric rank for singular irreducible cubic surfaces, and non-singular cubic surfaces for which Theorem 1.19 fails to give a decomposition. To prove Theorem 5.3 for real border rank, it remains to consider cubic surfaces whose rank and border rank differ, having real border rank five or six.

Chapter 6

Tensor hypernetworks

We have seen the importance of low rank structure in the study of matrices and tensors. Tensor networks are a flexible framework of different notions of rank for a tensor. A tensor network is a family of tensors that factor according to the adjacency structure of the graph, as I describe shortly. Tensor networks are widely used, in contexts ranging from numerical analysis [77, 138] to theoretical physics [29, 104, 137] to function approximation [12, 14].

In this chapter, I describe tensor hypernetworks: the set of tensors which can be factored according to the adjacency structure of a hypergraph. This is a broader framework than tensor networks, encompassing both tensor networks as well as the usual tensor rank. Moreover, it has close connections to statistics. I describe how tensor hypernetworks are dual to graphical models, multivariate statistical models based on graphs. In this chapter I usually denote a tensor by T to avoid confusion with random variables, which are denoted by the letter X. This chapter is based on joint work with Elina Robeva, published in Information and Inference: A Journal of the IMA [152].

I begin by considering a familiar example of a tensor network: the set of low rank matrices.

Example 6.1 (Matrix rank)**.** *Recall that a matrix M of format $n_1 \times n_2$ is rank r if and only if it can be written as the product of an $n_1 \times r$ matrix A and an $n_2 \times r$ matrix B, via*

$$M = AB^{\mathsf{T}} \qquad \textit{or, in coordinates,} \qquad m_{ij} = \sum_{k=1}^{r} a_{ik} b_{kj}. \tag{6.1}$$

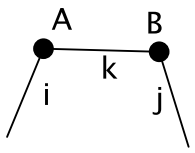


Figure 6.1: Matrix rank as a tensor network.

Rank r matrices factor according to the weighted graph in Figure 6.1. The vertices of the graph are labeled by the matrices A and B. The graph has two kinds of edges.

- *Dangling edges: The edges labeled i and j are only connected to one vertex, and are not summed over in the decomposition of M.*
- *Full edges: The edge between A and B is labeled by the an index k that is summed over in Equation* (6.1)*. This edge has weight r, because the index k is summed from* 1 *to r.*

As the weight r is increased, more matrices factor according to this tensor network. When $r = \min(n_1, n_2)$, all matrices can be decomposed as in Equation (6.1)*.*

A general tensor network is formed by taking building blocks of the form in Example 6.1. For a general graph, each *dangling edge* is labeled by an index of the original tensor, and each *full edge* is labeled by an index that is summed over in the decomposition. The edges are given weights that tell us how many values the index can take. The weights of the dangling edges give the format of tensors that the tensor network represents. The weights of the full edges give the tensor network ranks [190], which quantify the restrictiveness of the tensor network: the higher the weights, the more tensors can be represented by the tensor network.

A widely-studied tensor network arises from the Tucker decomposition. In the Tucker decomposition, we write a tensor as the product of a smaller core tensor C with a tuple of matrices $A^{(i)}$, via

$$X = [\![C; A^{(1)}, \ldots, A^{(d)}]\!], \qquad \text{where} \quad C \in \mathbb{K}^{m_1 \times \cdots \times m_d} \quad \text{and} \quad A^{(i)} \in \mathbb{K}^{n_1 \times m_i}. \tag{6.2}$$

See Equation (1.10) for the coordinate description. The singular values arising from such a decomposition were the focus of Chapter 4. Fixing the formats of the tensors involved gives us a tensor network. For example, we can consider the set of tensors X of format $n_1 \times n_2 \times n_3$ that have a factorization as in Equation (6.2), for some core tensor C of format $m_1 \times m_2 \times m_3$. The matrices $A^{(i)}$ that relate C to X have formats $n_i \times m_i$. The set of tensors which have such a decomposition parametrize the tensor network in Figure 1.7.

If we decompose a general tensor with respect to a tensor network, the edges will have large weights, and there will be poor reduction in the complexity of the set-up. The idea behind tensor networks is that the tensors of interest in an application can be accurately approximated by a well-chosen tensor network with fairly low edge weights. For example, in a Tucker approximation the format of the core tensor $m_1 \times m_2 \times m_3$ is a tradeoff between accuracy and conciseness. For the tensor network to exactly represent a tensor X, the m_i would need to be at least as large as the ranks of the principal flattenings of the tensor. For an approximate decomposition, the ranks m_i need to be at least the number of 'large' singular values in the flattenings.

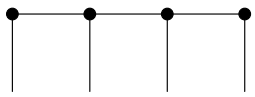
If a flattening of a tensor has low rank, it indicates that the indices in the rows are approximately uncoupled from the indices in the columns. A tensor network uses this information, across various flattenings, to decompose the tensor. The choice of graph is important, the

idea is to use a graph whose adjacency structure reflects the application at hand. In [190], the authors study how tensors with a low-complexity description with respect to one graph, can have high weights with respect to a different graph.

The computational properties of a tensor network are important for decomposing or approximating a tensor stably and accurately by a tensor network. In [77], the author describes the numerical stability properties of hierarchical tensor networks, those whose underlying graph is a tree. In [102], the authors show that the set of tensors in a tensor network may not be closed if the underlying graph contains a cycle.

The following example highlights a central motivating application of tensor networks.

Example 6.2 (Matrix Product States/Tensor Trains)**.** *The quantum state of a d particle system is given by a tensor of order d. For a system of qubits, we obtain a binary tensor, of format $2 \times \cdots \times 2$. More generally, we have d particles which can take two or more possible states. wave functionThe tensor entry at the index $(i_1, \ldots, i_d)$ is the value of the wave function at that tuple of states. A long-standing intuition in physics is that tensors arising from many body quantum systems have a special structure: particles further away are less entangled than the neighbouring particles. This means the wave function can be accurately approximated by a tensor network with underlying graph given by the following picture, for the example of four particles on a one-dimensional lattice.*



In the study of quantum many-body systems, such tensor networks are called matrix product states, and in numerical analysis they are called tensor trains [138].

The notion that quantum states are well-approximated by matrix product states can be formalized as an area law for one-dimensional quantum systems, first proved in [81] and strengthened in [10]. Such an approximation is important because it allows efficient computations with the wave function. I return to matrix product states in Section 6.2.

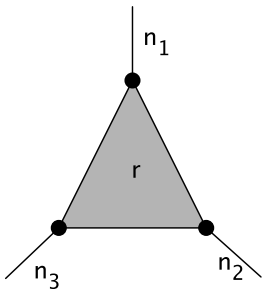
In Example 6.1, we saw that matrix rank gives a tensor network. It is natural to wonder whether tensor rank can also be thought of as a tensor network. This would allow a unified treatment of flattening ranks, tensor network ranks, and usual ranks, when deciding the best way to represent a tensor of interest. Recall that an order three tensor X of format $n_1 \times n_2 \times n_3$ has rank r if it can be written

$$X = \sum_{l=1}^{r} a_l \otimes b_l \otimes c_l, \qquad \text{or, in coordinates,} \qquad x_{ijk} = \sum_{l=1}^{r} a_{li} b_{lj} c_{lk},$$

where a_{li} is the ith entry of the vector a_l. We arrange the vectors a_l to form a matrix A of format $n_1 \times r$, and likewise we form the matrices B and C. The three indices (i, j, k) are not summed over, and the index l is summed over. However, the index l appears three times in the decomposition, so it cannot be thought of as a contraction of an index along

an edge. Instead, it is a contraction along a hyperedge connecting the vertices A, B and C. The set of tensors of rank at most r is parametrized by a *tensor hypernetwork*, a tensor network based on hyperedges. This representation does not require special structure (such as diagonal structure) on the matrices A, B, C appearing at the vertices, unlike ways to represent tensor rank via a usual tensor network.

Example 6.3 (Tensor rank as a tensor hypernetwork)**.** *Consider the following hypergraph.*



There is one dangling edge for each vertex. There is one more hyperedge, of weight r (represented by a shaded triangle) that connects the three vertices. The arrays at the vertices are matrices of formats $n_1 \times r$, $n_2 \times r$, and $n_3 \times r$.

The set of tensors which factor according to this tensor are exactly the tensors of format $n_1 \times n_2 \times n_3$ and rank $\leq r$. Extending this to d vertices, with a hyperedge of weight r, gives tensors of format $n_1 \times \cdots \times n_d$, and rank at most r.

I now give the general definition of a tensor hypernetwork, by first recalling the definition of a hypergraph.

Definition 6.4. *A hypergraph $G = (V, E)$ consists of a set of vertices V, and a set of hyperedges E. A hyperedge $e \in E$ is any subset of the vertices.*

Note that a graph is a hypergraph in which all hyperedges are subsets of size two. A hypergraph can be constructed from a matrix M of format $|V| \times |E|$ with entries in $\{0, 1\}$. Let the rows index the vertices and the columns index the hyperedges. The non-vanishing entries in each column give the vertices that appear in a hyperedge,

$$m_{ve} = \begin{cases} 1 & v \in e \\ 0 & \text{otherwise.} \end{cases} \tag{6.3}$$

The matrix M is the *incidence matrix* of the hypergraph. We allow nested or repeated hyperedges, as well as edges containing one or no vertices, so there are no restrictions on M. Alternatively, we can construct the hypergraph with incidence matrix M^{T}. This is the *dual hypergraph* to the one with incidence matrix M, see [27, Section 1.1].

We now add extra data to the matrix M. We attach positive integers $n_1, \ldots, n_d$ to each row. We assign tensors to each column of M whose format is the product of the n_i as i

ranges over the non-vanishing entries in the column. For example, the tensor associated to the column $(1, 1, 0, 1, 0, \ldots, 0)^{\mathsf{T}}$ would have format $n_1 \times n_2 \times n_4$. Filling in the entries of the tensors gives a tensor network state in a tensor hypernetwork, as I now describe.

Definition 6.5. *Consider a hypergraph $G = (V, E)$. To each hyperedge $e \in E$ we associate a positive integer n_e, called the size of the hyperedge. To each vertex $v \in V$ we assign a tensor $T_v \in \bigotimes_{e \ni v} \mathbb{K}^{n_e}$, where $\mathbb{K}$ is usually $\mathbb{R}$ or $\mathbb{C}$. The* tensor hypernetwork state *is obtained from $\bigotimes_{v \in V} T_v$ by contracting (summing over) the indices of all hyperedges in the graph that contain two or more vertices. We call hyperedges containing only one vertex dangling edges.*

The data of a tensor hypernetwork (up to global scaling constant) is its hypergraph along with the tensor at each vertex of the hypergraph. This is the tensor network before contracting the hyperedges. The distinction between the contracted and un-contracted tensor network generalizes the fact that a rank r matrix M of format $n_1 \times n_2$ can be represented either by its entries, or by two matrices A and B of formats $n_1 \times r$ and $n_2 \times r$ whose product is M, see Example 6.1.

Tensor hypernetworks were also introduced in the papers [16, 20]. Restricting the definition of a tensor hypernetwork to hyperedges with at most two vertices gives the usual definition of a tensor network. Tensor networks are sometimes assumed to have exactly one dangling edge per vertex, but I will not make that assumption here.

6.1 Duality to graphical models

The joint probability distribution of several finite random variables $X_1, \ldots, X_d$ can be naturally organized into a tensor, whose entry at the index $(i_1, \ldots, i_d)$ is the probability

$$P(X_1 = i_1, \ldots, X_d = i_d).$$

Statistical models are families of probability distributions that share some structure. Graphical models consist of distributions that factor according to the adjacency structure of a graph, see [30, 107]. Hence both graphical models and tensor networks are ways to represent families of tensors that factorize according to a graph structure.

The relationship between particular graphical models and tensor networks has been studied in the past. In [51], the authors reparametrize a hidden markov model to make a matrix product state tensor network. In [46], a map is constructed that sends a restricted Boltzmann machine graphical model to a matrix product state. In [140], an example of a directed graphical model is given with a related tensor network on the same graph, to highlight computational advantages of the graphical model in that setting. In [99, 111], the incidence matrix M from Equation (6.3) is considered as the biadjacency matrix of a bipartite graph, to obtain a factor graph construction.

The main results in this section is a duality between tensor hypernetworks and undirected graphical models on finite random variables. This close connection is a reminder of the

compatibility of graphical models and tensor networks as practical multi-dimensional tools, and as research areas. Before stating the duality correspondence, I define graphical models in terms of hypergraphs.

Definition 6.6. *Consider a hypergraph $H = (U, \mathscr{C})$. An* undirected graphical model *with respect to H is the set of probability distributions on the random variables $\{X_u, u \in U\}$ which factor according to the hyperedges in $\mathscr{C}$:*

$$P(x_1, \ldots, x_d) = \frac{1}{Z} \prod_{C \in \mathscr{C}} \psi_C(x_C).$$

Here, the random variable X_u takes values $x_u \in \mathfrak{X}_u$, the subset x_C equals $\{x_u : u \in C\}$, and the function ψ_C is a clique potential *with domain $\prod_{u \in C} \mathfrak{X}_u$ and range $\mathbb{R}_{\geq 0}$. The normalizing constant Z ensures that the probabilities sum to one.*

If we fix the values in the clique potentials, we obtain a particular distribution in the graphical model. We recover the description of the graphical model by a graph instead of a hypergraph by connecting pairs of vertices by an edge if they lie in the same hyperedge. When all random variables have finitely many states, the joint probabilities form a tensor of format $\times_{u \in U} |\mathfrak{X}_u|$ and the clique potentials are tensors of format $\times_{u \in C} |\mathfrak{X}_u|$ with all entries in $\mathbb{R}_{\geq 0}$. Graphical models are sometimes required to factorize according to the *maximal* cliques of a graph. We see later how our set-up specializes to this case. Models with cliques that are not necessarily maximal can be called *hierarchical models* [179].

The tensor hypernetwork on a hypergraph exactly corresponds to the graphical model given by the dual hypergraph.

Theorem 6.7. *A distribution in a discrete graphical model associated to a hypergraph $H = (U, \mathscr{C})$ with clique potentials $\psi_C : \prod_{u \in C} \mathfrak{X}_u \to \mathbb{K}$ is the same as the data of a tensor hypernetwork associated to its dual hypergraph H^* with tensors $T_C = \psi_C$ at each vertex of H^*.*

Proof. Consider a joint distribution (or tensor) P in the graphical model defined by the hypergraph H. As described above, the incidence matrix M of H has rows corresponding to the variables $u \in U$ and columns corresponding to the cliques $C \in \mathscr{C}$. The data of the distribution P also contains a potential function $\psi_C : \prod_{u \in C} \mathfrak{X}_u \to \mathbb{K}$ for each clique $C \in \mathscr{C}$, which is equivalently a tensor of format $\times_{u \in C} |\mathfrak{X}_u|$.

The dual hypergraph H^* has incidence matrix $M^\intercal$. It is a hypergraph with vertices $\{v_C : C \in \mathscr{C}\}$ and hyperedges $\{e_u : u \in U\}$. By definition of the dual hypergraph, $u \in C$ is equivalent to $v_C \in e_u$. Associating the tensors $T_C = \psi_C \in \bigotimes_{e_u \ni v_C} \mathbb{K}^{|\mathfrak{X}_u|}$ to each vertex v_C of H^* gives a tensor hypernetwork for H^*. Moreover, up to scaling by the normalization constant Z, the joint probability tensor P is given by

$$P(x_u : u \in U) \cdot Z = \prod_{C \in \mathscr{C}} \psi_C(x_C) = \prod_{C \in \mathscr{C}} (T_C)_{x_C}.$$

The last expression is the tensor hypernetwork state before contracting the hyperedges. □

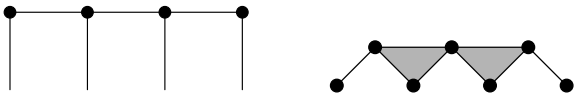
Denote the set of distributions on $\mathfrak{X} = \prod_{u \in U} \mathfrak{X}_u$ that are in the graphical model defined by the hypergraph $H = (U, F)$ by $\mathscr{G}(H, \mathfrak{X})$. These are the distributions which factor according to the hypergraph H, whose factor have formats determined by $\{|\mathfrak{X}_u| : u \in U\}$, and for which the entries of the factors can vary. Denote the set of non-contracted tensor hypernetwork states from a hypergraph $G = (V, E)$ with weights $\mathbf{n} = \{n_e : e \in E\}$ by $\mathscr{T}(G, \mathbf{n})$. Since $(M^{\mathsf{T}})^{\mathsf{T}} = M$, we obtain the following one-to-one correspondence.

Corollary 6.8. *There is a one-to-one correspondence between the graphical models $\mathscr{G}(H, \mathfrak{X})$ and the tensor hypernetwork states $\mathscr{T}(H^*, \{|\mathfrak{X}_u| : u \in U\})$ up to global scaling constant.*

In Corollary 6.8, we can impose that the tensors in both the graphical model and the tensor hypernetwork states have non-negative entries, to be in the setting of probabilities. We can also consider the tensors on both sides to have entries in a general field $\mathbb{K}$, since the definition and factorization of graphical models carries over to this case.

In the rest of this section I illustrate the duality results by showing the duals to some familiar examples of tensor network states and graphical models. I consider those tensor networks, or graphical models, which factor according to the hypergraphs shown, without fixing the formats of the factors or the entries of the tensors.

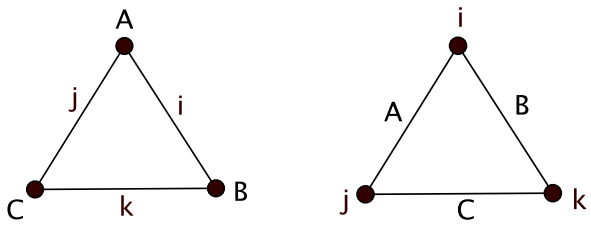
Example 6.9 (Matrix Product States/Tensor Trains)**.** *See Example 6.2. The MPS network on the left is dual to the graphical model on the right. The top row of edges in the tensor network is contracted. We see later that this corresponds to the top row of variables in the graphical model being hidden.*



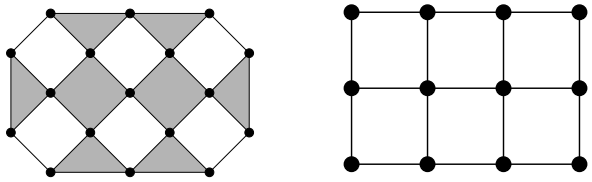
Example 6.10 (No three-way interaction model)**.** *This graphical model consists of all probability distributions that factor as $p_{ijk} = a_{ij}b_{ik}c_{jk}$, for clique potential matrices A, B, C. It is represented by a hypergraph in which all hyperedges have two vertices. The incidence matrix of the hypergraph is*

$$\begin{array}{c} \\ i \\ j \\ k \end{array} \begin{array}{c} \begin{array}{ccc} A & B & C \end{array} \\ \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix} \end{array}$$

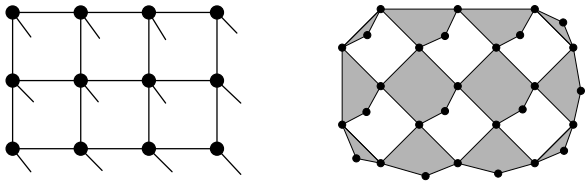
This matrix is symmetric. Hence the tensor network corresponding to this graphical model is given by the same triangle graph. We note that, up to dangling edges, this is also the shape of the tensor network that represents the matrix multiplication operator [104].



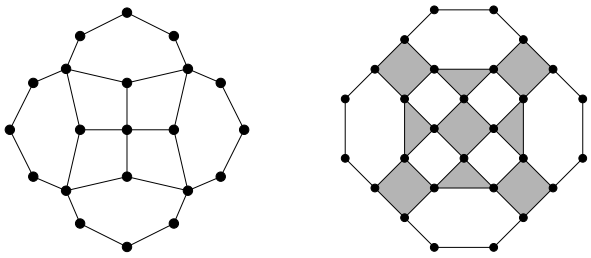
Example 6.11 (The Ising Model)**.** *This graphical model is defined by cliques which are the edges of a two-dimensional lattice such as the grid on the right. It is dual to the hypergraph on the left.*



Example 6.12 (Projected Entangled Pair States)**.** *This tensor network is a two-dimensional analogue of a Matrix Product State. It depicts two-dimensional quantum spin systems. Its hypergraph is depicted on the left, with its dual graphical model on the right. Note the similarity with Example 6.11.*



Example 6.13 (The Multi-scale Entanglement Renormalization Ansatz (MERA))**.** *This tensor network is popular in the quantum community, due to its favorable abilities to represent relevant tensors and compute efficiently with them. It is on the left, with its dual graphical model on the right.*



Properties and operations for graphical models and tensor hypernetworks relate under the duality map.

Restricting to graphs

Graphs are hypergraphs in which every hyperedge contains two vertices, also known as 2-uniform hypergraphs. Each column of the incidence matrix of such a hypergraph sums to two. The dual of a graph is a hypergraph in which every vertex has degree two, also known as a 2-regular hypergraph [27]. We call a hypergraph at-most-2-regular if every vertex has degree at most 2.

Proposition 6.14. *Tensor networks are dual to at-most-2-regular graphical models. Graph models (graphical models whose cliques are the edges of a graph) are dual to 2-regular tensor hypernetworks.*

Graphical models defined by the maximal cliques of a graph correspond to hypergraphs in which we introduce a hyperedge for each maximal clique. Their dual tensor hypernetworks have the following property.

Proposition 6.15. *Graphical models defined by the maximal cliques of a graph correspond to tensor hypernetworks whose hypergraphs have the property that whenever a set of hyperedges meet pairwise, the intersection of all of them is non-empty.*

Proof. Let $E' \subseteq E$ be a set of hyperedges of the hypergraph of the tensor hypernetwork that meet pairwise. Then, for all $e_1, e_2 \in E'$, the corresponding vertices u_{e_1}, u_{e_2} in the dual hypergraph (i.e. in the graphical model) are connected by an edge. Thus, the vertices $\{u_e : e \in E'\}$ form a clique in the graphical model, so there exists a maximal clique C in which this clique is contained. Thus, all hyperedges in E' contain the vertex corresponding to C. □

Trees

In this subsection I show that trees are preserved under the duality correspondence. The *homotopy type* of a hypergraph is the homotopy type of the simplicial complex whose maximal simplices are the maximal hyperedges. For topological purposes, we associate hypergraphs with their simplicial complexes. We see that the homotopy type of a hypergraph and its dual agree.

Definition 6.16 (see [82])**.** *Consider an open cover $\mathcal{V} = \{V_i : i \in I\}$ of a topological space. The* nerve *$N(\mathcal{V})$ of the cover is a simplicial complex with one vertex for each open set. A subset $\{V_j : j \in J\}$ spans a simplex in the nerve whenever $\cap_{j\in J} V_j \neq \emptyset$.*

Theorem 6.17 (The Nerve Lemma [34])**.** *The homotopy type of a space equals the homotopy type of the nerve of an open cover of the space, provided that all intersections $\cap_{j\in J} V_j$ of sets in the open cover are contractible.*

I consider the open cover of the simplicial complex in which open sets are ϵ-neighbourhoods of the maximal simplices. For ϵ sufficiently small, such an open cover has contractible intersections, since they are homotopy equivalent to intersections of simplices. Hence the homotopy type of the hypergraph is equal to that of its nerve. The following proposition relates the nerve to the dual hypergraph.

Proposition 6.18. *The nerve of the above open cover of the simplicial complex of a hypergraph is the simplicial complex of its dual hypergraph.*

Proof. Consider a hypergraph H with vertex set U and hyperedge set $\mathscr{C}$. We now construct the dual hypergraph. The edges are represented by rows in the original incidence matrix. A subset $\{C_j : j \in J\} \subseteq \mathscr{C}$ is contained in a hyperedge if there exists a vertex $u \in U$ that is in all hyperedges C_j in the subset. Hence the simplices that arise in the dual simplicial complex are given by subsets of hyperedges for which the intersection $\cap_{j\in J} C_j$ is non-empty. This is exactly the definition of the nerve. □

From this, the Nerve Lemma implies the following.

Theorem 6.19. *The hypergraph of a tensor hypernetwork and the hypergraph of its dual graphical model have the same homotopy type.*

A hypergraph cycle (see [27, Chapter 5]) is a sequence $(x_1, e_1, x_2, \ldots, x_k, e_k, x_1)$, where the e_i are distinct hyperedges and the x_j are distinct vertices, such that $\{x_i, x_{i+1}\} \subseteq e_i$ for all $i = 1, \ldots, k-1$, and $\{x_1, x_k\} \subseteq e_k$. A tree is a hypergraph with no cycles. Theorem 6.19 implies that trees are preserved under the duality correspondence.

Marginalization and contraction

The interpretations of marginalization and contraction are similar. The variables of a graphical model that are marginalized are often considered to be hidden, and the contracted edges of a tensor network represent entanglement, or unseen interaction. In the following proposition I give the mathematical relation between these two operations.

Let $H = (U, \mathscr{C})$ be a hypergraph and H^* its dual. Let P be a distribution in the graphical model on H with clique potentials $\psi_C : \prod_{u\in C}[n_u] \to \mathbb{K}$. The dual tensor hypernetwork has tensors $T_C = \psi_C \in \bigotimes_{u\in C} \mathbb{K}^{n_u}$ at the vertices of H^*.

Proposition 6.20 (Marginalization Equals Contraction)**.** *Let $W \subseteq U$ be a subset of the vertices of the graph H. Then, the marginal distribution of $\{X_u\}_{u\in W}$ equals*

$$P(x_W) = \sum_{\substack{x_u \in [n_u]: \\ u \notin W}} \prod_{C\in\mathscr{C}} (T_C)_{\{x_C : u\in C\}},$$

which is the contracted tensor hypernetwork along the hyperedges corresponding to W^c.

Proof. The proof follows from the chain of equalities:

$$P(x_W) = \sum_{\substack{x_u \in [n_u]: \\ u \notin W}} P(x) = \sum_{\substack{x_u \in [n_u]: \\ u \notin W}} \prod_{C \in \mathscr{C}} \psi_C(x_C) = \sum_{\substack{x_u \in [n_u]: \\ u \notin W}} \prod_{C \in \mathscr{C}} (T_C)_{\{x_C : u \in C\}}.$$

In words, summing over the values of all variables in W^c is the same as contracting the tensor hypernetwork along all hyperedges in W^c. □

The correspondence described in Proposition 6.20 allows us to translate algorithms for marginalization in graphical models to algorithms for contraction in tensor networks, see Section 6.2. Without care to order indices, marginalization and contraction involve summing exponentially many terms, but in many cases more efficient methods are possible.

Conditional distributions

Consider a probability distribution given by a fully observed graphical model. Conditioning a variable X_u to only take values in a given set $\mathfrak{Y}_u \subseteq \mathfrak{X}_u$ means restricting the probability tensor P to the slices $\mathfrak{Y}_u \times \prod_{b \in U \setminus \{u\}} \mathfrak{X}_b$ that contains only the values $\mathfrak{Y}_u$ for the variable X_u. This corresponds to restricting each of the potentials for hyperedges C containing u to the subset of elements $\mathfrak{Y}_u \times \prod_{b \in C \setminus \{u\}} \mathfrak{X}_b$. On the tensor networks side, we restrict the tensor corresponding to the given clique potential to the slice $\mathfrak{Y}_u \times \prod_{b \in C \setminus \{u\}} \mathfrak{X}_b$.

The equivalence of conditioning and restriction to slices of the probability tensor is due to the fact that the basis in which we view the probability tensor is fixed. The basis is given by the states of the random variables: graphical models are not basis invariant. On the other hand, basis invariance is a key property of tensor networks that crops up in many applications, e.g. often a basis, or gauge, is selected to make the computations efficient [137].

Entropy

Given a tensor network state represented by a tensor X, the *entanglement entropy* [137] equals

$$-\text{trace}(X \log X),$$

where $X \log X$ is a tensor, the same format as X, whose entry indexed by $\mathbf{i}$ is $x_{\mathbf{i}} \log x_{\mathbf{i}}$. On the other hand, if X represents the corresponding marginal distribution of the graphical model, the *Shannon entropy* [184] of X is defined as

$$H(X) = -\sum_{\mathbf{i}} x_{\mathbf{i}} \log x_{\mathbf{i}},$$

where $\mathbf{i}$ indexes all entries of X. Expanding out the formula $-\text{trace}(X \log T)$ shows that these two notions of entropy are the same.

6.2 Algorithms to contract tensor networks

Marginalization in graphical models is equivalent to contraction in tensor hypernetworks, see Proposition 6.20. This means we can contract tensor networks and hypernetworks using methods for marginalization of graphical models, which is widely studied [184]. I describe how this can be used to systematically find efficient ways to compute the expectation values of a tensor network, something that is usually addressed on a case-by-case basis depending on the structure of the tensor network.

The *belief propagation* (or *sum-product*) algorithm is a dynamic programming method for computing marginals of a distribution [184]. The *junction tree algorithm* [184] applies it to compute the marginals of a graphical model, defined with respect to a graphs with cycles. Expectation values of tensor hypernetwork states are obtained by contracting a tensor hypernetwork along all edges [137], while contracted tensor hypernetwork states contract along all full edges. In this section, I apply the junction tree algorithm to contract the matrix product state tensor networks from Example 6.9. I first recall the algorithm.

The junction tree algorithm

Input: A graphical model defined by a hypergraph H with clique potentials $\psi_C(x_C)$.
Output: The marginals at each hyperedge, $P(x_C) = \sum_{x_u : u \notin C} P(x)$.

The junction tree of a graph is a tree whose nodes are the maximal cliques of the graph. It has the *running intersection property*: the subset of cliques containing a given vertex forms a connected subtree. The junction tree algorithm works as follows. First, we construct the graph G associated to the hypergraph H by adding edge (i, j) whenever vertices i and j belong to the same hyperedge. If G is not chordal (or triangulated) we add edges until G is chordal. Then we form a junction tree of the graph G, noting that there are often multiple ways to construct a junction tree of a given graph G.

To each maximal clique C in G we associate a clique potential which equals the product of the potentials of the hyperedges contained in C. If a hyperedge is contained in more than one maximal clique, its clique potential is assigned to one of them. Each edge of the junction tree connects two cliques $C_1, C_2 \in \mathscr{C}$ in G. We associate to such an edge the *separator* set $S = C_1 \cap C_2$. We also assign a separator potential $\psi_S(x_S)$ to each S. It is initialized to the constant value 1. A *basic message passing operation* from C_1 to a neighbouring C_2 updates the potential functions at clique C_2 and separator $S = C_1 \cap C_2$

$$\tilde{\psi}_S(x_S) \leftarrow \sum_{x_{C_1 \setminus S}} \psi_{C_1}(x_{C_1}),$$

$$\tilde{\psi}_{C_2}(x_{C_2}) \leftarrow \frac{\tilde{\psi}_S(x_S)}{\psi_S(x_S)} \psi_{C_2}(x_{C_2}).$$

The algorithm chooses a root of the junction tree, and orients all edges to point from the root outwards. It then applies basic message passing operations step-by-step from the root

to the leaves until every node has received a message. Then we reverse the orientation of all edges, and update the clique and separator potentials from the leaves back to the root obeying the partial order given by the new orientations of the edges. After all messages have been passed, the final clique potentials equal the marginals, $\tilde{\psi_C}(x_C) = \sum_{x_u \notin C} \prod_{B \in \mathscr{C}} \psi_B(x_B)$, and likewise for the final separator potentials.

The complexity of the junction tree algorithm depends on the triangulation. It is exponential in the size of the largest clique of the chosen triangulation. Thus, at best, it is exponential in the *treewidth* of the graph, which is one less than the smallest size of the largest clique over all possible triangulations [184, Chapter 2].

Contracting tensor networks

A contracted tensor network state is formed by contracting all full edges. Consider the dual graphical model to the tensor hypernetwork. We make a new clique in the graphical model consisting of all vertices corresponding to the dangling edges of the tensor hypernetwork. The tensor hypernetwork state is the marginal distribution of that clique. Hence algorithms for marginalization of a graphical model can now be applied.

When the junction tree algorithm is used for probability distributions the clique potential functions are positive, but it works in the same way for complex valued functions. In this section I describe how to use the junction-tree algorithm to contract a tensor network.

Consider a Projected Entangled Pair States tensor network, as in Example 6.12, on four particles. The tensor network consists of four arrays

$$T_a \in \mathbb{C}^{n_1 \times n_2 \times n_5}, \quad T_b \in \mathbb{C}^{n_2 \times n_3 \times n_6}, \quad T_c \in \mathbb{C}^{n_3 \times n_4 \times n_7}, \quad T_d \in \mathbb{C}^{n_1 \times n_4 \times n_8},$$

where $n_1, n_2, \ldots, n_8$ are the dimensions of the vector spaces at the eight edges. Contracting the tensor network gives $T \in \mathbb{C}^{n_5 \times n_6 \times n_7 \times n_8}$ with entries:

$$t_{i_5,i_6,i_7,i_8} = \sum_{i_1,i_2,i_3,i_4} (t_a)_{i_1,i_2,i_5} (t_b)_{i_2,i_3,i_6} (t_c)_{i_3,i_4,i_7} (t_d)_{i_4,i_1,i_8}.$$

The junction tree algorithm gives a fast way to compute this sum. If the entanglement edge dimensions are $n_1 = n_2 = n_3 = n_4 = r$, and the dangling edge dimensions are $n_5 = n_6 = n_7 = n_8 = n$, we can compute the contracted tensor in time and space $O(n^3 r^2 + n^2 r^4)$, whereas summing term-by-term is $O(n^4 r^4)$. First, we find the junction tree of the tensor network. This process is illustrated in Figure 6.2. Then we assign clique potentials to each of the cliques and separators of the junction tree as follows:

$$\psi_{12458}(i_1, i_2, i_4, i_5, i_8) = (t_a)_{i_1,i_2,i_5} (t_d)_{i_4,i_1,i_8}, \quad \psi_{2458} = \psi_{245678} = \psi_{2467} = 1,$$

$$\psi_{23467}(i_2, i_3, i_4, i_6, i_7) = (t_b)_{i_2,i_3,i_6} (t_c)_{i_3,i_4,i_7}.$$

Next, we carry out the junction-tree algorithm. We choose the left vertex of the junction tree, i.e. 12458, as the root and proceed from left to right.

$$\tilde{\psi}_{2458}(i_2, i_4, i_5, i_8) = \sum_{i_1} \psi_{12458}(i_1, i_2, i_4, i_5, i_8),$$

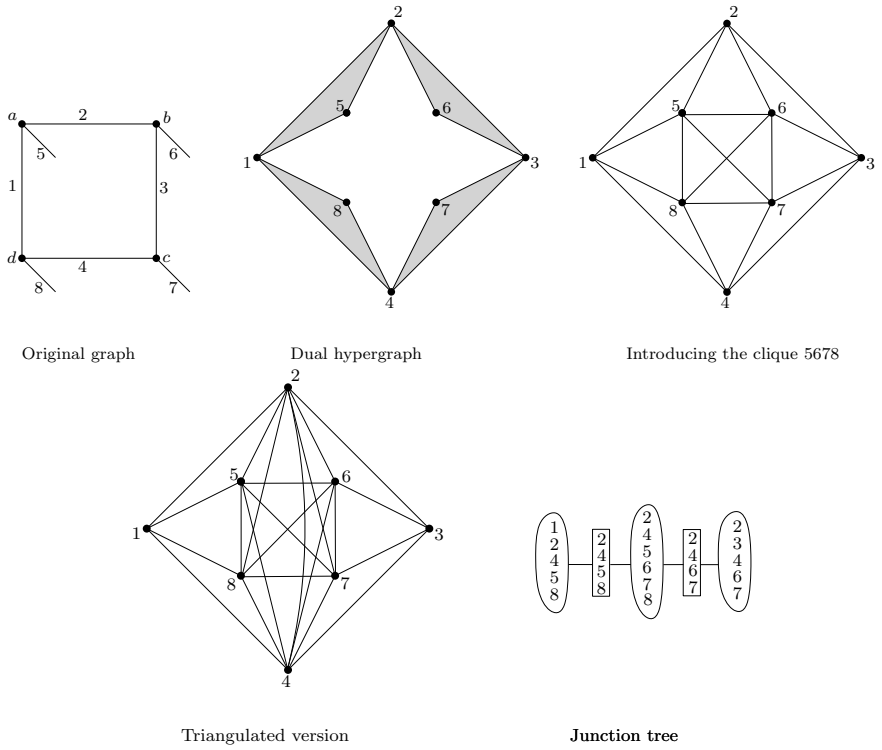


Figure 6.2: Steps to compute the junction tree of the projected pair entangled states tensor network on four particles.

$$\tilde{\psi}_{245678}(i_2, i_4, i_5, i_6, i_7, i_8) = \frac{\tilde{\psi}_{2458}(i_2, i_4, i_5, i_8)}{\psi_{2458}(i_2, i_4, i_5, i_8)} \psi_{245678}(i_2, i_4, i_5, i_6, i_7, i_8),$$

$$\tilde{\psi}_{2467}(i_2, i_4, i_6, i_7) = \sum_{i_5, i_8} \tilde{\psi}_{245678}(i_2, i_4, i_5, i_6, i_7, i_8),$$

$$\tilde{\psi}_{23467}(i_2, i_3, i_4, i_6, i_7) = \frac{\tilde{\psi}_{2467}(i_2, i_4, i_6, i_7)}{\psi_{2467}(i_2, i_4, i_6, i_7)} \psi_{23467}(i_2, i_3, i_4, i_6, i_7);$$

Then, we repeat the process going back to the root 12458. The first two steps of the updates, returning to the root, are as follows.

$$\tilde{\tilde{\psi}}_{2467}(i_2, i_4, i_6, i_7) = \sum_{i_3} \tilde{\psi}_{23467}(i_2, i_3, i_4, i_6, i_7),$$

$$\tilde{\tilde{\psi}}_{245678}(i_2, i_4, i_5, i_6, i_7, i_8) = \frac{\tilde{\tilde{\psi}}_{2467}(i_2, i_4, i_6, i_7)}{\tilde{\psi}_{2467}(i_2, i_4, i_6, i_7)} \tilde{\psi}_{245678}(i_2, i_4, i_5, i_6, i_7, i_8).$$

The desired marginal over $5, 6, 7, 8$ equals $\sum_{i_2, i_4} \tilde{\tilde{\psi}}_{245678}(i_2, i_4, i_5, i_6, i_7, i_8)$.

Note that for this particular graph, the complexity $O(n^3r^2 + n^2r^4)$ of the junction tree algorithm can also be achieved by factorizing T as

$$t_{i_5,i_6,i_7,i_8} = \sum_{i_2,i_4} \left(\sum_{i_1} (t_a)_{i_1,i_2,i_5} (t_d)_{i_4,i_1,i_8} \right) \left(\sum_{i_3} (t_b)_{i_2,i_3,i_6} (t_c)_{i_3,i_4,i_7} \right).$$

For more general graphs, finding a way to factor the contracted tensor network in order to match the performance of the junction tree algorithm is more difficult.

Expectation values for matrix product states

I now explain how to compute *expectation values* of matrix product state (MPS) tensor networks (see Figure 6.3) using the junction tree algorithm. The junction tree determines the order in which to contract the indices of the tensor network. Using Theorem 6.7, we contract the tensor network by applying the junction tree algorithm to the dual graphical model. I show that the junction tree algorithm used to marginalize the dual graphical model corresponds to the *bubbling* algorithms that are used to compute expectation values of a MPS [137].

In quantum applications a tensor network state is denoted $|\psi\rangle$. Its expectation value is the inner product $\langle\psi|A|\psi\rangle$ for some operator A. We consider the case that A is block diagonal, which means A transforms a tensor network state $|\psi\rangle$ by a linear transformation in each of its vector spaces of observable indices. The method we describe can be extended to operators of interest that are not block diagonal.

The expectation value of a MPS is computed by contracting the tensor network on the left in Figure 6.3, where the middle row of vertices correspond to the blocks of A. Equivalently, it is computed by marginalizing all variables of the graphical model on the right in Figure 6.3. The matrix product state is drawn with four observable indices, but repeating the pattern gives the results in the general case. The first step of the algorithm is to triangulate the graph of the graphical model, and to form the junction tree, see Figure 6.4.

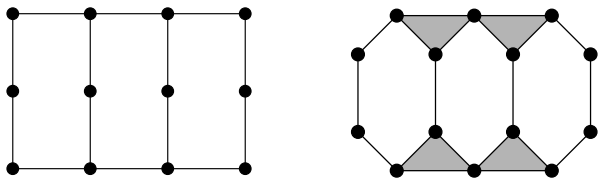


Figure 6.3: The matrix product state tensor network on four states contracted with itself (left) and its dual graphical model (right).

We choose the root of the tree to be the left-most vertex in the junction tree in Figure 6.4. We orient all edges to point away from the root, i.e. from left to right, and we perform

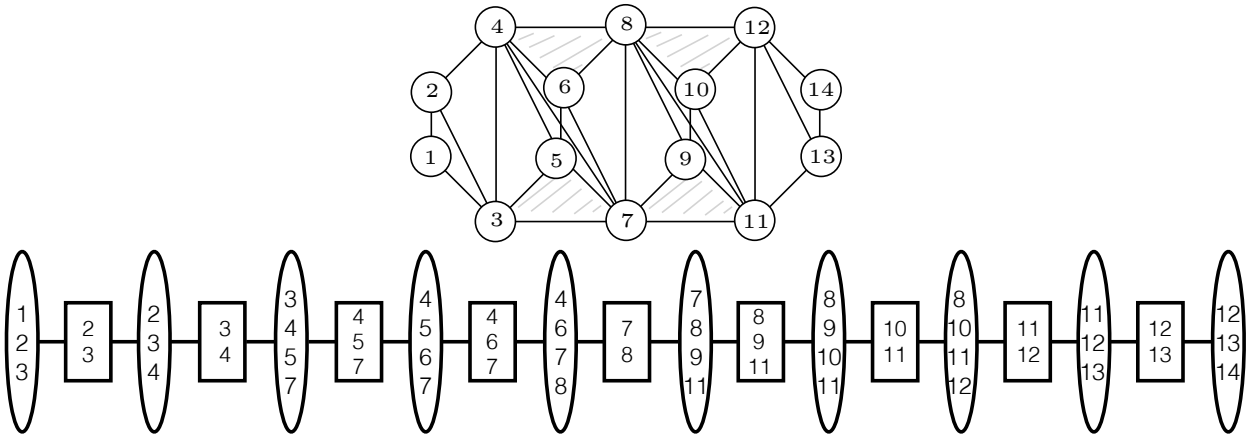


Figure 6.4: Triangulating the graphical model dual to a matrix product state (top). Its junction tree (bottom) where the cliques are in ovals and the separators are boxed.

basic message passing operations along the directed edges until every vertex has received a message from its parent. The function recorded at the clique $\{12, 13, 14\}$ is the marginal at that clique. In order to compute the total sum, we can simply sum over the three vertices 12, 13, and 14. In this special case, we do not need to run the second step of the junction tree algorithm that passes messages back to the root.

I now translate the junction tree algorithm to the language of tensor networks. At each step we sum over just one vertex of the dual graphical model (due to the structure of the junction tree in this case). This means we contract one edge at a time from the tensor network. The order of contractions is shown in Figure 6.5. For example, in the first message passing operation we have $C_1 = \{1, 2, 3\}$, $C_2 = \{2, 3, 4\}$, $S = \{2, 3\}$. We sum over the values of vertex 1, since it is the only variable in $C_1 \backslash S$. This corresponds to contracting the tensor along the edge corresponding to vertex 1 of the graphical model.

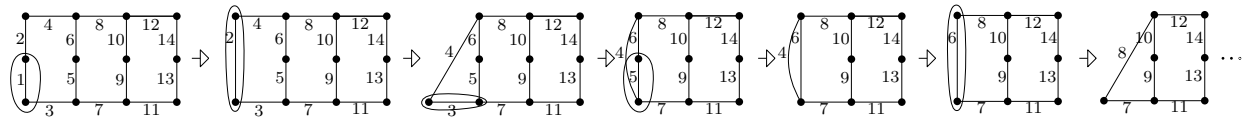


Figure 6.5: Order of contraction of indices in the matrix product state tensor network, to compute its expectation value using the junction tree algorithm.

The triangulated graph of the dual graph of MPS has a treewidth of size four, since we can continue the triangulation given in Figure 6.4 to an arbitrary number of steps. We can compute the complexity of the junction tree algorithm to be $O(|V|(nr^3 + n^2r^2))$ where $|V|$ is the number of vertices in the MPS, n is the weight of the dangling edges, and r is the

weight of the entanglement edges. It turns out that contracting the tensor in this way is what is usually done by the tensor networks community as well, a method sometimes called bubbling [137]. Similar algorithms are used in the case of MPS with periodic boundary conditions, e.g. the algorithm in [143] which runs in $O(|V|nr^5)$. A numerical algorithm for an infinitely long MPS chain which runs in $O(nr^3)$ is given in [157].

I conclude this chapter with a brief discussion of projected entangled pair states, the two-dimensional generalization of MPS. Projected entangled pair states (PEPS), see Example 6.12, have a two-dimensional lattice of entanglement interactions. Computing expectation values for the PEPS network takes exponential time in the number of states of the network [137]. On the graphical models side, it is possible in principle to find expectation values of a PEPS state using the junction tree algorithm. Since the triangulated graph of the dual hypergraph of PEPS has a tree-width that grows in the size of the network, the junction tree algorithm is exponential time.

In [118], the authors show that algorithms for computing expectation values are exponential in the treewidth of the tensor network. On the other hand, we have seen that the junction tree algorithm is exponential time in the treewidth of the dual graphical model. This indicates a similarity between the treewidth of a hypergraph and of its dual. A comparison of the treewidths of planar graphs and of their graph duals can be found in [150].

To avoid exponential running times, numerical approximations are used [132, 137, 189]. For graphical models, these are termed *loopy belief propagation*, see [184, Chapter 4] and references therein. A natural question is whether the algorithms for loopy belief propagation translate to known algorithms in the tensor networks community, e.g. for computing expectation values of PEPS, or whether they provide a new family of algorithms.

In this chapter we saw the connection between multivariate statistics and algorithms for contracting tensors. This brings me to the second part of my thesis, in which I study algorithms for tensor data.

Part II

Algorithms for tensor data

Chapter 7

Semi-algebraic statistics

The joint probabilities of several finite random variables $X_1, \ldots, X_d$ can be organized into a tensor P with entries

$$p_{i_1 \ldots i_d} = P(X_1 = i_1, \ldots, X_d = i_d).$$

If the random variable X_j has n_j possible states, we obtain a tensor of format $n_1 \times \cdots \times n_d$. In this chapter I usually denote a tensor by the letter P to emphasize that its entries $p_{i_1 i_2 \ldots i_d}$ represent probabilities, and to avoid confusion with random variables which are denoted by the letter X. The purpose of representing a distribution as a tensor is that the structure of the tensor P can be given statistical interpretation.

Example 7.1 (Rank one = full independence)**.** *A tensor is rank one if it can be written in the form $v^{(1)} \otimes \cdots \otimes v^{(d)}$. If the entries of P are non-negative and sum to one, then the vectors $v^{(j)}$ can also be chosen to be non-negative with entries summing to one. The full independence model on d random variables $X_1, \ldots, X_d$ is the set of distributions whose joint probabilities factor as*

$$P(X_1 = i_1, \ldots, X_d = i_d) = P(X_1 = i_1) \cdots P(X_d = i_d).$$

Hence, setting $v^{(j)}_{i_j} = P(X_j = i_j)$, we see that the full independence model is equal to the set of rank one non-negative tensors with entries summing to one. The model is given implicitly as the intersection of the probability simplex $\Delta_{n_1 \cdots n_d - 1}$ with the cone over the Segre variety $\mathrm{Seg}(\mathbb{P}^{n_1 - 1} \times \cdots \times \mathbb{P}^{n_d - 1})$.

More generally, a statistical model on the random variables X_j is a subset of the probability simplex. What might we want to know about a statistical model? Whether an empirical distribution of interest lies in, or close to, the model indicates whether the model is a suitable way to represent the empirical distribution. We could be interested in a membership test for the model, or in the representational power of the statistical model in general. We could also be interested in the distributions that the model represents poorly, those furthest away from the model. The maximum likelihood estimate of our empirical distribution with respect to the model gives the point in the model that is most likely to give rise to the empirical

data. This is the point on the model that minimizes the Kullback-Leibler divergence from the empirical distribution to the model.

Many statistical models can be parametrized by polynomials. For such statistical models, we can approach the above questions using algebraic tools. The use of algebraic methods in statistics is *algebraic statistics*. A statistical model that is parametrized by polynomials has a semi-algebraic parametric description. By Theorem 1.23, the set of distributions which lie in the model is a semi-algebraic subset of the probability simplex.

If we can find the semi-algebraic description of a statistical model, it offers the possibility of answering statistical questions for the model. The semi-algebraic description, also known as an implicit description, gives a membership test for the model, allows the computation of divergence to the model, and can suggest model-specific algorithms for parameter estimation. Methods from algebraic statistics often ignore the inequalities defining a semi-algebraic set of interest, focusing only on the equations. This gives an outer-approximation to the statistical model, and certain properties (such as the dimension) are preserved by this relaxation. Here, I seek a semi-algebraic description for statistical models, inequalities included. Statistical models with hidden variables offer better models of real-world settings but, for models with hidden variables, the semi-algebraic description is difficult to find. For more on algebraic statistics, see [179], and for semi-algebraic statistics in the context of phylogenetic models, see [194].

In this chapter, I present a case study in which semi-algebraic methods are used to describe two statistical models. One is a mixture model and the other is a product of mixtures model called a restricted Boltzmann machine, both on three binary random variables. Although the two models look different from their parametrizations, I show that they represent the same set of distributions on the interior of the probability simplex, and are equal up to closure, resolving a conjecture due to Montúfar and Morton [125]. I give a semi-algebraic description of the model in terms of six binomial inequalities. Exact maximum likelihood estimates could previously only be found for models with no hidden variables. For these models with hidden variables, we can also use the implicit description to give an exact description of the projection to the boundary strata of the model, and this leads to a closed form expression for the maximum likelihood estimate. This case study could be used for methods to find implicit descriptions of larger statistical models, and I briefly discuss such extensions. This chapter is joint work with Guido Montúfar, published in the Journal of Algebraic Statistics [163].

7.1 Mixture models and restricted Boltzmann machines

The joint probabilities of a multivariate probability distribution can be organized into a tensor, and the structure of the tensor encodes statistical information about the distribution. As we saw above in Example 7.1, a distribution is in the full independence model if and only

if its tensor has rank one. Extending this to higher ranks, we have the following.

Example 7.2 (Non-negative rank = mixture model)**.** *Consider a tensor P, of non-negative rank r, with entries summing to one. Then P has a decomposition*

$$P = \sum_{i=1}^{r} \lambda_i w_i^{(1)} \otimes \cdots \otimes w_i^{(d)},$$

where $w_i^{(j)} \in \mathbb{R}^{n_j}$ is a vector of non-negative entries summing to one, and the λ_i are non-negative scalars with $\sum_{i=1}^{r} \lambda_i = 1$. Non-negative rank was introduced in Definition 1.29. This re-writing of the decomposition to have vector summing to one allows us to interpret the weights λ_i of the rank one terms as the probabilities of a hidden variable, and the entries of the vectors $w_i^{(d)}$ as conditional distributions, which are independent conditional on the value of the hidden variable. The distribution P is a mixture of independent distributions, with a single hidden variable that has r states,

$$p_{i_1,\ldots,i_d} = P(X_1 = i_1, \ldots, X_d = i_d) = \sum_{j=1}^{r} P(X_1 = i_1 | Y = j) \cdots P(X_d = i_d | Y = j) P(Y = j).$$

The non-negative rank is the smallest possible number of states of the hidden variable. As we will see, fitting a distribution to this model, with a small number of hidden states, can be viewed as finding a low non-negative rank approximation of the tensor, see [110]. This model is called the Naive Bayes model. For three observed variables, X_1, X_2, X_3, the statistical model can be drawn as in Figure 7.1.

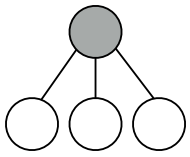


Figure 7.1: A mixture model on three observed (unshaded) variables that are independent, conditioned on the state of the hidden (shaded) variable.

Other arrangements of observed and hidden variables can be used to represent multivariate probability distributions. These are given by graphs of shaded and unshaded variables such as Figure 7.1. Such distributions are called graphical models; they are specified in terms of conditional independence relations between the variables. For fully observed graphical model (with no hidden variables) the implicit description encodes the conditional independence relations that hold for the model, see [73]. We saw in Chapter 6 that graphical models are dual to tensor hypernetworks.

Any probability distribution can be modeled by a graphical model, for instance a complete undirected graph imposes no constraints on the distribution. However, certain graphs involve

many more parameters than others to represent specific distributions. In the interest of concisely representing data and reducing computational costs, one would like to understand which graphs best represent which kinds of data. For example, deep architectures, with several layers of hidden variables, have become increasingly important in machine learning, see [26] and references therein. Following [125], I focus on two building blocks to such multi-layer architectures. I study these in the case where all random variables are binary.

1. One hidden variable with r states, connected to d observed variables. This is the *mixture of products* model, denoted $\mathcal{M}_{d,r}$, see Example 7.2. Up to scaling, it consists of $2 \times \cdots \times 2$ (d times) tensors of non-negative rank at most r,

$$P = \sum_{i=1}^{r} v_i^{(1)} \otimes \cdots \otimes v_i^{(d)}, \qquad v_i^{(j)} \in \mathbb{R}^2_{\geq 0}. \tag{7.1}$$

2. A layer of m hidden binary variables, each connected to d observed variables. This is the *restricted Boltzmann machine* (RBM) model $\mathrm{RBM}_{d,m}$. Up to scale, it consists of $2\times\cdots\times 2$ (d times) tensors that are the Hadamard product of m tensors of non-negative rank at most two,

$$P = \prod_{i=1}^{m} (v_i^{(1)} \otimes \cdots \otimes v_i^{(d)} + w_i^{(1)} \otimes \cdots \otimes w_i^{(d)}), \qquad v_i^{(j)}, w_i^{(j)} \in \mathbb{R}^2_{\geq 0}. \tag{7.2}$$

Note that $\mathcal{M}_{d,1}$ is the *independence model*, and $\mathcal{M}_{d,2} = \mathrm{RBM}_{d,1}$, since both have a single hidden variable with two states. In [6] the description of $\mathcal{M}_{d,2}$ is found. The authors describe the 'formidable obstacles' to extending their results to hidden variables with more than two states.

Remark 7.3 (Marginals of exponential families)**.** *The mixture of products model and RBM model are are often defined as marginals of exponential families, instead of in the polynomial parametrization above. The mixture model is parametrized by*

$$p(x) = \frac{1}{Z(W,b,c)} \sum_{y \in \{e_j \,:\, j=1,\ldots,r\}} \exp(y^\top W x + c^\top y + b^\top x)$$

and the RBM model is parametrized by

$$p(x) = \frac{1}{Z(W,b,c)} \sum_{y \in \{0,1\}^m} \exp(y^\top W x + c^\top y + b^\top x),$$

where the variable x ranges over $\{0,1\}^d$ and $Z(W,b,c)$ normalizes the entries to sum to one. In contrast to the exponential parametrization, the polynomial parametrization allows zeros in the decomposition, while requiring that the entries sum to one. The polynomial and exponential definitions are equivalent up to closure, see for example [125, Proposition 2.3].

The descriptions of the mixture model and RBM model in Equations (7.1) and (7.2) are parametric. Each value of the parameters gives a distribution that lies in the model. However, these descriptions do not allow us to test if an empirical distribution of interest lies in the model. For this, we seek the implicit description. In most previous work on the representational power of RBMs, membership in the model is determined by constructing parameters that realize certain probability distributions. In contrast, the implicit descriptions discussed here fully characterize distributions that are in the model.

7.2 Implicit descriptions of statistical models

In this section, I consider the restricted Boltzmann machine $\mathrm{RBM}_{3,2}$ and the mixture model $\mathcal{M}_{3,3}$. Both are models on three binary random variables, so they are sets of $2 \times 2 \times 2$ tensors P with entries p_{ijk} for $0 \leq i, j, k \leq 1$ that lie in the probability simplex $\Delta_{2^3-1} = \Delta_7$. Both models are over-parametrized in Δ_7, since they have 11 parameters. In [125], it is shown that $\mathcal{M}_{3,3}$ does not fill the simplex. The authors state 'we believe that $\mathcal{M}_{3,3}$ and $\mathrm{RBM}_{3,2}$ are very similar, if not equal.'

In this section, I give a implicit description of both models $\mathrm{RBM}_{3,2}$ and $\mathcal{M}_{3,3}$, and I show that the two models describe, up to closure, the same distributions in the probability simplex. The main theorems are the following.

Theorem 7.4. *The statistical model* $\mathrm{RBM}_{3,2}$ *is described on the interior of the simplex* Δ_7 *by the union of six basic semi-algebraic sets:*

$$\begin{aligned}
&\{p_{000}p_{011} \geq p_{001}p_{010}, \quad p_{100}p_{111} \geq p_{101}p_{110}\}\\
&\{p_{000}p_{011} \leq p_{001}p_{010}, \quad p_{100}p_{111} \leq p_{101}p_{110}\}\\
&\{p_{000}p_{101} \geq p_{001}p_{100}, \quad p_{010}p_{111} \geq p_{011}p_{110}\}\\
&\{p_{000}p_{101} \leq p_{001}p_{100}, \quad p_{010}p_{111} \leq p_{011}p_{110}\}\\
&\{p_{000}p_{110} \geq p_{100}p_{010}, \quad p_{001}p_{111} \geq p_{101}p_{011}\}\\
&\{p_{000}p_{110} \leq p_{100}p_{010}, \quad p_{001}p_{111} \leq p_{101}p_{011}\}.
\end{aligned}$$

These binomial inequalities correspond to determinants of slices of the tensor P. They record conditional correlations in the distribution.

Theorem 7.5. *We have the equality* $\mathcal{M}_{3,3} = \overline{\mathrm{RBM}_{3,2}}$. *Equality* $\mathcal{M}_{3,3} = \mathrm{RBM}_{3,2}$ *holds on the interior of the simplex.*

The notation $\overline{\mathrm{RBM}_{3,2}}$ refers to the topological closure of $\mathrm{RBM}_{3,2}$. The mixture model $\mathcal{M}_{3,3}$ and the RBM model $\mathrm{RBM}_{3,2}$ look quite different in their parametrization, but this result shows that they turn out to parametrize the same probability distributions (up to closure). The parametrization of $\mathrm{RBM}_{3,2}$ in Equation (7.2) does not describe a closed set on the boundary of the simplex. We describe $\mathrm{RBM}_{3,2}$ on the boundary of the simplex in Proposition 7.8. On the other hand, $\mathcal{M}_{3,3}$ is closed (see Proposition 7.10) and we have the following corollary.

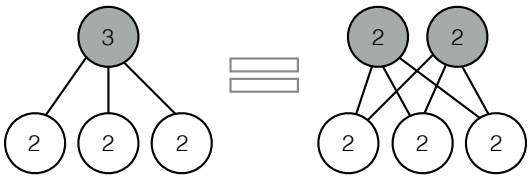


Figure 7.2: Theorem 7.5 gives the equality of these two graphical models. The label of a variable is its number of states; the shaded nodes are hidden.

Corollary 7.6. *The model $\mathcal{M}_{3,3}$ is described on Δ_7 by the inequalities in Theorem 7.4.*

Previous results showed that $\mathcal{M}_{3,3}$ has relative volume at most 96.4%, and $\mathrm{RBM}_{3,2}$ has relative volume at most 99.2% inside the simplex Δ_7 [125]. Simulations using Theorem 7.4 and Corollary 7.6 estimate the true volume of both of these models to be 75.3%.

We use Theorem 7.4 to prove a conjecture from [125, Section 3.5.1]:

Corollary 7.7. *No distribution in* $\mathrm{RBM}_{3,2}$ *has four modes.*

For a discrete distribution, a *mode* is a state with larger probability than any of its Hamming neighbor states. Corollary 7.7 is stated as a conjecture $\mathrm{RBM}_{3,2} \cap \mathcal{G}_3 = \emptyset$ in [125], where $\mathcal{G}_3$ denotes distributions on $\{0,1\}^3$ with four modes (the maximum possible number). Note that the models $\mathcal{M}_{3,4}$ and $\mathrm{RBM}_{3,3}$ fill the interior of the simplex Δ_7 [126, 127]. Corollary 7.7 also follows from Theorem 7.5, since no $P \in \mathcal{M}_{3,3}$ has four modes [125, Proposition 3.10]. I now prove Theorem 7.4.

The implicit description of $\mathrm{RBM}_{3,2}$

The semi-algebraic description of the non-negative rank at most two model $\mathcal{M}_{3,2}$ is given in [6]. The model is described in Δ_7 by the union of four basic semi-algebraic sets. On the interior of the simplex, one of the sets is given by the inequalities

$$\begin{array}{lll} p_{000}p_{011} \geq p_{010}p_{001}, & p_{000}p_{101} \geq p_{100}p_{001}, & p_{000}p_{110} \geq p_{100}p_{010}, \\ p_{100}p_{111} \geq p_{110}p_{101}, & p_{010}p_{111} \geq p_{110}p_{011}, & p_{001}p_{111} \geq p_{101}p_{011}. \end{array} \tag{7.3}$$

The other three sets are obtained by reversing the signs of the inequalities in *two out of the three* columns of Equation (7.3). For example:

$$\begin{array}{lll} p_{000}p_{011} \geq p_{010}p_{001}, & p_{000}p_{101} \leq p_{100}p_{001}, & p_{000}p_{110} \leq p_{100}p_{010}, \\ p_{100}p_{111} \geq p_{110}p_{101}, & p_{010}p_{111} \leq p_{110}p_{011}, & p_{001}p_{111} \leq p_{101}p_{011}. \end{array} \tag{7.4}$$

One way to get a distribution in $\mathrm{RBM}_{3,2}$ is to take the Hadamard product of a distribution satisfying Equation (7.3) with one satisfying Equation (7.4). We find the semi-algebraic description for all distributions expressible as such a Hadamard product. From this, swapping indices gives the full semi-algebraic description of the restricted Boltzmann machine $\mathrm{RBM}_{3,2}$ on the interior of the simplex. Note that the independence model $\mathcal{M}_{3,1}$ is obtained on the interior of Δ_7 by setting the inequalities in Equation (7.3) or Equation (7.4) to equalities.

On the interior of the simplex

The binomial inequalities above translate to linear inequalities in the log-probabilities. The inequalities are independent of scaling and we can work with unnormalized distributions. For a strictly positive distribution P, we take the log distribution $l_{ijk} = \log(p_{ijk})$. Taking the logarithm of the inequalities in Equation (7.3) gives the polyhedron

$$\mathcal{X} = \begin{Bmatrix} l_{000} + l_{011} - l_{001} - l_{010} \geq 0, & l_{100} + l_{111} - l_{101} - l_{110} \geq 0 \\ l_{000} + l_{101} - l_{001} - l_{100} \geq 0, & l_{010} + l_{111} - l_{011} - l_{110} \geq 0 \\ l_{000} + l_{110} - l_{010} - l_{100} \geq 0, & l_{001} + l_{111} - l_{011} - l_{101} \geq 0 \end{Bmatrix}.$$

Similarly, we define $\mathcal{Y}$ to be the log-probabilities satisfying the logarithms of Equation (7.4),

$$\mathcal{Y} = \begin{Bmatrix} l_{000} + l_{011} - l_{001} - l_{010} \geq 0, & l_{100} + l_{111} - l_{101} - l_{110} \geq 0 \\ l_{000} + l_{101} - l_{001} - l_{100} \leq 0, & l_{010} + l_{111} - l_{011} - l_{110} \leq 0 \\ l_{000} + l_{110} - l_{010} - l_{100} \leq 0, & l_{001} + l_{111} - l_{011} - l_{101} \leq 0 \end{Bmatrix}.$$

Taking the Hadamard product in probability space is the same as taking the sum in log-probability space. Therefore, showing Theorem 7.4 is equivalent to proving that the Minkowski sum $\mathcal{X} + \mathcal{Y} = \{x + y : x \in \mathcal{X}, y \in \mathcal{Y}\}$ is

$$\mathcal{W} = \{l_{000} + l_{011} - l_{001} - l_{010} \geq 0, \quad l_{100} + l_{111} - l_{101} - l_{110} \geq 0\}.$$

The two polyhedra $\mathcal{X}$ and $\mathcal{Y}$ are eight-dimensional in $\mathbb{R}^8$. The lineality spaces of a polyhedron is the space obtained by setting all the inequalities in their descriptions to equalities. For both $\mathcal{X}$ and $\mathcal{Y}$, the lineality space is the set of tensors (l_{ijk}) for which the tensor $(\exp(l_{ijk}))$ is rank one. It is spanned by the rows of the matrix

$$\begin{array}{c} \begin{matrix} l_{000} & l_{100} & l_{010} & l_{001} & l_{110} & l_{101} & l_{011} & l_{111} \end{matrix} \\ \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 & 1 & 1 \end{pmatrix}. \end{array}$$

The polyhedron $\mathcal{W}$ is also eight-dimensional. It has a six-dimensional lineality space that is spanned degenerately by the rows of the matrix

$$\begin{array}{c} \begin{matrix} l_{000} & l_{100} & l_{010} & l_{001} & l_{110} & l_{101} & l_{011} & l_{111} \end{matrix} \\ \begin{pmatrix} 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \end{pmatrix}. \end{array} \tag{7.5}$$

Using the software `polymake` [72], we can find a description for the quotient of $\mathcal{X}$ or $\mathcal{Y}$ by its lineality space. They are both triangular bipyramids.

Proof of Theorem 7.4. We aim to show that $\mathcal{W} = \mathcal{X} + \mathcal{Y}$. We begin with the containment $\mathcal{X} + \mathcal{Y} \subseteq \mathcal{W}$. Summing the first equations in $\mathcal{X}$ and $\mathcal{Y}$ yields

$$x_{000} + y_{000} + x_{011} + y_{011} - x_{001} - y_{001} - x_{010} - y_{010} \geq 0,$$

while summing the second equations from $\mathcal{X}$ and $\mathcal{Y}$ gives

$$x_{100} + y_{100} + x_{111} + y_{111} - x_{101} - y_{101} - x_{110} - y_{110} \geq 0.$$

Translating back to the l-coordinates, we get $l_{000} + l_{011} - l_{001} - l_{010} \geq 0$ and $l_{100} + l_{111} - l_{101} - l_{110} \geq 0$. Hence $\mathcal{X} + \mathcal{Y} \subseteq \mathcal{W}$.

For the reverse containment $\mathcal{W} \subseteq \mathcal{X} + \mathcal{Y}$ we require a spanning set for $\mathcal{W}$ in which every basis vector lies either in $\mathcal{X}$ or in $\mathcal{Y}$. The first four rows of the lineality space of $\mathcal{W}$ in Equation (7.5) lie in $\mathcal{X}$, while the last four rows lie in $\mathcal{Y}$. Hence any non-negative combination of the lineality space lies in $\mathcal{W}$. To extend to negative linear combinations we multiply the spanning set by -1. The first four rows of the negation of Equation (7.5) lie in $\mathcal{Y}$, and the last four are in $\mathcal{X}$.

It remains to find a basis for the two-dimensional polytope obtained by taking the quotient of $\mathcal{W}$ by its lineality space. The quotient is spanned by non-negative combinations of any two linearly independent vectors in $\mathcal{W}$ not in its lineality space. For example $l_{000} \in \mathcal{X}$ and $l_{100} \in \mathcal{Y}$. All non-negative combinations of these lie in $\mathcal{X} + \mathcal{Y}$. □

On the boundary of the simplex

We now have a semi-algebraic description for the restricted Boltzmann machine $\mathrm{RBM}_{3,2}$ on the interior of the simplex Δ_7. However, for P in the boundary of the simplex $\partial\Delta_7$, the inequalities in Theorem 7.4 are not sufficient for membership in $\mathrm{RBM}_{3,2}$.

Proposition 7.8. *The intersection* $\mathrm{RBM}_{3,2} \cap \partial\Delta_7$ *is given by distributions which satisfy*

$$\begin{array}{c} \textit{If the probability of a state vanishes, so does the} \\ \textit{probability of one of its Hamming neighbour states.} \end{array} \tag{7.6}$$

Proof. First we show that $P \in \mathrm{RBM}_{3,2} \cap \partial\Delta_7$ satisfies Condition (7.6). Since P lies on the boundary of Δ_7, one of its entries vanishes. Assume without loss of generality $p_{000} = 0$. Then Condition (7.6) means that $p_{100}p_{010}p_{001} = 0$. Since $P \in \mathrm{RBM}_{3,2}$, it is the product of two distributions in $\mathcal{M}_{3,2}$. That is,

$$p_{ijk} = (q_{ijk} + r_{ijk})(s_{ijk} + t_{ijk}),$$

where Q, R, S, T are rank one non-negative $2 \times 2 \times 2$ tensors. Up to swapping factors the $(0, 0, 0)$ entry of the tensor $Q + R$ must vanish. Hence $q_{000} = r_{000} = 0$. Since Q and R are

rank one, they must vanish on a slice. Both Q and R vanish in at least one of the locations $(0,0,1)$, $(0,1,0)$ and $(1,0,0)$, hence so does P.

For the converse, we consider some $P \in \partial\Delta_7$ satisfying Condition (7.6) and we aim to show that $P \in \mathrm{RBM}_{3,2}$. As before, we can assume $p_{000} = 0$. Condition (7.6) implies that one of $p_{001}, p_{010}, p_{100}$ must also vanish. We reorder indices such that p_{010} vanishes. The distribution admits the Hadamard factorization

$$P = \left[\begin{array}{cc||cc} 0 & 0 & p_{001} & p_{011} \\ p_{100} & p_{110} & p_{101} & p_{111} \end{array}\right] = \left[\begin{array}{cc||cc} 0 & 0 & p_{001} & p_{011} \\ p_{101} & p_{111} & p_{101} & p_{111} \end{array}\right] * \left[\begin{array}{cc||cc} 0 & 0 & 1 & 1 \\ \frac{p_{100}}{p_{101}} & \frac{p_{110}}{p_{111}} & 1 & 1 \end{array}\right].$$

If $p_{101}, p_{111} \neq 0$, both factors are non-negative rank two and the distribution lies in $\mathrm{RBM}_{3,2}$. If $p_{101} = 0$, then $p_{111}p_{100}p_{001} = 0$ and if $p_{111} = 0$ then $p_{110}p_{101}p_{011} = 0$. In both of these cases the distribution consists of two pairs of non-zero adjacent entries, hence lies in $\mathcal{M}_{3,2}$, which is a subset of $\mathrm{RBM}_{3,2}$. Hence in all cases the distribution lies in $\mathrm{RBM}_{3,2}$. □

Condition (7.6) is stricter than the restriction of the inequalities in Theorem 7.4 to the boundary of the simplex. The model $\mathrm{RBM}_{3,2}$ is a semi-algebraic subset of the simplex that is not closed. I give an example of a distribution that lies in the closure of the model, but not in the model.

Example 7.9. *Consider the distribution*

$$p_{ijk} = \begin{cases} \frac{1}{3}, & (i,j,k) \in \{(0,0,1),(0,1,0),(1,0,0)\} \\ 0, & \textit{otherwise.} \end{cases}$$

Observe that $P \in \mathcal{M}_{3,3}$, since $P = \frac{1}{3}(e_0 \otimes e_0 \otimes e_1 + e_0 \otimes e_1 \otimes e_0 + e_1 \otimes e_0 \otimes e_0)$ has non-negative rank three and entries summing to one. Since P does not satisfy the conditions in Proposition 7.8, $P \notin \mathrm{RBM}_{3,2}$. I give a sequence of distributions $(P_\epsilon) \subset \mathrm{RBM}_{3,2}$, such that $P = \lim_{\epsilon\to 0} P_\epsilon$. Consider

$$P_\epsilon \propto \left[\begin{array}{cc||cc} \epsilon & 1 & 1 & \epsilon \\ 1 & \epsilon & \epsilon & \epsilon^4 \end{array}\right].$$

As $\epsilon \to 0$, $P_\epsilon \to P$. The scaling factor can be subsumed to either factor in the following decomposition.

$$\begin{aligned} P_\epsilon &\propto \left[\begin{array}{cc||cc} \epsilon & 1 & \epsilon^2 & \epsilon \\ 1 & \epsilon & \epsilon & \epsilon^2 \end{array}\right] * \left[\begin{array}{cc||cc} 1 & 1 & \epsilon^{-2} & 1 \\ 1 & 1 & 1 & \epsilon^2 \end{array}\right] \\ &= \left(\begin{bmatrix}\epsilon\\1\end{bmatrix} \otimes \begin{bmatrix}1\\0\end{bmatrix} \otimes \begin{bmatrix}1\\\epsilon\end{bmatrix} + \begin{bmatrix}1\\\epsilon\end{bmatrix} \otimes \begin{bmatrix}0\\1\end{bmatrix} \otimes \begin{bmatrix}1\\\epsilon\end{bmatrix}\right) * \left(\begin{bmatrix}1\\1\end{bmatrix} \otimes \begin{bmatrix}1\\1\end{bmatrix} \otimes \begin{bmatrix}1\\0\end{bmatrix} + \begin{bmatrix}\epsilon^{-1}\\\epsilon\end{bmatrix} \otimes \begin{bmatrix}\epsilon^{-1}\\\epsilon\end{bmatrix} \otimes \begin{bmatrix}0\\1\end{bmatrix}\right) \end{aligned}$$

This decomposition shows that $P_\epsilon \in \mathrm{RBM}_{3,2}$ for each ϵ. Hence $\mathrm{RBM}_{3,2}$ is not closed.

In the example above, the entries of one of the tensors in the decomposition are unbounded as $\epsilon \to 0$. They are multiplied by very small entries in the other term so that the limiting tensor P is bounded. Such situations can be avoided on the interior of the simplex, where the model $\mathrm{RBM}_{3,2}$ is closed, and can also be avoided for the mixture model $\mathcal{M}_{3,3}$.

Proposition 7.10. *The model $\mathcal{M}_{d,r}$ is closed for all d and r.*

Proof. Consider a convergent sequence of tensors $P_n \to P$, where each $P_n \in \mathcal{M}_{d,r}$. We show that the limiting tensor P also lies in $\mathcal{M}_{d,r}$. By definition, each P_n can be written as the sum of r non-negative rank one tensors $P_n = A_n + B_n + \cdots + C_n$. Since the entries of P_n are bounded above by 1, and the entries of $A_n, B_n, \ldots, C_n$ are non-negative, the entries of $A_n, B_n, \ldots, C_n$ are also bounded above by 1. By the Bolzano Weierstrass Theorem, there exists a subsequence of the A_n, call it A_{n_j}, that converges. Its limit, A, is a non-negative rank one tensor. Taking $P_{n_j} \to P$ as our new convergent sequence, we repeat the argument to find a convergent subsequence of the B_{n_j} which converges to a non-negative rank one tensor B. Repeating r times we obtain a subsequence of the P_n whose limit is $A + B + \cdots + C$. Hence $P = A + B + \cdots + C \in \mathcal{M}_{d,r}$. The result also follows directly from topological considerations, since $\mathcal{M}_{d,r}$ is the image of the closed set $(\Delta_1)^{dr} \times \Delta_{r-1}$ under a polynomial map. □

Equality of $\mathrm{RBM}_{3,2}$ and $\mathcal{M}_{3,3}$

We prove Theorem 7.5 by proving the two directions of the containment in two lemmas. Equality on the interior of the simplex follows from equality of the model closures by the fact that $\mathrm{RBM}_{3,2}$ is closed on the interior of the simplex.

Lemma 7.11. *We have the containment of statistical models $\mathrm{RBM}_{3,2} \subseteq \mathcal{M}_{3,3}$.*

Proof. Consider a distribution $P \in \mathrm{RBM}_{3,2}$. If $P \in \partial\Delta_7$ then it satisfies Condition (7.6) and we can assume without loss of generality $p_{000} = p_{001} = 0$. Then

$$P = \left[\begin{array}{cc||cc} 0 & 0 & 0 & 0 \\ p_{100} & 0 & p_{101} & 0 \end{array}\right] + \left[\begin{array}{cc||cc} 0 & p_{010} & 0 & p_{011} \\ 0 & 0 & 0 & 0 \end{array}\right] + \left[\begin{array}{cc||cc} 0 & 0 & 0 & 0 \\ 0 & p_{110} & 0 & p_{111} \end{array}\right]$$

is an expression for P as the sum of three non-negative rank one terms, hence $P \in \mathcal{M}_{3,3}$.

It remains to consider distributions P with no entries vanishing. We name the six determinants by $d_{i,j}$ where $i \in \{1,2,3\}$ denotes which index is fixed in the determinant, and $j \in \{0,1\}$ gives the value of the fixed index:

$$\begin{aligned} d_{1,0} &= p_{000}p_{011} - p_{001}p_{010}, \quad d_{1,1} = p_{100}p_{111} - p_{101}p_{110}, \\ d_{2,0} &= p_{000}p_{101} - p_{001}p_{100}, \quad d_{2,1} = p_{010}p_{111} - p_{011}p_{110}, \\ d_{3,0} &= p_{000}p_{110} - p_{010}p_{100}, \quad d_{3,1} = p_{001}p_{111} - p_{011}p_{101}. \end{aligned} \tag{7.7}$$

As we will see in Section 7.4 and Figure 7.4, we can relabel indices such that determinants $d_{2,1}$ and $d_{1,1}$ have opposite signs. We can write P as

$$P = \left[\begin{array}{cc||cc} p_{000} & 0 & p_{001} & 0 \\ 0 & 0 & 0 & 0 \end{array}\right] + \left[\begin{array}{cc||cc} 0 & 0 & 0 & 0 \\ p_{100} & x & p_{101} & \frac{p_{101}}{p_{100}}x \end{array}\right] + \left[\begin{array}{cc||cc} 0 & p_{010} & 0 & p_{011} \\ 0 & y & 0 & \frac{p_{011}}{p_{010}}y \end{array}\right],$$

where $x = \frac{p_{100}p_{111}\cdot d_{2,1}}{p_{101}d_{2,1} - p_{011}d_{1,1}}$ and $y = \frac{p_{010}p_{111}\cdot d_{1,1}}{p_{011}d_{1,1} - p_{101}d_{2,1}}$. Since the signs of $d_{2,1}$ and $d_{1,1}$ are different this expression for P is non-negative rank three, hence $P \in \mathcal{M}_{3,3}$. The denominator of x

and y is non-zero, provided that $d_{2,1}$ or $d_{1,1}$ is non-zero. If some determinant vanishes, a non-negative rank three decomposition is obtained from the rank one tensor of that face plus the non-negative rank two representation of the opposite face. Note that x and y are not both non-negative for $P \notin \mathrm{RBM}_{3,2}$, by Figure 7.4d, there is no way to rotate or reflect the cube such that determinants $d_{2,1}$ and $d_{2,2}$ have opposite sign. □

Lemma 7.12. *We have the containment of statistical models* $\mathcal{M}_{3,3} \subseteq \overline{\mathrm{RBM}_{3,2}}$.

Proof. Consider a distribution $P + Q \in \mathcal{M}_{3,3}$ where P is non-negative rank two, Q is non-negative rank one, and no entries of P or Q vanish. Up to swapping values 0 and 1 in one index, P being non-negative rank two means it satisfies the six binomial inequalities in Equation (7.3). Equivalently, its determinants $d_{i,j}$ from Equation (7.7) have sign pattern $(+,+,+,+,+,+)$, meaning that $d_{i,j} \geq 0$ for all i, j. We assume for contradiction that $P + Q \notin \mathrm{RBM}_{3,2}$. This means $P + Q$ has three "$-$" in its sign pattern, $d_{i,j} < 0$ for these pairs i, j. After adding tensor Q, *three* determinants have swapped sign: $d_{1,h}$, $d_{2,h}$, $d_{3,h}$ for $h = 0$ or 1.

Take non-negative vectors $a, b, c \in \mathbb{R}^2_{\geq 0}$ such that $q_{ijk} = a_i b_j c_k$. Assume that the determinant $d_{3,h}$ of $P+Q$ is negative: $(p_{00h}+a_0b_0c_h)(p_{11h}+a_1b_1c_h)-(p_{01h}+a_0b_1c_h)(p_{10h}+a_1b_0c_h) < 0$. Multiplying this expression out, and using $p_{00h}p_{11h} \geq p_{01h}p_{10h}$, gives

$$p_{00h}a_1b_1 + p_{11h}a_0b_0 < p_{10h}a_0b_1 + p_{01h}a_1b_0. \tag{7.8}$$

Hence either $p_{00h}b_1 < p_{01h}b_0$ or $p_{11h}b_0 < p_{10h}b_1$ must hold, and likewise either $p_{00h}a_1 < p_{10h}a_0$ or $p_{11h}a_0 < p_{01h}a_1$ must hold. Furthermore, rearranging Equation (7.8) yields

$$\frac{1}{p_{00h}}(p_{00h}a_1 - p_{10h}a_0)(p_{00h}b_1 - p_{01h}b_0) + \left(p_{11h} - \frac{p_{10h}p_{01h}}{p_{00h}}\right) a_0b_0 < 0.$$

Since the last term is non-negative, this implies that $\frac{1}{p_{00h}}(p_{00h}a_1-p_{10h}a_0)(p_{00h}b_1-p_{01h}b_0) < 0$, hence exactly one of $p_{00h}a_1 < p_{10h}a_0$ and $p_{00h}b_1 < p_{01h}b_0$ holds. Similarly, Equation (7.8) yields

$$\frac{1}{p_{11h}}(p_{11h}a_0 - p_{01h}a_1)(p_{11h}b_0 - p_{10h}b_1) + \left(p_{00h} - \frac{p_{01h}p_{10h}}{p_{11h}}\right) a_1b_1 < 0,$$

implying exactly one of $p_{11h}a_0 < p_{01h}a_1$ and $p_{11h}b_0 < p_{10h}b_1$ holds. Repeating the above for determinants $d_{2,h}$ and $d_{1,h}$ gives the following $2^3 = 8$ options:

$$\begin{aligned}
I^{(1)}_{ab} &= \{p_{00h}b_1 < p_{01h}b_0, \quad p_{11h}a_0 < p_{01h}a_1\}, & I^{(2)}_{ab} &= \{p_{11h}b_0 < p_{10h}b_1, \quad p_{00h}a_1 < p_{10h}a_0\}, \\
I^{(1)}_{ac} &= \{p_{0h0}a_1 < p_{1h0}a_0, \quad p_{1h1}c_0 < p_{1h0}c_1\}, & I^{(2)}_{ac} &= \{p_{1h1}a_0 < p_{0h1}a_1, \quad p_{0h0}c_1 < p_{0h1}c_0\}, \\
I^{(1)}_{bc} &= \{p_{h00}c_1 < p_{h01}c_0, \quad p_{h11}b_0 < p_{h01}b_1\}, & I^{(2)}_{bc} &= \{p_{h11}c_0 < p_{h10}c_1, \quad p_{h00}b_1 < p_{h10}b_0\}.
\end{aligned}$$

If either inequality from $I^{(1)}_{ab}$ holds, the inequalities of $I^{(2)}_{ab}$ cannot hold, and likewise for I_{ac} and I_{bc}. To conclude the proof, we derive a contradiction from these options.

Let $h = 0$. Assume the inequalities in $I^{(1)}_{ab}$ hold. Then one of the inequalities from $I^{(2)}_{bc}$ holds, hence $I^{(1)}_{bc}$ cannot hold. If $I^{(1)}_{ac}$ also holds, combining $p_{110}a_0 < p_{010}a_1$ from $I^{(1)}_{ab}$ with $p_{000}a_1 < p_{100}a_0$ from $I^{(1)}_{ac}$ gives $p_{110}p_{000} < p_{010}p_{100}$, contradicting the hypothesis that P satisfies the inequalities in (7.3). If $I^{(2)}_{ac}$ holds, combining inequalities involving c gives $p_{000}p_{011} < p_{001}p_{010}$, also a contradiction. Likewise, if $I^{(2)}_{ab}$ holds then $I^{(1)}_{ac}$ must hold. If $I^{(1)}_{bc}$ also holds, combining the inequalities involving c implies $p_{101}p_{000} < p_{100}p_{001}$, a contradiction. If $I^{(2)}_{bc}$ holds, combining inequalities involving b gives $p_{110}p_{000} < p_{100}p_{010}$, also a contradiction. The case $h = 1$ follows by analogous reasoning.

This shows that an open dense subset of $\mathcal{M}_{3,3}$ is contained in $\mathrm{RBM}_{3,2}$. It remains to consider when P or Q has some vanishing entry. Such cases are in the closure of the above, hence they lie in the closure of $\mathrm{RBM}_{3,2}$. □

Proof of Theorem 7.5. Lemma 7.11, and the fact that $\mathcal{M}_{3,3}$ is closed, implies the inclusion of closures $\overline{\mathrm{RBM}_{3,2}} \subseteq \mathcal{M}_{3,3}$. Combining with the inclusion in Lemma 7.12 gives $\mathcal{M}_{3,3} \subseteq \overline{\mathrm{RBM}_{3,2}} \subseteq \mathcal{M}_{3,3}$, hence the two models are equal up to closure. Theorem 7.4 implies that $\mathrm{RBM}_{3,2}$ is closed on the interior of the simplex, hence we have $\mathcal{M}_{3,3} = \mathrm{RBM}_{3,2}$ on the interior of the simplex. □

7.3 Maximum likelihood estimation

In this section I give a closed-form formula for maximum likelihood estimation (MLE) to the statistical model of consideration in this chapter,

$$\mathcal{M} = \mathcal{M}_{3,3} = \overline{\mathrm{RBM}_{3,2}}.$$

The MLE is found by giving a description of the boundary of the model, which is a union of exponential families. The MLE to each boundary piece can then be given in closed form, and the MLE to the model as a whole is given by a minimization over the boundary pieces.

Consider an empirical probability distribution coming from some data. The maximum likelihood estimation problem is the distribution in a statistical model with smallest Kullback-Leibler (KL) divergence to the data distribution. The KL divergence from P to Q is defined as $D(P\|Q) := \sum_x p_x \log \frac{p_x}{q_x}$, where x ranges over the possible states of P and Q. This is zero if and only if $P = Q$ and it is set to $+\infty$ when $\mathrm{supp}(P) \not\subseteq \mathrm{supp}(Q)$. The distributions in the closure of a model that minimize the KL divergence are called *reverse information projections* [52]. In general they are not unique, but they are unique for exponential families.

The boundary of the model

We saw that $\mathcal{M}$ is defined by the binomial inequalities in Theorem 7.4. Setting the inequalities in Theorem 7.4 to equalities gives the Zariski closure of the boundary of the model. This is also the Zariski closure of the boundary of the model $\mathcal{M}_{3,2}$ from [6].

Proposition 7.13. *Distributions on the boundary of $\mathcal{M}$ are $2 \times 2 \times 2$ tensors with a 2×2 slice of rank ≤ 1.*

The following is a converse result. It implies that $\mathrm{RBM}_{3,2}$ is closed on the interior of the simplex. Furthermore, within the simplex of probability distributions, the Zariski closure of the boundary is contained in the closure of the model. This result (which fails for $\mathcal{M}_{3,2}$) is useful for simplifying the study of maximum likelihood estimation for the model.

Lemma 7.14. *Every distribution of three binary random variables with a rank one 2×2 slice, and strictly positive entries, lies in the models $\mathrm{RBM}_{3,2}$ and $\mathcal{M}_{3,3}$.*

Proof. As in the proof of Lemma 7.11, if the determinant of a distribution P vanishes, a non-negative rank three decomposition is obtained from the rank one tensor of that slice plus the non-negative rank two representation of the opposite slice. This proves the result for $\mathcal{M}_{3,3}$.

It remains to build a decomposition of P as $(Q + R)(S + T)$ where Q, R, S, T are rank one non-negative $2 \times 2 \times 2$ tensors, and multiplication is entry-wise, as in Equation (7.2). Assume without loss of generality that $d_{3,1} = 0$. Let Q be the rank one tensor with slices Q_{**1} and P_{**1} equal, where Q_{**0} is set to be the smallest scalar multiple of P_{**1} that zeros out an entry of P_{**0}. The notation P_{**0} refers to the slice P_{ij0} for $i, j \in \{0, 1\}$. Then $P - Q$ consists of at most three non-zero entries. Let R be the tensor which satisfies $r_{ijk} = p_{ijk} - q_{ijk}$ for two of the three entries at which $P \neq Q$. Since these two entries can be chosen to be Hamming neighbors, R is rank one. And since $P - Q$ is non-negative, R is non-negative. There remains at most one entry where equality $P = Q + R$ does not hold: let i, j, k be such that $p_{ijk} > q_{ijk} + r_{ijk}$. Let S be the all ones tensor, and let T be the tensor with just one non-zero entry, $t_{ijk} = \frac{p_{ijk}}{q_{ijk}+r_{ijk}} - 1$. Then T is also non-negative and rank one, and $P = (Q + R)(S + T)$ as required. □

In the log-probability coordinates, the boundary of $\mathcal{M}$ is the union of hyperplanes:

$$\begin{array}{ll} \mathcal{L}_{1,0} = \{l_{000} + l_{011} - l_{001} - l_{010} = 0\}, & \mathcal{L}_{1,1} = \{l_{100} + l_{111} - l_{101} - l_{110} = 0\}, \\ \mathcal{L}_{2,0} = \{l_{000} + l_{101} - l_{001} - l_{100} = 0\}, & \mathcal{L}_{2,1} = \{l_{010} + l_{111} - l_{011} - l_{110} = 0\}, \\ \mathcal{L}_{3,0} = \{l_{000} + l_{110} - l_{010} - l_{100} = 0\}, & \mathcal{L}_{3,1} = \{l_{001} + l_{111} - l_{011} - l_{101} = 0\}. \end{array} \tag{7.9}$$

The intersection poset of a hyperplane arrangement is the set of all intersections of hyperplanes, ordered by reverse inclusion [173]. In Figure 7.3, I give the intersection poset of the pieces of the boundary of $\mathcal{M}$. The lowest node is the ambient space $\mathbb{R}^8$. At the first level are the six boundary pieces. At the second level are the 15 pairwise intersections. The enlarged nodes are intersections of the form $\mathcal{L}_{i,0} \cap \mathcal{L}_{i,1}$. The third level contains the 11 distinct codimension three intersections. The enlarged nodes are codimension three, despite being intersections of four hyperplanes, which highlight the non-generic structure of the hyperplane arrangement. The top intersection corresponds to the independence model. The nodes are labeled with their Möbius function value.

We can study the combinatorics of the arrangement using its characteristic polynomial $\chi(t) = \sum_f \mu(f) t^{\dim(f)}$. The summation is taken over all flats in the arrangement, where μ is the Möbius function. Evaluating the characteristic polynomial at $t = -1$ gives the number of full dimensional regions of the ambient space defined by the arrangement (see [173])

$$|\chi(-1)| = 46.$$

For comparison, a generic four dimensional central arrangement of six hyperplanes defines 52 regions. Ours is a central arrangement (the origin is in all hyperplanes) hence all 46 regions are unbounded cones. Of the 46 regions, the model $\mathcal{M}$ occupies 44 and the smaller model $\mathcal{M}_{3,2}$ occupies four.

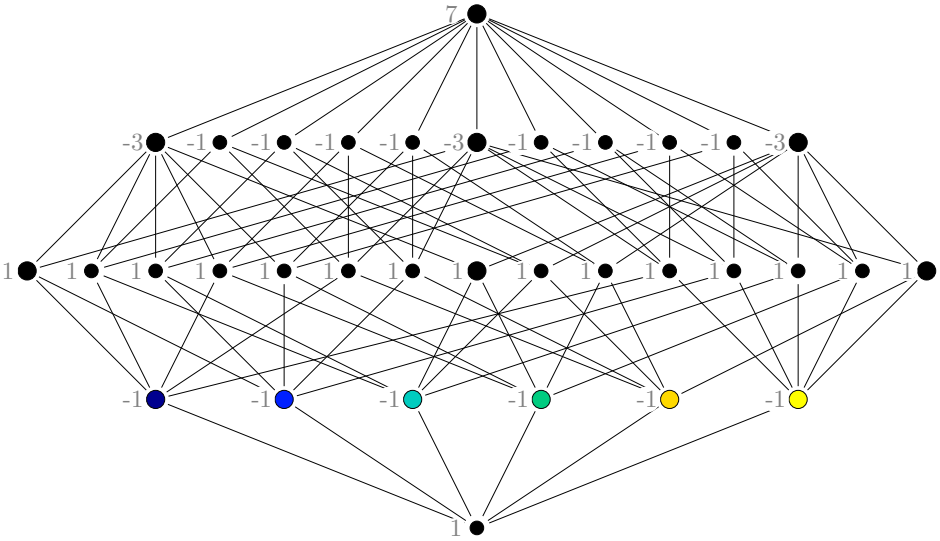


Figure 7.3: Intersection poset of the boundary pieces of the statistical model $\mathcal{M}$.

Since the six boundary pieces in Equation (7.9) are linear equations in log probability space, they define exponential families. For instance, the exponential family $\mathcal{L}_{1,0}$ consists of all distributions whose log-probabilities have a vanishing inner product with the vector $\begin{bmatrix} 1 & -1 & -1 & 1 & 0 & 0 & 0 & 0 \end{bmatrix}^T$. Since intersections of exponential families are exponential families, each element in the intersection poset in Figure 7.3 is also an exponential family.

Reversed information projections

To study the maximum likelihood estimation problem for the model $\mathcal{M}$, we first find the reversed information projections (rI-projections) to each boundary piece of the model. We only need to consider projections onto the six boundary pieces, not onto the entire intersection poset, by Proposition 7.13. In contrast, for the model $\mathcal{M}_{3,2}$ we have to consider projections to the whole poset of boundary pieces, see [5]. For a distribution $P \in \Delta_7 \backslash \mathcal{M}$,

the rI-projection to each boundary piece will lie in the model, and there is at most one projection point in each boundary piece. Taking the projection that minimizes divergence, over the six boundary pieces, gives the rI-projection to the whole model.

Let $\mathcal{P}_{i,j}$ be the toric hypersurface in the simplex obtained by exponentiating the hyperplane $\mathcal{L}_{i,j}$ in log-probability space and normalizing. The following proposition gives the maximum likelihood estimation for that toric model.

Proposition 7.15. *The unique rI-projection of $P \in \Delta_7$ onto $\mathcal{P}_{1,0}$ is found by taking the best rank one approximation in the slice P_{0**}, and leaving the other slice unchanged. In symbols, its entries are*

$$P(X_2|X_1)P(X_3|X_1)P(X_1), \text{ for } X_1 = 0 \quad \text{and} \quad P(X), \text{ for } X_1 = 1,$$

where X is the random variable on state space $\{0,1\}^3$ and X_i is its ith coordinate. The divergence from P to $\mathcal{P}_{1,0}$ is

$$D(P\|\mathcal{P}_{1,0}) = P(X_1 = 0) \cdot I_P(X_2; X_3|X_1 = 0),$$

where $I_P(X_2; X_3|X_1 = 0) = D(P(X_2X_3|X_1 = 0)\|P(X_2|X_1 = 0)P(X_3|X_1 = 0))$ is the conditional mutual information of the two variables X_2 and X_3, given $X_1 = 0$. The rI-projections to the five other pieces follow analogously.

Proof. This follows by applying [128, Lemma 3.2] to the exponential family described in Proposition 7.13 and using the fact that the rI-projection of a distribution to an independence model is given by the product of its marginals. □

The rI-projection to the entire model is the boundary projection with smallest divergence value. It has divergence

$$D(P\|\mathcal{M}) = \min_{i=1,2,3,\ j=0,1} D(P\|\mathcal{P}_{i,j}).$$

The rI-projection of any P to an exponential family is unique, so there are at most six rI-projections to $\mathcal{M}$. The distributions whose rI-projections to $\mathcal{P}_{1,0}$ coincide are those with the same values p_{1jk}, $j,k \in \{0,1\}$ and fixed marginals on p_{0jk}, $j,k \in \{0,1\}$.

Remark 7.16. *For the $\mathcal{M}_{3,3}$ and $\mathrm{RBM}_{3,2}$ parametrizations of $\mathcal{M}$, each rI-projection may be realized by several distinct choices of the parameters: there are several choices of parameters associated with each local maximizer of the likelihood function.*

Divergence maximizers

The maximum divergence to a statistical model is a measure of the representational power of that model. Here I describe the distributions with the largest divergence to the model $\mathcal{M}$.

Proposition 7.17. *The maximum divergence to $\mathcal{M}$ is $\frac{1}{2}\log 2$. The maximizers are the uniform distributions on the odd or even parity states, $u^+ := \frac{1}{4}(\delta_{000} + \delta_{011} + \delta_{101} + \delta_{110})$ and $u^- := \frac{1}{4}(\delta_{001} + \delta_{010} + \delta_{100} + \delta_{111})$. There are six rI-projections of u^+, one in each boundary piece:*

$$\begin{aligned}
u^+_{\mathcal{P}_{1,0}} &= \frac{1}{8}(\delta_{000} + \delta_{001} + \delta_{010} + \delta_{011}) + \frac{1}{4}(\delta_{101} + \delta_{110}) \\
u^+_{\mathcal{P}_{1,1}} &= \frac{1}{8}(\delta_{100} + \delta_{101} + \delta_{110} + \delta_{111}) + \frac{1}{4}(\delta_{011} + \delta_{000}) \\
u^+_{\mathcal{P}_{2,0}} &= \frac{1}{8}(\delta_{000} + \delta_{001} + \delta_{100} + \delta_{101}) + \frac{1}{4}(\delta_{011} + \delta_{110}) \\
u^+_{\mathcal{P}_{2,1}} &= \frac{1}{8}(\delta_{010} + \delta_{011} + \delta_{110} + \delta_{111}) + \frac{1}{4}(\delta_{000} + \delta_{101}) \\
u^+_{\mathcal{P}_{3,0}} &= \frac{1}{8}(\delta_{000} + \delta_{010} + \delta_{100} + \delta_{110}) + \frac{1}{4}(\delta_{011} + \delta_{101}) \\
u^+_{\mathcal{P}_{3,1}} &= \frac{1}{8}(\delta_{001} + \delta_{011} + \delta_{101} + \delta_{111}) + \frac{1}{4}(\delta_{000} + \delta_{110}).
\end{aligned}$$

The projection points of u^- are given in a similar way.

Proof. Proposition 7.15 shows that the indicated distributions are the rI-projections of u^+ onto the individual boundary pieces of $\mathcal{M}$. There can be no more than six projection points and hence we have a complete list. The fact that $\frac{1}{2}\log 2$ is the maximum possible divergence to $\mathcal{M}$ follows from upper bounds for mixtures of products and RBMs given in [129]. Both u^+ and u^- attain this upper bound.

Now I show that u^+ and u^- are the only divergence maximizers. Assume without loss of generality that some maximizer P has an rI-projection onto $\mathcal{M}$ in $\mathcal{P}_{1,0}$. Then $D(P\|\mathcal{P}_{1,0}) = P(X_1 = 0)I_P(X_2; X_3|X_1 = 0) \leq D(P\|\mathcal{P}_{1,1}) = P(X_1 = 1)I_P(X_2; X_3|X_1 = 1) \leq (1 - P(X_1 = 0))\log 2$. The last inequality follows since, for two binary variables, the mutual information is maximized by a uniform distribution on strings of Hamming distance two (see [13]). The maximum value $\frac{1}{2}\log 2$ is attained only if $P(X_1 = 0) = P(X_1 = 1) = \frac{1}{2}$ and both $P(X_2X_3|X_1 = 0)$ and $P(X_2X_3|X_1 = 1)$ are uniform on pairs of Hamming distance two. If these two conditional distributions were equal, then $P \in \mathcal{M}$, and P is not a divergence maximizer. Hence the pairs are different. This shows that P is a uniform distribution on four strings of equal parity. □

Remark 7.18. *Proposition 7.17 shows that the upper bound on the maximum divergence to mixtures of products and RBMs from [129, Theorems 1 and 2] is tight in the case of $\mathcal{M}_{3,3}$ and $\mathrm{RBM}_{3,2}$. Moreover it shows that for a given data point, $\mathrm{RBM}_{3,2}$ can have up to 6 global maximizers of the likelihood, and that generically this will be the number of local maximizers.*

An interesting question is whether we can characterize the points in the probability simplex that project to the different boundary pieces of the model. That is, to provide a *decision boundary* separating the regions of the simplex that are closer to each part of the

model, with respect to the KL divergence. In this case, these decision boundaries are neither linear families nor exponential families.

7.4 Connection to triangulations

A positive distribution $P \in \Delta_7$ induces a triangulation of the three-cube. We can characterize the statistical model $\mathcal{M}$ in terms of triangulations, because the implicit description for $\mathcal{M}$ consists of binomial expressions that impose constraints on the associated triangulation. Membership of a distribution P in the model $\mathcal{M}$ depends on properties of the triangulation. We use the connection between model membership and triangulations to prove Corollary 7.7.

Consider a generic, strictly positive distribution $P \in \Delta_7$. Its tensor of log-probabilities $l_{ijk} = \log(p_{ijk})$ induces a triangulation of the three-cube, as follows. Assign the height l_{ijk} to each vertex $(i, j, k) \in \{0, 1\}^3$. Take the upper part of the convex hull (the upper hull) of the points (i, j, k, l_{ijk}) in four-dimensional space and project it back to the three-dimensional cube. The facets in the upper hull project to tetrahedra that triangulate the cube. Figure 7.4 illustrates different triangulations of the three-dimensional cube by showing how the triangulations restrict to certain faces of the cube.

Proposition 7.19. *The model $\mathcal{M}$ contains distributions with triangulations as in Figure 7.4b and 7.4c. Distributions with triangulations as in Figure 7.4a are a special case that lies in $\mathcal{M}_{3,2}$. Distributions in Figure 7.4d lie outside of $\mathcal{M}$.*

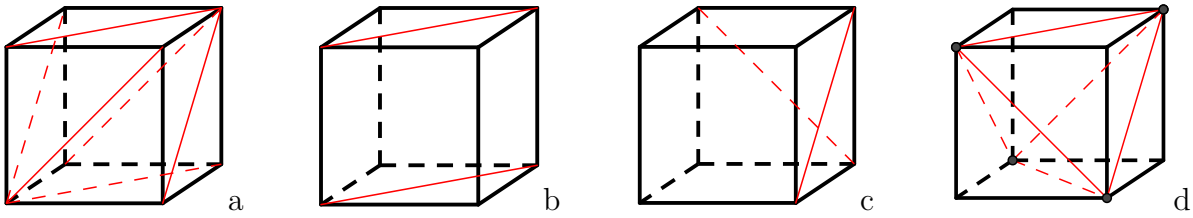


Figure 7.4: Membership in statistical models in terms of triangulations, see Theorem 7.19.

Rotating or reflecting the cubes in Figure 7.4 corresponds to relabeling indices of the distributions, and in particular does not affect membership in any of the statistical models we consider. Some of the 2^6 possible triangulations of the six faces do not come from a triangulation of the whole cube. The empty faces in Figure 7.4 can be triangulated in either of the two possible directions, provided that the triangulation of the faces is a restriction of a triangulation of the whole cube.

Proof of Proposition 7.19. There are 20 linear expressions in the coordinates l_{ijk} whose signs determine the triangulation, see [25, page 1325]. Six of these equations determine how the triangulation restricts to the faces of the cube. These are the logarithms of the binomial equations that define $\mathcal{M}$. Hence we can see whether $\exp(l_{ijk})$ lies in $\mathcal{M}$ by looking at how

the triangulation induced by the l_{ijk} restricts to the faces of the cube. The equations that define $\mathcal{M}_{3,2}$ and $\mathcal{M}_{3,1}$ are also of this form.

In the language of triangulations, being in $\mathcal{M}$ means we triangulate *at least one pair of opposite faces in the same direction*, as in Figure 7.4b. The condition for being in $\mathcal{M}_{3,2}$ is that every pair of opposite faces is triangulated in the same direction, with sign compatibility as in Figure 7.4a. Triangulations of distributions not in $\mathrm{RBM}_{3,2}$ *triangulate every pair of opposite faces in opposing directions*, as in Figure 7.4d. An alternate characterization of such triangulations is that every pair of adjacent faces is triangulated in a continuous way. If, conversely, a pair of adjacent faces is triangulated in a discontinuous way, as in Figure 7.4c, the distribution lies in $\mathcal{M}$. □

Proof of Corollary 7.7. I show that distributions with four modes restrict to the faces of the cube as shown in Figure 7.4d. Assume we have a distribution with four modes. Without loss of generality, the four numbers l_{000}, l_{011}, l_{101}, and l_{110} exceed the values of their neighbours. Consider a face of the cube, for example the face $\langle l_{000}, l_{001}, l_{010}, l_{011}\rangle$. Since $l_{000} \geq l_{001}$ and $l_{011} \geq l_{010}$, we have

$$l_{000} + l_{011} - l_{010} - l_{001} \geq 0,$$

which determines how the triangulation of (l_{ijk}) restricts to the face. Repeating for the other faces gives the triangulation of the faces shown in Figure 7.4d. Distributions on $\partial\Delta_7 \cap \mathrm{RBM}_{3,2}$ have at least two adjacent entries vanishing, by Equation (7.6). This excludes the possibility of having four modes. □

Visualizing the model

In this subsection, I explain how to draw the seven-dimensional model $\mathcal{M}$ using a three dimensional figure. In [162, Figure 3], a first attempt was made to visualize the model $\mathcal{M}$. I make use of the following change of basis (corresponding to the basis of characters) in the log-probabilities:

$$\begin{pmatrix} m_{\emptyset} \\ m_{\{3\}} \\ m_{\{2\}} \\ m_{\{2,3\}} \\ m_{\{1\}} \\ m_{\{1,3\}} \\ m_{\{1,2\}} \\ m_{\{1,2,3\}} \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 \\ 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 \\ 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 \\ 1 & -1 & -1 & 1 & -1 & 1 & 1 & -1 \end{pmatrix} \begin{pmatrix} l_{000} \\ l_{001} \\ l_{010} \\ l_{011} \\ l_{100} \\ l_{101} \\ l_{110} \\ l_{111} \end{pmatrix}.$$

The boundary pieces of the model can be written in terms of just four of these coordinates:

$$\begin{aligned} \mathcal{L}_{1,0} &= \{m_{\{2,3\}} + m_{\{1,2,3\}} = 0\}, \quad \mathcal{L}_{1,1} = \{m_{\{2,3\}} - m_{\{1,2,3\}} = 0\}, \\ \mathcal{L}_{2,0} &= \{m_{\{1,3\}} + m_{\{1,2,3\}} = 0\}, \quad \mathcal{L}_{2,1} = \{m_{\{1,3\}} - m_{\{1,2,3\}} = 0\}, \\ \mathcal{L}_{3,0} &= \{m_{\{1,2\}} + m_{\{1,2,3\}} = 0\}, \quad \mathcal{L}_{3,1} = \{m_{\{1,2\}} - m_{\{1,2,3\}} = 0\}. \end{aligned}$$

Hence it suffices to visualize the combinations of coordinates $(m_{\{1,2\}}, m_{\{1,3\}}, m_{\{2,3\}}, m_{\{1,2,3\}})$ that lie in the model. Furthermore, if a vector satisfies the inequalities above, then so does any scalar multiple, so we need only consider vectors $(m_{\{1,2\}}, m_{\{1,3\}}, m_{\{2,3\}}, m_{\{1,2,3\}})$ of norm one. The value of $m_{\{1,2,3\}}$ can be found up to sign from the other three coordinates. We can draw the model in coordinates

$$(\overline{m}_{\{1,2\}}, \overline{m}_{\{1,3\}}, \overline{m}_{\{2,3\}}) = \frac{(m_{\{1,2\}}, m_{\{1,3\}}, m_{\{2,3\}})}{\|(m_{\{1,2\}}, m_{\{1,3\}}, m_{\{2,3\}}, m_{\{1,2,3\}})\|},$$

where $\|\cdot\|$ denotes the Euclidean norm, with separate panels for the different signs of $m_{\{1,2,3\}}$. Figure 7.5 shows pieces $\mathcal{L}_{1,0}$ and $\mathcal{L}_{1,1}$. The set $\mathcal{L}_{1,0}$ is in dark blue, and $\mathcal{L}_{1,1}$ is in light blue. The points enclosed by the surface correspond to distributions in the complement of the two basic semi-algebraic sets of $\mathrm{RBM}_{3,2}$ enclosed by $\mathcal{L}_{1,0}$ and $\mathcal{L}_{1,1}$. The black line is $\{m_{\{2,3\}} = m_{\{1,2,3\}} = 0\}$, along which $\mathcal{L}_{1,0}$ and $\mathcal{L}_{1,1}$ meet. The non-linearity of the surfaces is due to the normalization. The whole model is shown in Figure 7.6. From this picture, we also obtain a visualization of $\mathcal{M}_{3,2}$. Within each orthant, the part of the sphere outside all three surfaces is a triangular bipyramid. Four of these bipyramids make up the model $\mathcal{M}_{3,2}$.

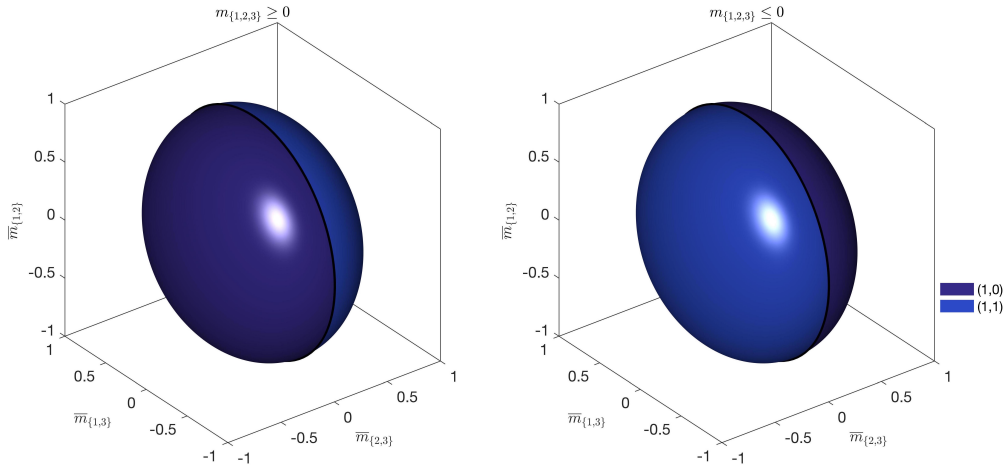


Figure 7.5: Two boundary pieces of the statistical model $\mathcal{M}$.

Outlook

In this chapter, I proved the rather surprising fact that a mixture of products and a product of mixtures represent the same set of probability distributions. The semi-algebraic description allows the computation of maximum likelihood estimates and divergence maximizers, both of which appear quite difficult to obtain via other methods.

The natural next step is to extend the analysis to larger models. However, the description for larger models involves complicated equality constraints. For example, in [53] the Zariski

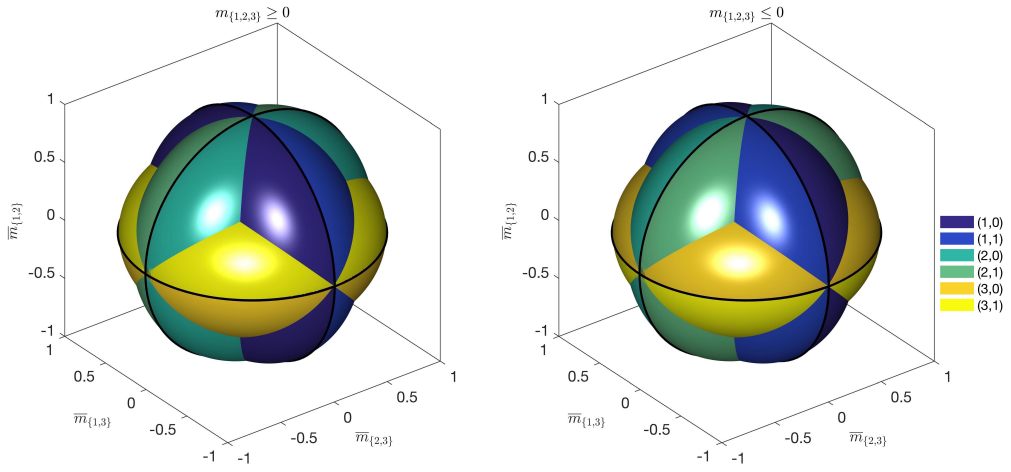


Figure 7.6: The statistical model $\mathcal{M}$ is the space inside the three-sphere and outside any of the blue, green, or yellow surfaces, the six boundary pieces of the model.

closure of the model $\text{RBM}_{4,2}$ is found. It is the zero set of a single degree 110 polynomial with at least 17,214,912 terms. The binomial inequalities we obtain here are more tractable. In light of this, it appears natural to consider approximate descriptions of larger RBM models in terms of inequality constraints only. A relaxation of larger statistical models, given in terms of inequalities only, would provide lower bounds on the maximal divergence and the minimal size of universal approximators.

In [6] the authors show that the model $\mathcal{M}_{d,2}$ consists of supermodular distributions with flattening rank at most two. Distributions in larger RBM models are Hadamard products of non-negative tensors of rank at most two (products of tensors proportional to distributions in $\mathcal{M}_{d,2}$). Ignoring the equations, we have the set of supermodular tensors, which consists of basic semi-algebraic sets satisfying binomial quadratic inequalities as in Equation (7.3). Hence the algebraic boundary of Hadamard products of supermodular tensors is again a union of exponential families, for which we may hope to obtain maximum likelihood estimates in closed form.

Chapter 8

Learning paths from signature tensors

In many applied contexts, tensors can be used to encode features of geometric data. The tensors are then used as the input to algorithms aimed at classifying and understanding the original data. The signature is collection of tensors that encodes a path. For a path $\psi : [0, 1] \to \mathbb{R}^n$, the signature is a formal series of tensors

$$\sigma(\psi) = \sum_{d=1}^{\infty} \sigma^{(d)}(\psi),$$

whose dth term is an order d tensor in $\mathbb{R}^{n\times\cdots\times n}$ with entries that are iterated integrals of ψ:

$$(\sigma^{(d)}(\psi))_{i_1\cdots i_d} = \int_0^1 \cdots \left(\int_0^{t_3} \left(\int_0^{t_2} \mathrm{d}\psi_{i_1}(t_1) \right) \mathrm{d}\psi_{i_2}(t_2) \right) \cdots \mathrm{d}\psi_{i_d}(t_d). \tag{8.1}$$

Assume that the path has coordinates that are piecewise continuous differentiable functions. The first tensor in the signature is a vector in $\mathbb{R}^n$. Evaluating Equation (8.1) for $d = 1$ shows that the first signature tensor is the increment of the path, $\sigma^{(1)}(\psi) = \psi(1) - \psi(0)$. The second tensor in the signature is a matrix in $\mathbb{R}^{n\times n}$. Evaluating Equation (8.1) for $d = 2$ shows that $\sigma^{(2)}(\psi) = \frac{1}{2}(\psi(1) - \psi(0))^{\otimes 2} + Q$, where Q is skew-symmetric. The entry q_{ij} is the *Lévy area* of the projection of the path ψ onto the plane indexed by i and j, the signed area between the planar path and the segment connecting its endpoints. The dth tensor in the sequence gives, as d increases, a finer encoding of the path, provided that the path is not a loop.

Signature tensors were introduced in [48] and are important in the theory of rough paths in stochastic analysis [71, 113]. A path is determined by its infinite encoding as a signature, up to a natural equivalence [47, 79]. There has been extensive interest in recovering a path from its signature [114, 115]. This has applications to time series data, including temporal medical data [60, 94, 95]. Many recovery results focus on recovery of a path from its infinite signature. The perspective of focusing on a single tensor in the signature was introduced in [7].

In this chapter, I study the problem of recovering a low-complexity path from its signature tensor of order three, the third term in the signature. I consider paths whose coordinates can be written as linear combinations of functions in a fixed dictionary. The family of piecewise linear paths is a main example. In the setting of a fixed dictionary, path recovery is closely related to the study of tensors under congruence action by the special linear group. I study conditions under which the stabilizer under congruence action is finite or trivial, which leads to identifiability results on recovering a path from the third order signature tensor, including a proof of part of [7, Conjecture 6.10]. I give a numerical optimization algorithm for path recovery, which gives accurate and unique recovery of low-complexity paths, and I also address the problem of finding the shortest path with given third signature tensor. This chapter is based on joint work with Max Pfeffer and Bernd Sturmfels, published in the SIAM Journal on Matrix Analysis and Applications [141].

8.1 Tensors under congruence

In this section I consider the congruence action for tensors, in which a matrix $A \in GL_n$ acts on a tensor X of format $n \times \cdots \times n$ (d times), via

$$X \mapsto [\![X; A, \ldots, A]\!]. \tag{8.2}$$

The entries of the transformed tensor are

$$[\![X; A, \ldots, A]\!]_{i_1 \ldots i_d} = \sum_{j_1, \ldots, j_d = 1}^{n} x_{j_1 \ldots j_d} a_{i_1 j_1} \cdots a_{i_d j_d}.$$

I give a necessary condition for a tensor to have trivial stabilizer under congruence action, and a Jacobian criterion that is sufficient for a tensor to have finite stabilizer under congruence. I also discuss extensions of the congruence action to rectangular matrices, and study the identifiability of recovering a rectangular matrix $A \in \mathbb{R}^{n \times m}$ from the tensor matrix product $[\![X; A, \ldots, A]\!]$.

So far in this thesis, we have seen the change of basis action for a tensor of format $n_1 \times \cdots \times n_d$, in which a product of matrices $(A^{(1)}, \ldots, A^{(d)}) \in GL_{n_1} \times \cdots \times GL_{n_d}$ acts on a tensor X via $X \mapsto [\![X; A^{(1)}, \ldots, A^{(d)}]\!]$. For symmetric tensors X, corresponding to polynomials f via Example 1.7, we have also seen the symmetric chance of basis $f(z) \mapsto f(A \cdot z)$, for $A \in GL_n$. For tensors, while there is a literature on the above two change of basis actions, much less is known about the congruence action. We note the relation between the congruence action in Equation (8.2) and the Tucker decomposition from Equation (1.10). The difference here is that the core tensor is fixed, and the tensor is multiplied on each side by the same matrix.

The *stabilizer* of X under the group action is the subgroup of matrices $A \in GL_n$ that satisfy $[\![X; A, \ldots, A]\!] = X$. It can be considered in the space of real or complex invertible matrices. The stabilizer is defined by a system of polynomial equations of degree d in the

entries of A. Matrices ηI with $\eta^d = 1$ are always among the solutions. It is an open problem to characterize the tensors X of format $n \times \cdots \times n$ (d times) whose stabilizer under congruence is non-trivial, i.e. strictly contains $\{\eta I : \eta^k = 1\}$. The stabilizer of a matrix under congruence is always infinite, see [141, Proposition 5.1]. The stabilizer of a tensor will be important for the identifiability of path recovery from the family of paths whose dictionary has signature tensor X. I introduce an important notion for stabilizers under congruence.

Definition 8.1 (Symmetrically concise)**.** *A tensor X in $V^{\otimes d}$ is symmetrically concise if there is no subspace $W \subsetneq V$ such that $X \in W^{\otimes k}$. We can equivalently define symmetrically concise in terms of flattenings. For $\dim(V) = n$, the tensor X has n^d entries and d principal flattenings, matrices of format $n \times n^{d-1}$. Concatenate the d flattening matrices to form a single matrix of format $n \times dn^{d-1}$. The tensor is symmetrically concise if this matrix has full rank n.*

The reason for the name symmetrically concise comes from the definition in terms of flattenings. A tensor X of format $n \times \cdots \times n$ (d times) is called *concise* if it has flattening ranks $(n, \ldots, n)$ [177]. For symmetric tensors, concise and symmetrically concise are equivalent, because the $n \times dn^{d-1}$ matrix consists of d identical blocks of format $n \times n^{d-1}$. However, symmetrically concise is weaker than concise for non-symmetric tensors. For example, the $3 \times 3 \times 3$ basis tensor $X = e_1 \otimes e_2 \otimes e_3$ is symmetrically concise but not concise: there exist subspaces $W_i \subsetneq \mathbb{K}^3$ with $X \in W_1 \otimes W_2 \otimes W_3$, but we cannot find the same subspace $W \subsetneq \mathbb{K}^3$ across all modes such that $X \in W^{\otimes 3}$.

We can also define symmetrically concise from a decomposition into rank one terms. Consider a decomposition of a tensor X of rank r as a sum of r rank one terms. This is called a minimal decomposition, since a minimal number of rank one terms have been used. A tensor is symmetrically concise if the dr vectors in any minimal decomposition span the ambient space $\mathbb{K}^n$.

Proposition 8.2. *Let X be a tensor that is not symmetrically concise. Then the stabilizer of X under the congruence action is non-trivial.*

Proof. Let X have format $n \times \cdots \times n$ (d times). Since X is not symmetrically concise, there exists a vector $v \in \mathbb{K}^n$ of norm one such that $v^\mathsf{T} X^{(i)} = 0$ for all $i = 1, \ldots, d$, where $X^{(i)}$ denotes the ith principal flattening of X. Hence $[\![X; I + vv^\mathsf{T}, \ldots, I + vv^\mathsf{T}]\!] = X$, and the invertible matrix $I + vv^\mathsf{T}$ is in the stabilizer of X. □

Tensors with trivial stabilizer are symmetrically concise but not always concise:

Example 8.3. *Consider the rank-one tensor $X = e_1 \otimes e_2 \otimes (e_1 + e_2)$. Each 2×4 flattening matrix of X is rank-deficient. This means that X has flattening ranks $(1, 1, 1)$, so X is not concise. However, the 2×12 matrix we obtain by concatenating the three flattening matrices has full rank, hence the tensor X is symmetrically concise. The stabilizer of X is directly computed to be trivial.*

I next derive a Jacobian criterion which gives a sufficient condition for the stabilizer of a tensor under the congruence action to be finite. For notational simplicity I state the criterion only for order three tensors. The Jacobian $\nabla f(A) \in \mathbb{K}^{n^3 \times n^2}$ of the function $f(A) = [\![X; A, A, A]\!]$ has entries:

$$\nabla f(A)_{(i,j,k),(u,v)} = \frac{\partial f_{ijk}}{\partial a_{uv}} = \sum_{\alpha,\beta} (x_{v\alpha\beta}\delta_{ui}a_{j\alpha}a_{k\beta} + x_{\alpha v\beta}\delta_{uj}a_{i\alpha}a_{k\beta} + x_{\alpha\beta v}\delta_{uk}a_{i\alpha}a_{j\beta}),$$

where δ_{ij} is the Kronecker delta. The entries of the Jacobian at $A = I$ are

$$\nabla f(I)_{(i,j,k),(u,v)} = x_{vjk}\delta_{ui} + x_{ivk}\delta_{uj} + x_{ijv}\delta_{uk}. \tag{8.3}$$

Consider the $n^2 \times n^2$ submatrix J_1 of the Jacobian obtained by setting $k = 1$ in Equation (8.3). The entry of J_1 in row (i, j) and column (u, v) is the linear form

$$J_1((i,j),(u,v)) = \delta_{ui}x_{vj1} + \delta_{uj}x_{iv1} + \delta_{u1}x_{ijv}.$$

Proposition 8.4. *Let X be a tensor whose $n^2 \times n^2$ matrix J_1 as above is invertible. Then the stabilizer of X under the congruence action by GL_n is finite.*

Proof. The stabilizer of X under congruence is infinite when the map $f : A \mapsto [\![X; A, A, A]\!]$ has positive-dimensional fibers. If the matrix J_1 is invertible then the Jacobian ∇f has full rank at $A = I$. This implies that a connected component of the stabilizer consists of the single matrix I. Consider another connected component of the stabilizer, containing a matrix Z. Applying Z^{-1} to the component gives a connected component of the stabilizer containing I, which therefore must be the single matrix I. Hence all connected components are zero-dimensional, and the stabilizer is finite. □

The same conclusion is obtained from any maximal minor of the Jacobian in $\mathbb{K}^{n^3 \times n^2}$. Since the stabilizer of a matrix under congruence is infinite, computing the Jacobian in this case gives a matrix of format $n^2 \times n^2$ with vanishing determinant.

Multiplication by rectangular matrices

In this section, I relate the stabilizer of a tensor under congruence action to the recovery of the matrix A from the tensor $[\![X; A, \ldots, A]\!]$, where $X \in \mathbb{K}^{m \times \cdots \times m}$ and $A \in \mathbb{K}^{n \times m}$ and $m \leq n$. The tensor is called *identifiable* if the matrix A can be recovered up to scale. The tensor is called *algebraically identifiable* if A can be recovered up to a finite list of choices.

The following result compares minimal decompositions of smaller tensors with those of larger tensors in which they appear as a block. Any decomposition of the larger tensor is obtained from a decomposition of the smaller tensor by adding zeros.

Lemma 8.5. *Let X be a tensor of format $n \times \cdots \times n$ (d times), where $d \geq 3$ and all non-zero entries are in a block of format $m \times \cdots \times m$ (d times) for some $m \leq n$. Then any rank one term in a minimal decomposition of X is also zero outside of the block.*

Proof. Let $X = \sum_{l=1}^{r} v_l^{(1)} \otimes \cdots \otimes v_l^{(d)}$ be a minimal decomposition. Assume that a rank one term is non-zero outside of the block, i.e. the α coordinate of $v_l^{(j)}$ is non-zero for some l, j, and index α not in the block. The terms $v_l^{(1)} \otimes \ldots \otimes v_l^{(j)}(\alpha) \otimes \cdots \otimes v_l^{(d)}$ sum to zero. However, the order $d-1$ tensors in a minimal decomposition, resulting from removing the jth vector from each rank one term, are linearly independent, a contradiction. □

We note that Lemma 8.5 also appears as [105, Proposition 3.1.3.1]. We use this lemma to relate the stabilizer of a tensor X under the congruence action, to recovery of the matrix A from $[\![X; A, \ldots, A]\!]$.

Theorem 8.6. *Fix a symmetrically concise tensor X of format $m \times \cdots \times m$ (d times). Let $\mathrm{Stab}_{\mathbb{K}}(X)$ denote its stabilizer under the congruence action by matrices in GL_m with entries in $\mathbb{K}$. For any matrix $A \in \mathbb{K}^{n \times m}$ of rank $m \leq n$, we have*

$$\big\{B \in \mathbb{K}^{n\times m} : [\![X; A, \ldots, A]\!] = [\![X; B, \ldots, B]\!]\big\} = \big\{AZ : Z \in \mathrm{Stab}_{\mathbb{K}}(X)\big\}. \tag{8.4}$$

Proof. Suppose $[\![X; A, \ldots, A]\!] = [\![X; B, \ldots, B]\!]$. Let $\tilde{X}$ be the $n \times \cdots \times n$ tensor with entries

$$\tilde{x}_{i_1 \ldots i_d} = \begin{cases} x_{i_1 \ldots i_d} & \text{if } 1 \leq i_1, \ldots, i_d \leq m, \\ 0 & \text{otherwise.} \end{cases}$$

Let $\tilde{A}$ be an invertible $n \times n$ matrix whose first m columns are A, and likewise construct $\tilde{B}$. Then $[\![\tilde{X}; \tilde{A}, \ldots, \tilde{A}]\!] = [\![\tilde{X}; \tilde{B}, \ldots, \tilde{B}]\!]$. We multiply by $\tilde{A}^{-1}$ to get $\tilde{X} = [\![\tilde{X}; \tilde{Z}, \ldots, \tilde{Z}]\!]$ where $\tilde{Z} = \tilde{A}^{-1}\tilde{B}$ and the top-left $m \times m$ block of $\tilde{Z}$, denoted Z, satisfies $[\![X; Z, \ldots, Z]\!] = X$.

Let $\tilde{X} = \sum_{l=1}^{r} \tilde{X}_l$ be a minimal decomposition, where $\tilde{X}_l = \tilde{v}_l^{(1)} \otimes \cdots \otimes \tilde{v}_l^{(d)}$ with $\tilde{v}_l^{(j)} \in \mathbb{K}^m \times \{0\}^{n-m} \subseteq \mathbb{K}^n$. We obtain another minimal decomposition of $\tilde{X}$, by acting with $\tilde{Z}$, with rank one terms $[\![\tilde{X}_l; \tilde{Z}, \ldots, \tilde{Z}]\!] = (\tilde{Z}\tilde{v}_l{}^{(1)}) \otimes \cdots \otimes (\tilde{Z}\tilde{v}_l{}^{(d)})$. By Lemma 8.5, all minimal decompositions of $\tilde{X}$ come from those of X by adjoining zeros. This means that the $n-m$ row vectors in the $(n-m) \times m$ lower-left block of $\tilde{Z}$ have dot product zero with every vector appearing in a minimal decomposition of X. Since X is symmetrically concise, these row vectors must be zero. The identity $\tilde{B} = \tilde{A}\tilde{Z}$ now implies $B = AZ$. □

Corollary 8.7. *Let X be a symmetrically concise tensor of format $m \times \cdots \times m$ (d times) whose stabilizer under congruence by matrices in GL_m with entries in $\mathbb{K}$ has size s. Then, for any matrix $A \in \mathbb{K}^{n \times m}$ of rank m, there are s matrices in $\mathbb{K}^{n \times m}$ with*

$$[\![X; A, \ldots, A]\!] = [\![X; B, \ldots, B]\!].$$

Proof. Let B be a matrix in $\mathbb{K}^{n \times m}$ that satisfies $[\![X; A, \ldots, A]\!] = [\![X; B, \ldots, B]\!]$. By Theorem 8.6, we have $B = AZ$ where Z is in the stabilizer under the congruence action. If there are n choices for Z, then there are n choices for B. □

If X has trivial stabilizer under congruence, then it is already symmetrically concise by Proposition 8.2. We can thus simplify Corollary 8.7 as follows.

Corollary 8.8. *If X of format $m \times \cdots \times m$ (d times) has trivial stabilizer under congruence then rank m matrices $A \in \mathbb{K}^{n\times m}$ are identifiable from the tensor $[\![X; A, \ldots, A]\!]$.*

The following example illustrates why Theorem 8.6 and Corollary 8.7 fail when C is not symmetrically concise.

Example 8.9. *Fix the $2 \times 2 \times 2$ tensor $X = e_1 \otimes e_1 \otimes e_1$. Its stabilizer in GL_2 is*

$$Z = \begin{bmatrix} 1 & * \\ 0 & * \end{bmatrix},$$

where the $*$ entries can take any value in $\mathbb{K}$. Setting $m = 2, n = 3$, I also introduce

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \qquad \text{and} \qquad B = \begin{bmatrix} 1 & * \\ 0 & * \\ 0 & * \end{bmatrix}.$$

The left hand side of Equation (8.4) is the set of all matrices of the form B. This set strictly contains the right hand side of Equation (8.4), because not all matrices B are expressible as AZ for some Z. This happens because the last row of B has dot product zero with all vectors in the minimal decomposition of X, without being zero itself, i.e. the tensor X is not symmetrically concise.

8.2 Signature tensors

Consider a *dictionary* $\psi = (\psi_1, \ldots, \psi_m)$ of piecewise differentiable functions $\psi_i : [0, 1] \to \mathbb{R}$. In this section, I consider the signature tensors of paths whose coordinates can be written as a linear combination of functions in a fixed dictionary. This holds for many useful families of paths, such as piecewise linear paths with a bounded number of steps. As we will see, once the signature of the dictionary ψ is known, integrals no longer need to be computed: signature tensors of paths are obtained by tensor matrix multiplication.

The dictionary corresponds to a path in $\mathbb{R}^m$, also denoted ψ, whose ith coordinate is ψ_i. The image of the path ψ under the linear map given by $A \in \mathbb{R}^{n\times m}$ is the path $A\psi : [0, 1] \to \mathbb{R}^n$ given by

$$t \mapsto \left(\sum_{j=1}^{m} a_{1j}\psi_j(t), \ldots, \sum_{j=1}^{m} a_{dj}\psi_j(t) \right).$$

The following key lemma relates the linear transformation of a path to the induced linear transformation of its signature tensor. The proof follows directly from the iterated integrals in Equation (8.1), bearing in mind that integration is a linear operation.

Lemma 8.10. *The signature map is equivariant under linear transformations,*

$$\sigma(A\psi) = A(\sigma(\psi)).$$

The action of the linear map A on the signature $\sigma(\psi)$ is the congruence action

$$\sigma^{(d)}(A\psi) = [\![\sigma^{(d)}(\psi); A, \ldots, A]\!] \qquad \text{for } d = 1, 2, 3, \ldots \tag{8.5}$$

For $d = 1$ this is the matrix-vector product $\sigma^{(1)}(A\psi) = A \cdot \sigma^{(1)}(\psi)$. For $d = 2$ the rectangular matrix A acts on the square signature matrix via $\sigma^{(2)}(A\psi) = A \cdot \sigma^{(2)}(\psi) \cdot A^{\mathsf{T}}$. In this chapter, I focus on the third order signature $\sigma^{(3)}(\psi)$, because it is the lowest order signature for which paths can be recovered uniquely from third signatures under the assumption that they come from a dictionary.

The third order signature tensor of the path ψ is denoted by $C_\psi = \sigma^{(3)}(\psi) \in \mathbb{R}^{m\times m\times m}$. It has entries

$$c_{ijk} = \int_0^1 \left(\int_0^{t_3} \left(\int_0^{t_2} \mathrm{d}\psi_i(t_1) \right) \mathrm{d}\psi_j(t_2) \right) \mathrm{d}\psi_k(t_3) \quad \text{for all } 1 \le i, j, k \le m. \tag{8.6}$$

The first and second signature of a real path are determined by the third signature, provided the path is not a loop, just as any lower order signature tensor can be recovered up to scale from higher order signatures. This follows from the *shuffle relations* [7, Lemma 4.2]. Writing c_i and c_{ij} for the entries of the first and second signature respectively, we have the identities

$$c_i c_j = c_{ij} + c_{ji} \quad \text{and} \quad c_i c_{jk} = c_{ijk} + c_{jik} + c_{jki}. \tag{8.7}$$

Given a $n \times m$ matrix A, the third signature of the image path $A\psi$ in $\mathbb{R}^n$ is denoted by $\sigma^{(3)}(A)$, as shorthand for $\sigma^{(3)}(A\psi)$. Following Equation (8.5), this $n \times n \times n$ tensor is $[\![C_\psi; A, A, A]\!]$, with entries

$$[\![C_\psi; A, A, A]\!]_{\alpha\beta\gamma} = \sum_{i=1}^{m}\sum_{j=1}^{m}\sum_{k=1}^{m} c_{ijk} a_{\alpha i} a_{\beta j} a_{\gamma k}.$$

I next discuss some specific dictionaries and the paths they encode, starting with two dictionaries studied in [7]. The first dictionary is $\psi(t) = (t, t^2, \ldots, t^m)$. Multiplying this dictionary by matrices A of format $n \times m$ gives all *polynomial paths* of degree at most m that start at the origin in $\mathbb{R}^n$. The core tensor of ψ is denoted by C_{mono} to indicate the monomials t^i. By [7, Example 2.2], its entries are

$$c_{ijk} = \frac{j}{i+j} \cdot \frac{k}{i+j+k}.$$

Our second dictionary comes from an axis path in $\mathbb{R}^m$. It encodes all *piecewise linear paths* with $\le m$ steps. The ith entry in the dictionary is the piecewise linear basis function

$$\psi_i(t) = \begin{cases} 0 & \text{if } t \le \frac{i-1}{m}, \\ mt - (i-1) & \text{if } \frac{i-1}{m} < t < \frac{i}{m}, \\ 1 & \text{if } t \ge \frac{i}{m}. \end{cases} \tag{8.8}$$

By [7, Example 2.1], the associated core tensor C_{axis} is "upper-triangular", namely

$$c_{ijk} \quad = \quad \begin{cases} 1 & \text{if } i < j < k, \\ \frac{1}{2} & \text{if } i < j = k \text{ or } i = j < k, \\ \frac{1}{6} & \text{if } i = j = k, \\ 0 & \text{otherwise.} \end{cases} \tag{8.9}$$

The tensors C_{mono} and C_{axis} are real points in the *universal variety* $\mathcal{U}_{m,3} \subset (\mathbb{C}^m)^{\otimes 3}$, studied in [7]. The universal variety consists of all third order signatures of paths in $\mathbb{C}^m$, or equivalently all core tensors of dictionaries of size m. At present, we do not know whether all real points in $\mathcal{U}_{m,3}$ are in the topological closure of the signature tensors of real paths.

Example 8.11 (Generic Dictionaries)**.** *We describe a method for sampling real points in the universal variety $\mathcal{U}_{m,3}$, assuming [7, Conjecture 6.10]. Pick M random vectors $Y_1, Y_2, \ldots, Y_M$ in $\mathbb{R}^m$, where M exceeds $\frac{1}{3}m^2 + \frac{1}{2}m + \frac{1}{6} = \dim(\mathcal{U}_{m,3})$, and take the piecewise linear path with steps $Y_1, Y_2, \ldots, Y_M$. By [7, Example 5.4], the resulting* generic core tensor *equals*

$$C_{\text{gen}} = \frac{1}{6} \cdot \sum_{i=1}^{M} Y_i^{\otimes 3} + \frac{1}{2} \cdot \sum_{1 \le i < j \le M} \left(Y_i^{\otimes 2} \otimes Y_j + Y_i \otimes Y_j^{\otimes 2} \right) + \sum_{1 \le i < j < k \le M} Y_i \otimes Y_j \otimes Y_k. \tag{8.10}$$

The coefficients in Equation (8.10) *match the tensor entries in Equation* (8.9)*. By Chen's Formula [7, Equation (38)], the signature tensor C_{gen} is the degree 3 component in the tensor series $\sigma(\psi) = \exp(Y_1) \otimes \cdots \otimes \exp(Y_M)$, where $\exp(Y_i) = \sum_{k=0}^{\infty} \frac{1}{k!} Y_i^{\otimes k}$.*

An alternative method for sampling from $\mathcal{U}_{m,3}$ uses the Gröbner basis in [7, Theorem 4.10]. We write σ_{lyndon} for the vector of all signatures σ_i, σ_{ij} and σ_{ijk} whose indices are Lyndon words. This includes all m first order signatures σ_i, all $\binom{m}{2}$ second order signatures σ_{ij} with $i < j$, and all $\frac{1}{3}(m^3 - m)$ third order signatures σ_{ijk} satisfying $i < \min(j, k)$ or $i = j < k$. We pick these $m + \binom{m}{2} + \frac{1}{3}(m^3 - m)$ signature values to be random real numbers and substitute these numbers into the vector σ_{lyndon}. The non-Lyndon signatures σ_{ijk} are then computed by evaluating $\phi_{ijk}(\sigma_{\text{lyndon}})$, where ϕ_{ijk} is the normal form polynomial in [7, Theorem 4.10].

I now define what I mean by "learning paths" in the title of this chapter. Let C be a fixed core tensor of format $m \times m \times m$, such as C_{axis}, C_{mono} or C_{gen}. The data is an $n \times n \times n$ tensor X that is the third signature of some path in $\mathbb{R}^n$. Our hypothesis is that the path can be represented by the dictionary ψ, i.e. it is the image of ψ under a linear map. We seek an $n \times m$ matrix A such that $X = \sigma^{(3)}(A)$. In other words, given X and C, we wish to solve the tensor equation

$$[\![C; A, A, A]\!] = X,$$

the system of n^3 cubic equations in mn unknowns a_{ij}

$$\sum_{i=1}^{m} \sum_{j=1}^{m} \sum_{k=1}^{m} c_{ijk} a_{\alpha i} a_{\beta j} a_{\gamma k} = x_{\alpha\beta\gamma} \quad \text{for} \quad 1 \le \alpha, \beta, \gamma \le n. \tag{8.11}$$

The system in Equation (8.11) has a solution A if and only if the dictionary with core tensor C admits a path with signature tensor X. For the dictionaries we consider, the solution A is conjectured to be unique among real matrices provided $m < \frac{1}{3}n^2 + \frac{1}{2}n + \frac{1}{6}$, and unique up to scaling by a third root of unity if we allow complex matrices. The inequality means that the dimension of the universal variety $\mathcal{U}_{n,3}$ exceeds the number mn of unknowns, which is necessary for identifiability. For piecewise linear and polynomial paths, this is presented in [7, Conjecture 6.12 and Lemma 6.16].

8.3 Tensor congruence identifiability

Exact identifiability

In this subsection I prove identifiability properties for tensors coming from dictionaries of paths of interest. I consider generic dictionaries and piecewise linear paths. I then compare this notion of identifiability with other uniqueness properties that are studied for tensors.

Generic dictionaries

Consider an $m \times m \times m$ core tensor C that is generic in the universal variety $\mathcal{U}_{m,3}$. This is the third signature of a dictionary ψ which is generic in the sense of Example 8.11. In the following result, the field $\mathbb{K}$ can be either $\mathbb{R}$ or $\mathbb{C}$.

Theorem 8.12. *Let C be an $m \times m \times m$ tensor that is a generic point in the variety $\mathcal{U}_{m,3}$. The stabilizer of C is trivial, under congruence action by matrices in GL_m with entries in $\mathbb{K}$.*

Proof. I show that the complex stabilizer of the real tensor C is trivial: it consists only of the scaled identity matrices ηI, where $\eta^3 = 1$. For $m \leq 3$, this result is established by a direct Gröbner basis computation in maple. For $m \geq 4$ we use the following parametrization of the universal variety $\mathcal{U}_{m,3}$. Let P be a generic vector in $\mathbb{C}^m$, and let Q be a generic skew-symmetric $m \times m$ matrix. Following the definitions in [7, §4.1], we take L to be a generic element in the space $\mathrm{Lie}^{[3]}(\mathbb{C}^m)$ of homogeneous Lie polynomials of degree 3. Then

$$C = \frac{1}{6}P^{\otimes 3} + \frac{1}{2}(P \otimes Q + Q \otimes P) + L. \tag{8.12}$$

Indeed, $P + Q + L$ is a general Lie polynomial of degree ≤ 3, and the right hand side in Equation (8.12) is the degree 3 component in the expansion of its exponential, see [7, Example 5.15]. The constituents P, Q and L are recovered from C by taking the logarithm of C in the tensor algebra and extracting the homogeneous components of degree 1, 2 and 3. In particular, since these computations are equivariant with respect to the congruence action by GL_m, the stabilizer of C is contained in the stabilizer of L.

By [7, Proposition 4.7], a basis for the vector space $\mathrm{Lie}^{[3]}(\mathbb{C}^m)$ consists of the bracketings of all Lyndon words of length three on the alphabet $\{1, \ldots, m\}$. The number of these Lyndon

triples is $\frac{1}{3}(m^3 - m)$. The group $G = GL_m$ acts irreducibly on $\mathrm{Lie}^{[3]}(\mathbb{C}^m)$. By comparing dimensions, we see that

$$\mathrm{Lie}^{[3]}(\mathbb{C}^m) \simeq S_{(2,1)}(\mathbb{C}^m). \tag{8.13}$$

The right hand side is the irreducible G-module associated with the partition $(2,1)$ of the integer 3; see [105]. By the Hook Length Formula, the vector space dimension of Equation (8.13) equals $\frac{1}{3}(m^3 - m)$. This exceeds the dimension m^2 of the group, since $m \geq 4$. The map $C \mapsto L$ from the universal variety $\mathcal{U}_{m,3}$ to the G-module in Equation (8.13) is surjective, since the homogeneous Lie polynomial L in Equation (8.12) can be chosen arbitrarily.

I now apply Popov's classification [9, 142] of irreducible G-modules with non-trivial generic stabilizer. A special case of these results says that the stabilizer of a generic point L in the G-module $S_{(2,1)}(\mathbb{C}^m)$ is trivial. This implies that the stabilizer of the core tensor C under the congruence action of G is trivial. □

Theorem 8.12 and Corollary 8.8 imply the following.

Corollary 8.13. *Paths that are representable in a generic dictionary are identifiable from their third order signature. That is, let $m \leq n$ and let $C \in \mathcal{U}_{m,3}$ be a generic dictionary. Given $A \in \mathbb{R}^{n\times m}$ of rank m, the only real solution to $[\![C; A, A, A]\!] = [\![C; B, B, B]\!]$ is $A = B$.*

Piecewise linear paths

I now prove that piecewise linear paths in $\mathbb{R}^n$ with $m \leq n$ steps are uniquely recoverable from their third order signatures. As before, I take $\mathbb{K}$ to be $\mathbb{R}$ or $\mathbb{C}$. Let $C_{\text{axis}} \in \mathbb{K}^{m\times m\times m}$ be the piecewise linear core tensor in Equation (8.9) and X any tensor in the orbit

$$\big\{[\![C_{\text{axis}}; A, A, A]\!] \in \mathbb{K}^{n\times n\times n} : A \in \mathbb{K}^{n\times m}\big\}.$$

I show that there is a unique matrix A, up to third root of unity, such that $X = [\![C_{\text{axis}}; A, A, A]\!]$. This proves Conjectures 6.10 and 6.12 in [7] for $m \leq n$. In particular, if the field $\mathbb{K}$ is the real numbers $\mathbb{R}$, the matrix A can be uniquely recovered from X.

Lemma 8.14. *Let $A \in \mathbb{K}^{m\times m}$ be in the stabilizer of C_{axis} under congruence. If $e_m = (0, \ldots, 0, 1)^{\mathsf{T}}$ is an eigenvector of A then e_m is also an eigenvector of A^{T}.*

Proof. Any matrix A in the stabilizer of $C = C_{\text{axis}}$ also stabilizes the first and second order signatures, up to scaling by third root of unity, by Equation (8.7). The core tensor C represents a path from $(0, \ldots, 0)$ to $(1, \ldots, 1)$, so the first order signature is $b = (1, \ldots, 1)^{\mathsf{T}}$. The matrix A satisfies $Ab = \eta b$, where $\eta^3 = 1$, so b is an eigenvector of A. By [7, Example 2.1], the signature matrix of the piecewise linear dictionary is

$$C_2 = \begin{bmatrix} \frac{1}{2} & 1 & \cdots & 1 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \frac{1}{2} & 1 \\ 0 & \ldots & 0 & \frac{1}{2} \end{bmatrix}.$$

Since C_2 differs from AC_2A^T by a third root of unity, denoted η', we have

$$AC_2A^\mathsf{T} = \eta' C_2 \quad \Longrightarrow \quad \eta' C_2^{-1} = A^\mathsf{T} C_2^{-1} A \quad \Longrightarrow \quad \eta' C_2^{-1} b = \eta A^\mathsf{T} C_2^{-1} b. \tag{8.14}$$

Hence $v := \frac{1}{2} C_2^{-1} b = (\pm 1, \mp 1, \ldots, -1, 1)^\mathsf{T}$ is an eigenvector of A^T.

Consider the matrix obtained from C by multiplying by v along the second index, $D = [\![C; \cdot, v, \cdot]\!]$ or $d_{ik} = \sum_j c_{ijk} v_j$. The matrix D is diagonal, by the following direct computation. If $i < k$, we get

$$d_{ik} = \sum_{j=1}^{m} (-1)^{m+j} c_{ijk} = (-1)^m \left(\frac{1}{2}(-1)^i + \sum_{i<j<k} (-1)^j + \frac{1}{2}(-1)^k \right) = 0.$$

If $i > k$ all entries c_{ijk} vanish hence the sum is zero. If $i = k$, the only non-zero entry of C that appears in the sum is c_{iii} and we obtain $d_{ii} = \frac{1}{6}(-1)^{m+i}$. Also, by definition of D, we find that the matrix A is in its stabilizer under congruence, up to scaling by third root of unity, $D = \frac{\eta'}{\eta} ADA^\mathsf{T}$.

Suppose that e_m is an eigenvector of A. By the same argument as in Equation (8.14) we find that $D^{-1} e_m = 6 e_m$ is an eigenvector of A^T. Hence, e_m is an eigenvector of A^T. □

Theorem 8.15. *The stabilizer of the piecewise linear core tensor $C = C_{\text{axis}}$ is trivial, under the congruence action $C \mapsto [\![C; A, A, A]\!]$ by matrices A in GL_m with entries $\mathbb{K}$.*

Proof. Let A be in the stabilizer of C under GL_m. Evaluating $C = [\![C; A, A, A]\!]$ implies that c_{ijk} equals

$$\sum_{1 \le \alpha \le m} \frac{1}{6} a_{i\alpha} a_{j\alpha} a_{k\alpha} + \sum_{1 \le \alpha < \beta \le m} \frac{1}{2} a_{i\alpha} a_{j\alpha} a_{k\beta} + \sum_{1 \le \alpha < \beta \le m} \frac{1}{2} a_{i\alpha} a_{j\beta} a_{k\beta} + \sum_{1 \le \alpha < \beta < \gamma \le m} a_{i\alpha} a_{j\beta} a_{k\gamma}.$$

Here the constants in Equation (8.9) were substituted for the entries $c_{\alpha\beta\gamma}$ of C. We can express this equation as the dot product $c_{ijk} = f_{jk} \cdot A_i^\mathsf{T} = \sum_{\alpha=1}^{m} f_{jk}(\alpha) A_i(\alpha)$, where A_i is the ith row of A and f_{jk} denotes the row vector with α-coordinate

$$f_{jk}(\alpha) = \frac{1}{6} a_{j\alpha} a_{k\alpha} + \sum_{\beta > \alpha} \frac{1}{2} a_{j\alpha} a_{k\beta} + \sum_{\beta > \alpha} \frac{1}{2} a_{j\beta} a_{k\beta} + \sum_{\gamma > \beta > \alpha} a_{j\beta} a_{k\gamma}.$$

When $j > k$, the entry c_{ijk} vanishes, for all $1 \le i \le m$. Hence the vector f_{jk} for $j > k$ has dot product zero with all rows of A. Since the rows of A span $\mathbb{K}^m$, we see that f_{jk} is the zero vector, and the last entry $f_{jk}(m) = \frac{1}{6} a_{jm} a_{km}$ vanishes for all $j \neq k$.

We can express the entries c_{ijk} as a different dot product. Namely, factoring out the terms involving the jth row, we obtain $c_{ijk} = g_{ik} \cdot A_j^\mathsf{T}$, where g_{ik} is the row vector with β-coordinate

$$g_{ik}(\beta) = \frac{1}{6} a_{i\beta} a_{k\beta} + \sum_{\gamma > \beta} \frac{1}{2} a_{i\beta} a_{k\gamma} + \sum_{\alpha < \beta} \frac{1}{2} a_{i\alpha} a_{k\beta} + \sum_{\gamma > \beta > \alpha} a_{i\alpha} a_{k\gamma}.$$

For all $i > k$, the entry c_{ijk} vanishes. This means that the dot product of g_{ik} with all rows of A is zero, hence g_{ik} is the zero vector. Its mth entry $g_{ik}(m)$ equals

$$\frac{1}{6}a_{im}a_{km} + \sum_{\alpha=1}^{m-1}\frac{1}{2}a_{i\alpha}a_{km} = \frac{a_{km}}{2}\left(\sum_{\alpha=1}^{m} a_{i\alpha} - \frac{2}{3}a_{im}\right).$$

Since A stabilizes the first order signature, up to scaling by third root of unity η, the rows of A sum to η, hence $g_{ik}(m) = \frac{\eta}{2}a_{km} - \frac{1}{3}a_{km}a_{im}$ for all $i > k$. By the previous paragraph, the second term vanishes and, setting $i = m$, we deduce that $a_{km} = 0$ for all $1 \leq k < m$. This implies that the last column of A is parallel to the mth standard basis vector e_m, and hence that e_m is an eigenvector of A.

By Lemma 8.14, e_m is also an eigenvector of A^{T}. Thus, all entries in the last row of A vanish except the last. This means that A has the block diagonal structure

$$A = \begin{bmatrix} * & \cdots & * & 0 \\ \vdots & & \vdots & \vdots \\ * & \cdots & * & 0 \\ 0 & \cdots & 0 & 1 \end{bmatrix}.$$

The $*$ entries represent unknowns in an $(m-1) \times (m-1)$ block which I call A'.

Observe that A' stabilizes C', the axis core tensor in $\mathbb{K}^{(m-1)\times(m-1)\times(m-1)}$ arising from C by restricting to indices $1 \leq i, j, k \leq m-1$. From $C = [\![C; A, A, A]\!]$ we have $c_{ijk} = \sum_{\alpha,\beta,\gamma=1}^{m} c_{\alpha\beta\gamma}a_{i\alpha}a_{j\beta}a_{k\gamma}$. Since $a_{uv} = 0$ whenever $u < m = v$, this simplifies to

$$c_{ijk} = \sum_{\alpha,\beta,\gamma=1}^{m-1} c_{\alpha\beta\gamma}a_{i\alpha}a_{j\beta}a_{k\gamma} \quad \text{for} \quad 1 \leq i, j, k \leq m-1.$$

Hence A' is in the stabilizer of C'. The proof of Theorem 8.15 is concluded by induction on m, given that the assertion can be tested for small m by a direct computation. □

We deduce the following from Theorem 8.15 and Corollary 8.8, using the fact that the upper triangular tensor C_{axis} is symmetrically concise.

Corollary 8.16. *Piecewise linear paths in $\mathbb{R}^n$, consisting of at most n steps, can be uniquely recovered from the third order signature. That is, let $m \leq n$, let $C = C_{\text{axis}}$, and let A be a matrix of format $n \times m$ and of rank m. The only real solution to the tensor equation $[\![C; A, A, A]\!] = [\![C; B, B, B]\!]$ is the matrix $B = A$.*

Comparison with other notions of tensor identifiability

Identifiability for tensors is usually studied in the context of minimal decompositions, see e.g. [98]. The following result gives conditions under which algebraic identifiability of a minimal decomposition implies algebraic identifiability under congruence. We consider two decompositions of a tensor to be the same if they differ by a re-ordering of the rank one terms.

Theorem 8.17. *Let ψ be a dictionary that is not a loop. Suppose that its core tensor $C = C_\psi \in \mathbb{R}^{m\times m\times m}$ is symmetrically concise, has* $\operatorname{rank}(C) = r$*, and the number of minimal decompositions of C, denoted δ, is finite. Given a generic matrix $A \in \mathbb{R}^{n\times m}$ with $m \leq n$, there are at most $\delta \cdot \frac{r!}{(r-m)!}$ matrices $B \in \mathbb{R}^{n\times m}$ that have the same third order signature tensor as A.*

Proof. We determine the number of solutions B to the tensor equation

$$\sigma^{(3)}(A) = [\![C; A, A, A]\!] = [\![C; B, B, B]\!] = \sigma^{(3)}(B).$$

Consider a change of basis of $\mathbb{R}^m$ such that all standard basis vectors $e_1, \ldots, e_m$ occur in the minimal decomposition of C. This exists because C is symmetrically concise. Let W be the change of basis matrix. By Theorem 8.6, it suffices to count $m \times m$ matrices Z which stabilize $C' = [\![C; W, W, W]\!]$. This is the third order signature of the path $W\psi$, and it also has δ minimal decompositions.

Consider a minimal decomposition of C'. We have at most r choices for the image of e_1 (up to scale) in this decomposition. Then, we have at most $r - 1$ choices for e_2 up to scale, etc. This gives at most $\frac{r!}{(r-m)!}$ choices of $m \times m$ matrices N with $Z = N\Lambda$, where Λ is diagonal and invertible. Since ψ is not a loop, the first order signature $v = \psi(1) - \psi(0)$ is recoverable from the diagonal entries of the third order signature and $v \neq 0$ is also fixed by Z. Hence $Zv = v$, so $\Lambda v = N^{-1}v$. Evaluating the right hand side allows us to find Λ. □

The following example attains the bound in Theorem 8.17 non-trivially.

Example 8.18 ($m = n = 2$)**.** *Fix the dictionary $\psi = (\psi_1, \psi_2)$ with basis functions $\psi_1(t) = t - 10t^2 + 10t^3$ and $\psi_2(t) = 11t - 20t^2 + 10t^3$. By Equation* (8.6)*, the core tensor equals*

$$C_\psi = \frac{1}{42}\left[\begin{array}{cc||cc} 7 & -8 & 37 & -8 \\ -8 & 37 & -8 & 7 \end{array}\right].$$

Using Macaulay2 [76], we find that this tensor is symmetrically concise and has rank two, and a unique rank two decomposition. The stabilizer consists of two matrices:

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.$$

Our upper bound of $\delta \cdot \frac{r!}{(r-m)!} = 1 \cdot 2 = 2$ on the size of the stabilizer is attained. The stabilizer shows that C_ψ is unchanged under swapping the coordinates ψ_1 and ψ_2.

Numerical identifiability

In this subsection I quantify the numerical identifiability of recovering paths from third order signatures. A path in $\mathbb{R}^n$, coming from a dictionary of size m, can be recovered from its signature tensor $X \in \mathbb{R}^{n\times n\times n}$ by solving a system of n^3 equations in mn unknowns, see

Section 8.2. This can be done in principle using Gröbner basis methods. However, such methods are infeasible when m and n are large, or when the data X is inexact or noisy. In such cases we instead minimize the distance between X and the set of tensors $[\![C; A, A, A]\!]$ as A ranges over $\mathbb{R}^{n\times m}$. We seek the global minimum of the cost function

$$g : \mathbb{R}^{n\times m} \to \mathbb{R},\ A \mapsto \big\|[\![C; A, A, A]\!] - X\big\|^2, \tag{8.15}$$

where $\|\cdot\|$ denotes the Euclidean norm in tensor space, the *Frobenius norm*.

Fix a core tensor C with trivial stabilizer under congruence action. Consider the set $\mathcal{N}(C, n)$ of signature tensors X for which $\{A \in \mathbb{R}^{n\times m} : X = [\![C; A, A, A]\!]\}$ has cardinality at least two. The set $\mathcal{N}(C, n)$ consists of all *ill-posed instances*, for which path recovery is non-identifiable. However, even if a path is identifiable from its signature tensor in the exact sense of the previous subsection, different paths may lead to numerically indistinguishable signatures.

Definition 8.19 (Numerical non-identifiability)**.** *The numerical non-identifiability of a pair* (C, A) *is*

$$\kappa(A, C) \quad = \quad \frac{\|A\|^3 \cdot \|C\|}{\inf_{X\in\mathcal{N}(C,n)} \|[\![C; A, A, A]\!] - X\|},$$

where $\|\cdot\|$ *denotes the Frobenius norm.*

When the numerical non-identifiability is large, this reflects the proximity of $[\![C; A, A, A]\!]$ to a non-identifiable tensor. Conversely, a small value of the numerical non-identifiability means that all close-by tensors $[\![C; A, A, A]\!]$ are also identifiable. I give upper and lower bounds on $\kappa(A, C)$, in terms of the flattenings of C and the condition number of the rectangular matrix A, $\kappa(A) = \|A\| \cdot \|A^+\|$ where A^+ denotes the pseudo-inverse.

I first remark on connections between the numerical non-identifiability and the condition number. The condition number records how much the recovered matrix can change with small changes to the signature tensor. Following [40, Section O.2], and setting $X = [\![C; B, B, B]\!]$, the condition number of our recovery problem is

$$\mathrm{cond}(A, C) = \lim_{\delta\to 0} \sup_{\|[\![C;A,A,A]\!]-X\|\le\delta} \frac{\|A - B\|}{\|[\![C; A, A, A]\!] - X]\!]\|} \cdot \frac{\|[\![C; A, A, A]\!]\|}{\|A\|}. \tag{8.16}$$

When the condition number is finite, the matrix can be recovered uniquely using symbolic computations. However, when the condition number is large, small changes in the signature induce large changes in the recovered matrix, a problem for numerical computations. Following the approach introduced by [149] in the context of linear programming, the condition number is often determined by the inverse distance to the set of instances with infinite condition number [40, 59]. On the set of ill-posed instances $\mathcal{N}(C, n)$ the condition number is infinite, because the problem is non-identifiable. Hence the numerical non-identifiability gives a lower bound on the inverse distance to the instances with infinite condition number. Condition numbers for algebraic identifiability can be defined using the local set-up described

in [38]. Proving a *condition number theorem* to relate Equation (8.16) to the inverse distance would be an interesting topic for further study.

Theorem 8.20. *The numerical non-identifiability of the pair (A, C), for a matrix $A \in \mathbb{R}^{n\times m}$ and a tensor $C \in \mathbb{R}^{m\times m\times m}$ with trivial stabilizer, satisfies the upper bound*

$$\kappa(A,C) \quad \leq \quad \kappa(A)^3 \left(\frac{\|C\|}{\max(\varsigma_m^{(1)}, \varsigma_m^{(2)}, \varsigma_m^{(3)})} \right), \tag{8.17}$$

where $\varsigma_m^{(i)}$ denotes the smallest singular value of the ith flattening of the tensor C.

Proof. We aim to bound $\kappa(A,C)^{-1}$, the distance of $[\![C; A, A, A]\!]$ to the locus of non-identifiable tensors, from below. Since C has trivial stabilizer, Corollary 8.8 implies that all non-identifiable tensors must be of the form $[\![C; B, B, B]\!]$ where $B \in \mathbb{R}^{n\times m}$ has rank strictly less than m. The flattenings of such tensors have rank strictly less than m, so it suffices to lower-bound the distance of the flattenings to the much larger set $\{B \in \mathbb{R}^{n\times n^2} : \operatorname{rank}(B) < m\}$. We have

$$\begin{aligned}
\|A\|^3 \cdot \|A^+\|^3 \cdot \|C\| \cdot \kappa(A,C)^{-1} &\geq \min_{\operatorname{rank}(B)<m} \|AC^{(i)}(A\otimes A)^{\mathsf{T}} - B\| \cdot \|A^+\|^3 \\
&= \min_{\operatorname{rank}(B')<m} \|A(C^{(i)} - B')(A\otimes A)^{\mathsf{T}}\| \cdot \|A^+\|^3 \\
&\geq \min_{\operatorname{rank}(B')<m} \|C^{(i)} - B'\| \\
&= \quad \varsigma_m^{(i)},
\end{aligned}$$

where $A \otimes A \in \mathbb{R}^{n^2\times m^2}$ is the Kronecker product of the matrix A with itself, $B' \in \mathbb{R}^{m\times m^2}$, and $\|\cdot\|$ is the Frobenius norm. The chain of inequalities holds for $i = 1, 2, 3$, and the claim follows. □

We quantify the suitability of a core tensor C, with trivial stabilizer under congruence, for path recovery. We define the numerical non-identifiability of C to be the smallest number $\kappa(C)$ satisfying

$$\kappa(C) \quad \geq \quad \frac{\kappa(A,C)}{\kappa(A)^3}$$

for all $A \in \mathbb{R}^{n\times m}$ of rank m and all n. From Theorem 8.20, we obtain the following.

Corollary 8.21. *The numerical non-identifiability of $C \in \mathbb{R}^{m\times m\times m}$, with trivial stabilizer under congruence, satisfies*

$$\kappa(C) \quad \leq \quad \frac{\|C\|}{\max(\varsigma_m^{(1)}, \varsigma_m^{(2)}, \varsigma_m^{(3)})}.$$

Proof. Divide Equation (8.17) by $\kappa(A)^3$. The supremum of the left hand side, as A ranges over $\mathbb{R}^{n\times m}$ for all n, is equal to $\kappa(C)$. Hence $\kappa(C)$ is bounded by the right hand side. □

I now bound the numerical non-identifiability of the piecewise linear dictionary.

Corollary 8.22. *The numerical non-identifiability of C_{axis} is at most $6\|C_{\text{axis}}\|$.*

Proof. We show that the singular values of the second flattening $C^{(2)} \in \mathbb{R}^{m\times m^2}$ are at least $\frac{1}{6}$. The $j \times (i,k)$ entry is c_{ijk}. Since the entries of C are zero unless $i \le j \le k$, the flattening has an $m \times m$ block, indexed by $j \times (i,i)$, which equals $\frac{1}{6}$ times the identity matrix I. Let B denote the $m \times (m^2 - m)$ matrix obtained by removing these m columns. Then $C^{(2)}(C^{(2)})^{\mathsf{T}} = \frac{1}{36}I + BB^{\mathsf{T}}$. The singular values of $C^{(2)}$are the square roots of the eigenvalues of $C^{(2)}(C^{(2)})^{\mathsf{T}}$. Consider an eigenvector v of BB^{T} with eigenvalue λ. Then $\lambda \ge 0$ because BB^{T} is positive semi-definite and v is an eigenvector of $C^{(2)}(C^{(2)})^{\mathsf{T}}$ with eigenvalue $\frac{1}{36} + \lambda$. Hence the singular values of $C^{(2)}$ are bounded from below by $\frac{1}{6}$. □

The recovery problem is ill-posed if the tensor is not symmetrically concise, by Proposition 8.2. The following is a numerical analogue to Proposition 8.2.

Proposition 8.23. *Let $C \in \mathbb{R}^{m\times m\times m}$ and $C^{(\text{all})}$ be the $m \times 3m^2$ matrix obtained by concatenating the three flattening matrices $C^{(i)}$. If ς_m is the smallest singular value of $C^{(\text{all})}$ then*

$$\kappa(C) \quad \ge \quad \frac{\|C\|}{7m^{3/2}\varsigma_m}.$$

Proof. We compute the distance to a tensor X in the orbit of C that is not symmetrically concise. This gives an upper bound for the minimal distance to the set of ill-posed instances. Consider $X = [\![C; I - vv^{\mathsf{T}}, I - vv^{\mathsf{T}}, I - vv^{\mathsf{T}}]\!]$, where v is the left singular vector corresponding to the singular value ς_m of $C^{(\text{all})}$. Then v is in the kernel of all three flattenings of X, which means that X is not symmetrically concise.

We have $v^{\mathsf{T}}C^{(\text{all})} = \varsigma_m w^{\mathsf{T}}$ where w is the right singular vector of length $3m^2$, corresponding to singular value ς_m. We define w_i such that w is the stacking of w_1, w_2, w_3 with each w_i of length m^2. Then $v^{\mathsf{T}}C^{(i)} = \varsigma_m w_i^{\mathsf{T}}$ hence $\|v^{\mathsf{T}}C^{(i)}\| = \varsigma_m\|w_i\| \le \varsigma_m\|w\| \le \varsigma_m$. We use this to upper bound the distance from C to X, as follows:

$$\begin{aligned}\|C - X\| &= \|[\![C; vv^{\mathsf{T}}, I, I]\!] + [\![C; I, vv^{\mathsf{T}}, I]\!] + [\![C; I, I, vv^{\mathsf{T}}]\!] - [\![C; vv^{\mathsf{T}}, vv^{\mathsf{T}}, I]\!] \\ &\qquad - [\![C; I, vv^{\mathsf{T}}, vv^{\mathsf{T}}]\!] - [\![C; vv^{\mathsf{T}}, I, vv^{\mathsf{T}}]\!] + [\![C; vv^{\mathsf{T}}, vv^{\mathsf{T}}, vv^{\mathsf{T}}]\!]\| \\ &\le \left(\sum_{i=1}^{3} \|v^{\mathsf{T}}C^{(i)}\| + \|v^{\mathsf{T}}C^{(i)}\|\|vv^{\mathsf{T}}\|\right) + \|v^{\mathsf{T}}C^{(1)}\|\|vv^{\mathsf{T}}\|^2 \quad \le \quad 7\varsigma_m.\end{aligned}$$

We have $\kappa(C) \ge \frac{\kappa(A,C)}{\kappa(A)^3}$ for all matrices A of rank m. In particular, $\kappa(C) \ge \frac{\kappa(I,C)}{\kappa(I)^3}$. By definition of the numerical non-identifiability

$$\kappa(I, C) = \frac{\|I\|^3\|C\|}{\inf_{\tilde{X}\in\mathcal{N}(C,m)} \|C - \tilde{X}\|} \ge \frac{m^{3/2}\|C\|}{\|C - X\|} \ge \frac{m^{3/2}\|C\|}{7\varsigma_m},$$

since $\|I\| = \sqrt{m}$. The condition number of I is m, so the claim follows. □

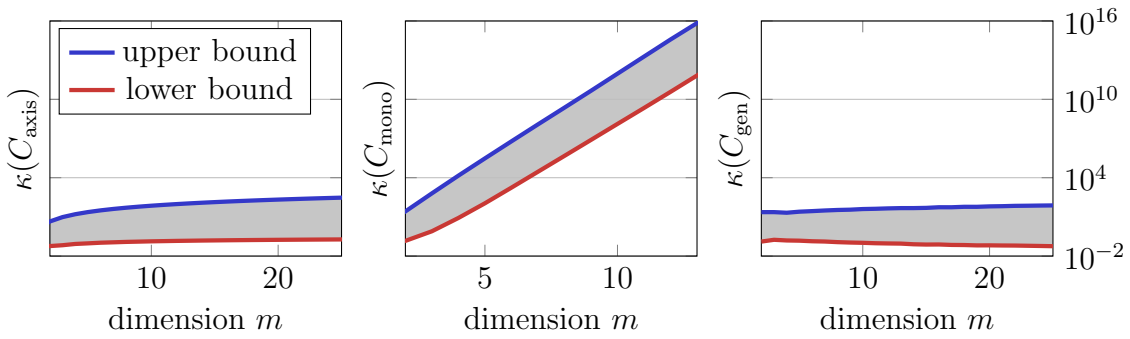


Figure 8.1: Bounds on the numerical non-identifiability of the piecewise linear core tensor (left), the monomial core tensor (middle), and generic core tensors (right).

The condition number quantifies the numerical identifiability of path recovery for the dictionary with core tensor C. I have derived informative upper and lower bounds on the related notion $\kappa(C)$, in terms of singular values of the flattenings of C. Figure 8.1 shows these bounds for small m. The lower bound is that in Corollary 8.21. The upper bound is that in Proposition 8.23. We see that the numerical non-identifiability of the monomial dictionary grows exponentially with m. We also empirically observe such a trend for other bases of polynomial functions, such as the Chebyshev functions. On the other hand, the piecewise linear dictionary is much more stable. Corollary 8.22 shows that the numerical non-identifiability of the piecewise linear dictionary remains below $6\|C_{\text{axis}}\| = 6\sqrt{\frac{m}{36} + \frac{1}{2}\binom{m}{2} + \binom{m}{3}}$, as seen on the left in Figure 8.1. The numerical non-identifiability of Generic dictionaries seems to remain below 100, independently of m. The right diagram in Figure 8.1 shows the average for 100 generic signature tensors C_{gen}, created using the first method in Example 8.11.

We can conclude that for piecewise linear paths, well-conditioned matrices A have signature tensors $[\![C_{\text{axis}}; A, A, A]\!]$ that are reasonably far from the non-identifiable locus. The same holds for paths from generic dictionaries, in a certain range of m. However, polynomial paths send well-conditioned matrices A to tensors $[\![C_{\text{mono}}; A, A, A]\!]$ which are very close to being non-identifiable, even for relatively small values of m. This suggests the possibility of numerical challenges for path recovery from such tensors, as I confirm in the numerical experiments in the next section.

8.4 Path recovery algorithms

Low-complexity paths

Given a fixed dictionary, we aim to compute a path represented by the dictionary whose signature most closely matches an input signature. In addition to the issues of numerical identifiability discussed in Section 8.3, numerical optimization has several well-documented

drawbacks, Since the objective function is non-convex, an abundance of local minima can be expected. The problem of local minima is inherent in almost all optimization methods, but there are some heuristic ways to overcome the problem. A thorough overview and application of state-of-the-art theory is out of the scope of this article. See [133].

I now discuss computational experiments for a range of values of m and n, using piecewise linear, polynomial, and generic paths, which were created using the first method in Example 8.11. For each pair (m, n), generate 100 random matrices $A \in \mathbb{R}^{n\times m}$ with entries $a_{ij} \sim N(0, 1)$ to represent the path $A\psi$. The tensor $X = \sigma^{(3)}(A)$ is then computed up to machine precision, and then the function in Equation (8.15) is minimized.

I used the Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm with an Armijo backtracking line search in Matlab 2018a. This was followed by a trust region Newton method for improved convergence, taken from the Manopt toolbox [36] which allows for direct implementation for matrix inputs. The algorithm stopped if $\|\nabla g(A)\| < 10^{-10}$ or after 100 steps of the BFGS procedure and 1000 steps of the trust region algorithm, allowing 10 reinitializations $a_{ij} \sim N(0, 1)$ to try to eliminate local minima and other numerical issues that arise from the relatively high degree of the objective function. Let A^* denote the result of this computation. The recovery was declared successful if $\|A^* - A\|/\|A^*\| < 10^{-5}$. Tables 8.1 and 8.2 show the percentages of successful recoveries. The success rate for piecewise linear paths is 100% for small m but it becomes slightly worse for larger m. For paths represented by a generic dictionary, the results are also close to 100%.

Ill-conditioning has occurred when a matrix A^* is recovered with a large distance to the original A, but whose signatures are very similar. I call an instance a *failure due to ill-conditioning* if the relative error between the matrices $\|A^* - A\|/\|A^*\|$ exceeds 10^{-5} but the relative distance between the signatures is less than 10^{-8}. This indicates a condition number exceeding 1000. Such failures never occurred for piecewise linear paths and generic paths, in over 10000 experiments. The situation is dramatically worse for polynomial paths: the subscripts in Table 8.2 count the failures due to ill-conditioning. For $m \geq 6$ if a matrix with sufficiently close signature was found then in all cases it was a failure due to ill-conditioning. The machine precision inaccuracy in the signature leads to large differences in the recovered matrix. The overall recovery rates for polynomial dictionaries are low. Although many of the other failures are not counted as being due to ill-conditioning under our requirements stated above, they often yield a relatively far away matrix with closeby signature tensor.

In conclusion, the experimental findings are consistent with the theoretical results on the numerical identifiability in Section 8.3. Generic paths and piecewise linear dictionaries behave best in numerical algorithms for recovering paths. The middle diagram in Figure 8.1 showed that the numerical non-identifiability of the monomial core tensor C_{mono} grows rapidly with m. The experiments confirm the difficulty of path recovery from the monomial dictionary.

$m\backslash d$	2	3	4	5	6	7	8	9	10	11	12	13	14	15
2	100	100	100	100	100	100	100	100	100	100	100	100	100	100
3		100	100	100	100	100	99	100	100	100	100	100	100	100
4			100	97	100	100	100	100	100	100	100	100	100	100
5				97	99	99	100	100	100	100	100	100	100	100
6					97	95	98	96	97	100	100	100	100	100
7						91	92	95	96	97	99	99	100	100
8							90	92	95	98	99	99	98	100
9								93	90	94	98	95	95	96
10									85	96	94	97	93	93

$m\backslash d$	2	3	4	5	6	7	8	9	10	11	12	13	14	15
2	100	100	100	100	100	100	100	100	100	100	100	100	100	99
3		99	100	100	100	100	100	98	100	100	100	100	99	100
4			99	99	100	100	100	98	98	98	99	98	98	100
5				100	100	100	100	100	99	99	100	100	100	100
6					98	99	100	100	100	100	99	100	100	99
7						100	98	97	99	99	99	100	100	100
8							99	99	99	100	99	98	98	99
9								97	92	98	97	97	99	98
10									100	98	97	97	100	99

Table 8.1: Percentage of successful path recoveries for random piecewise linear paths (top) and random paths represented by generic dictionaries (bottom).

Shortest paths

So far I have considered paths of low complexity in a space of high dimension. Such paths are identifiable from their third signature. In this final subsection, I consider a situation where the number of functions in the dictionary, m, is much larger than the dimension of the space, n. The paths are represented by a dictionary ψ, but identifiability no longer holds for the paths $A\psi$ because there are too many parameters to recover the matrix A from its third order signature. We can impose extra constraints to select a meaningful path among those with the same signature. A natural constraint is the length of the path. This leads to the problem of finding the shortest path for a given signature.

In this subsection I address the task of computing shortest paths when the third signature tensor is fixed. Recall that the length of a path $\psi : [0, 1] \to \mathbb{R}^m$ equals

$$\mathrm{len}(\psi) = \int_0^1 \sqrt{\langle \dot{\psi}(t), \dot{\psi}(t) \rangle} \mathrm{d}t,$$

$m\backslash d$	2	3	4	5	6	7	8	9	10	11	12
2	100_0	100_0	100_0	100_0	100_0	100_0	100_0	100_0	100_0	100_0	100_0
3		98_2	99_1	100_0	100_0	100_0	100_0	100_0	100_0	100_0	100_0
4			85_{13}	87_{13}	94_6	91_9	95_5	98_2	99_1	99_1	98_2
5				5_{31}	12_{24}	20_{30}	29_{37}	35_{40}	47_{41}	53_{38}	57_{37}
6					0_1	0_1	0_2	0_3	0_6	0_5	0_9
7						0_{21}	0_{24}	0_{35}	0_{29}	0_{28}	0_{35}

Table 8.2: The path recovery rate for polynomial paths is low once the condition number becomes too big. Subscripts count the failures due to ill-conditioning.

where $\dot{\psi}(t) = \frac{d\psi}{dt}$. This is a rather complicated function to evaluate in general. However, things are much easier for piecewise linear paths. For the m-step path given by the dictionary in Equation (8.8) and the matrix A, the length is given by the formula

$$\text{len}(A) := \text{len}(A\psi) = \sum_{j=1}^{m} \sqrt{\sum_{i=1}^{n} a_{ij}^2}.$$

Note that this function is piecewise differentiable. We can therefore regularize the objective function in Equation (8.15) with a length constraint. This leads to the new function

$$h(A, \lambda) = \text{len}(A) + \lambda g(A),$$

where λ is a parameter. A necessary condition for a minimum is that both the gradient in A and the gradient in λ equal zero. The latter requirement ensures that A yields the required signature. A problem with this method is that critical points are usually saddle points, which cannot be easily obtained using standard gradient-related techniques. This holds because h is not bounded from below for $\lambda \to -\infty$. A trick from [133] circumvents this problem. We fix λ_0 and minimize

$$h(A, \lambda_0)/\lambda_0 = \lambda_0^{-1}\text{len}(A) + g(A).$$

Once a minimum A_0 is found, we set $\lambda_1 = 2\lambda_0$ and minimize again with λ_1 and A_0 as a starting point for the iteration. We repeat, setting $\lambda_N = 2\lambda_{N-1}$ until λ_N is sufficiently large and the impact of the length constraint is negligible. Then, for some A_N, the function g is minimal, i.e. A_N has the correct signature up to machine precision. Local minima might occur – a guarantee that A_N gives *the* shortest path cannot be made. However, this method has proved to be satisfactory for the application.

Chapter 9

Tensor clustering with algebraic constraints

Clustering is the task of partitioning data into meaningful subsets, within which the data share some similarity, and between which they possess some difference. A wide range of clustering algorithms exist, such as agglomerative clustering, k-means, and spectral methods. Most clustering methods are designed for data points $x^{(i)}$ that are labeled by a single index i. There are also bi-clustering methods, which simultaneously cluster the rows and columns of a matrix in a compatible way: a clustering of two coupled datasets, $x^{(i)}$ and $y^{(j)}$.

Multi-dimensional datasets, which compare multiple factors simultaneously, are increasingly prevalent across the sciences. Analyzing them requires algorithms that preserve the multi-dimensional structure of the data. Usual clustering algorithms can be used, but they do not conserve the multi-dimensional structure. This flattens the insights that can be made, and hampers the interpretability of the results. Multi-dimensional data sets are prevalent in the biological sciences, and tensor methods have been developed to study them. For example, in [155], the authors use a singular value decomposition for tensors to analyze ovarian cancer survival. In [186], the authors study neuron dynamics using tensor decompositions. In [165], the authors study a tensor of factors associated with antibiotic resistance.

In this chapter, I introduce a clustering algorithm for multi-dimensional datasets, in which the data points $x^{(i_1,\ldots,i_d)}$ are labeled by multiple indices. The method uses multi-indexed information to produce shape-constrained clusters that are amenable to interpretation in the context of the application at hand. I describe the problem of clustering tensor data, and the benefits of structured clustering. I then describe two implementations of the structured clustering algorithm, one a standalone clustering tool, and the other for use in combination with an existing method, to add shape constraints to pre-existing clusters. Then I describe how the method is applied to a biological dataset from [131]. This chapter is based on joint work with Mariano Beguerisse-Díaz, Birgit Schoeberl, Mario Niepel and Heather Harrington, published in the Journal of the Royal Society Interface [166].

9.1 Tensor clustering

Consider a dataset of points $x^{(i)} \in \mathbb{R}^p$ for $i \in \{1, \ldots, n\}$. The p entries of the vector $x^{(i)}$ are measurements with respect to p different sensors. We can organize this data into a matrix X of format $n \times p$, whose rows are the vectors $x^{(i)}$. For example, the n data points could correspond to n different cell lines and the vector $x^{(i)}$ could be the measurements of a certain kinase (a type of enzyme) across p timepoints. Clustering means dividing the data points $x^{(i)}$ into subsets, i.e. partitioning the set of indices $\{1, \ldots, n\}$. See the section on principal component analysis on Page 12 for more details on extracting low-rank structure from matrix data.

Using modern experimental techniques, it is possible to record far more than a single output measurement. Instead of only measuring a single kinase over time, we can record many different output measurements simultaneously. Moreover, we can repeat the experiment under various experimental conditions. For each combination of independent variables, an experiment is performed with that gives various measurements, the outputs of multiple sensors, possibly across multiple timepoints. This leads to a multi-dimensional dataset such as the one in Example 1.4.

Multi-dimensional data can be organized into a tensor X of format

$$n_1 \times \cdots \times n_d \times p_1 \times \cdots \times p_h,$$

where each experiment is specified by d independent variables, and there are h output measurements. Each data point $x^{(i_1,\ldots,i_d)}$ is a point in the tensor space $\mathbb{K}^{p_1 \times \cdots \times p_h}$, where $\mathbb{K}$ is the field in which the measurements take values. As for the matrix case, the data points are combined to form the tensor X whose entry $x_{i_1 \ldots i_d j_1 \ldots j_h}$ is the $(j_1, \ldots, j_h)$ entry of the data point $x^{(i_1,\ldots,i_d)}$.

How can we extract structure from data that looks like this? One possibility is to flatten the data to a matrix of format $n \times p$ where $n = \prod_{i=1}^{d} n_d$ and $p = \prod_{j=1}^{h} p_j$. The rows of the flattening matrix are labeled by multi-indices $(i_1, \ldots, i_d)$ and the columns are labeled by $(j_1, \ldots, j_h)$. Once the data has been flattened to a matrix, matrix data analysis methods can be used. Flattening only reshapes the data into a matrix, and does not alter the data measurements. However, it loses structure of the tensor, as we saw in Chapter 4. In order to preserve the multi-dimensional structure of the data, we require methods that apply directly to the tensor. For more about flattenings, see Page 15.

Clustering the data points $x^{(i_1,\ldots,i_d)}$ means partitioning the set

$$\{(i_1, \ldots, i_d) : \quad i_1 \in \{1, \ldots, n_1\}, \ldots, i_d \in \{1, \ldots, n_d\}\}. \tag{9.1}$$

A partition of this set into m clusters can be encoded in multiple ways. Here I consider two encodings that will be useful.

1. A tensor Y of format $(n_1 \times \ldots \times n_d) \times (n_1 \times \ldots \times n_d)$ in which

$$y_{\mathbf{ij}} = \begin{cases} 0 & \text{if } x^{(\mathbf{i})} \text{ and } x^{(\mathbf{j})} \text{ are in the same cluster,} \\ 1 & \text{otherwise,} \end{cases} \tag{9.2}$$

where $\mathbf{i} = (i_1, \ldots, i_d)$ and $\mathbf{j} = (j_1, \ldots, j_d)$ are multi-indices labeling the data points. The tensor Y can be thought of as a Boolean approximation of the distances between pairs of data points: $y_{\mathbf{ij}} = 0$ if the data points $x^{(\mathbf{i})}$ and $x^{(\mathbf{j})}$ are 'close together' (in the same cluster), and $y_{\mathbf{ij}} = 1$ if they are 'far apart' (in different clusters). For the tensor Y to encode a valid clustering of the data, the three conditions of an equivalence relation must be met. These conditions are given by the following linear equations and inequalities, which must hold for all multi-indices $\mathbf{i}, \mathbf{j}$, and $\mathbf{k}$.

$$\begin{aligned} \text{Reflexivity:} \quad & y_{\mathbf{ii}} = 0, \\ \text{Symmetry:} \quad & y_{\mathbf{ij}} = y_{\mathbf{ji}}, \\ \text{Transitivity:} \quad & 0 \leq -y_{\mathbf{ik}} + y_{\mathbf{ij}} + y_{\mathbf{jk}} \leq 2. \end{aligned} \tag{9.3}$$

2. A tensor Z of format $n_1 \times \ldots \times n_d \times m$ in which

$$z_{\mathbf{i}k} = \begin{cases} 1 & \text{if the data indexed by } \mathbf{i} \text{ belongs to cluster } k, \\ 0 & \text{otherwise.} \end{cases} \tag{9.4}$$

Since each data point is assigned to exactly one cluster, the entries of the tensor Z satisfy the equation $\sum_{k=1}^{m} z_{\mathbf{i}k} = 1$.

The tensors Y and Z are related by the equation

$$1 - y_{\mathbf{ij}} = \sum_{k=1}^{m} z_{\mathbf{i}k} z_{\mathbf{j}k}.$$

9.2 Structured clustering

In this section I introduce an algorithm for the structured clustering of multi-dimensional data. I describe two implementations. First, the algorithm can be applied directly to a dataset as a standalone clustering tool. Second, the algorithm can also be used in combination with other clustering methods to impose constraints onto pre-existing clusters. The method to find pre-existing clusters must be chosen carefully to fit the application and should not be viewed as merely an initialization of the algorithm.

Among the wide range of clustering methods, constrained clustering is an active field of research [22, 43, 54, 108, 185]. The most common approaches incorporate pairwise *must-link* and *cannot-link* constraints to indicate whether two items must or cannot be in the same cluster [56, 183]. Here I consider the more general setting of shape-constrained clusters. I begin by motivating structured clustering, and describing its applicability to the setting of multi-dimensional data.

A clustering is a partition of data points. In the multi-dimensional setting, a clustering is given by a partition of the set of indices in Equation (9.1). Not all partitions of the data will be interpretable for the context at hand and the multi-indexed structure makes interpreting

the clusters more subtle. For example, assume that $x^{(1,1)}$ is in the same cluster as $x^{(2,2)}$. How can we interpret this similarity? How does our interpretation depend on the clustering assignment of $x^{(1,2)}$ and $x^{(2,1)}$?

I introduce a method to cluster multi-dimensional data with shape constraints on the clusters. I focus on the following shape constraint, which I describe in the case that each data point is labeled by two indices, $x^{(i_1,i_2)}$, i.e. that the data tensor has format $n_1 \times n_2 \times p_1 \times \cdots \times p_h$. The condition can be extended to the setting of three or more labeling indices. I focus on this case to simplify the notation and pictures, and because this is the setting that will later be used to analyze the dataset in Section 9.3.

The shape constraint is the following *rectangular* condition, see Figure 9.1. Assume that the data points labeled by (c_h, l_i) and (c_k, l_j) belong to the same cluster. Then the experiments labeled (c_h, l_j) and (c_k, l_i) must also be in this cluster. Each data point corresponds to a position on a two-dimensional grid. The constraint leads to clusters that are rectangular on the grid, Such clusters are more amenable to interpretation than unconstrained clusters, since they match a subset of one indexing set with a subset of another indexing set. The rectangular condition can be extended to the setting of three or more indices, pairing subsets of each index. Compared to the clustering methods outlined in [116], this method has the strength that it does not require the clusters to be connected rectangles on the grid. An ordering of the indexing set is artificial, and we seek clustering results that are not biased by this choice.

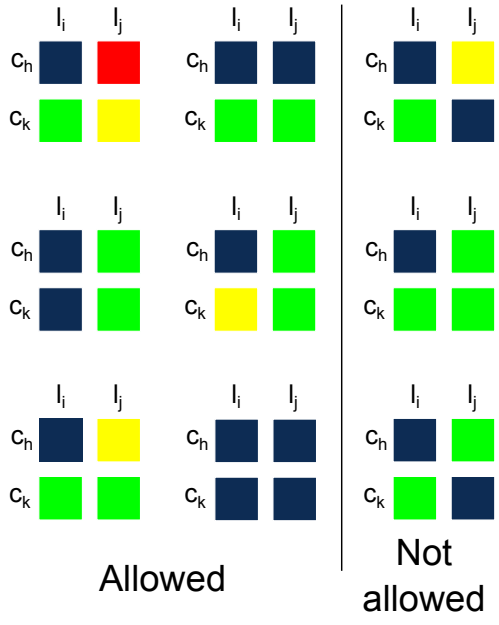


Figure 9.1: Examples of clusters that are allowed, and not allowed, with the rectangular shape constraint.

The interpretability constraints take the form of inequalities on the entries of the tensors Y or Z that encode the clustering partition, as I describe in more detail later. The optimal partition of the data can be obtained using integer linear optimization, by searching over unknown tensors satisfying these constraints. Specifically, I use the branch and cut algorithm [124] in the IBM ILOG CPLEX Optimization Studio [89].

I describe how the clustering algorithm can be applied to a biological dataset in Section 9.3, where I also describe the biological motivation for the rectangular constraint, and how it can be used to draw conclusions about mechanisms involved in breast cancer. I now give details for the implementation of the algorithm.

Standalone clustering

Assume we have a data tensor of format $n_1 \times \cdots \times n_d \times p_1 \times \cdots \times p_h$ where each data point $x^{(i_1,\ldots,i_d)}$ is in the space $\mathbb{R}^{p_1 \times \cdots \times p_h}$. We can construct a similarity tensor S, which records the similarity of the data points $x^{(\mathbf{i})}$ and $x^{(\mathbf{j})}$. For example, we can vectorize the two data points to give vectors of length $p = \prod_{k=1}^{h} p_k$, denoted by $v^{(\mathbf{i})}$ and $v^{(\mathbf{j})}$, and then compute the cosine dissimilarity between the two vectors

$$1 - \frac{\langle v^{(\mathbf{i})}, v^{(\mathbf{j})} \rangle}{\|v^{(\mathbf{i})}\| \|v^{(\mathbf{j})}\|},$$

where $\langle \cdot, \cdot \rangle$ is the usual inner product in a real vector space and $\| \cdot \|$ is the Euclidean norm.

I now describe how to partition the data into rectangular clusters. The clustering assignments will be recorded by the tensor Y from Equation (9.2). The rectangular condition corresponds to three types of algebraic constraint on the entries of the tensor Y. In the case of a data tensor of format $n_1 \times n_2 \times p_1 \times \cdots \times p_h$, the rectangular conditions are the following.

$$\begin{aligned} y_{i_1 i_2 j_1 j_2} &= y_{i_1 j_2 j_1 i_2}, \\ 0 \leq y_{i_1 i_2 j_1 j_2} - y_{i_1 i_2 j_1 i_2} &\leq 1, \\ 0 \leq y_{i_1 i_2 j_1 j_2} - y_{i_1 i_2 i_1 j_2} &\leq 1. \end{aligned} \tag{9.5}$$

The conditions must hold for all values $i_1, i_2 \in \{1, \ldots, n_1\}$ and $j_1, j_2 \in \{1, \ldots, n_2\}$.

The clustering method works by maximizing the similarity between experiments in the same cluster, over arrays Y that satisfy these conditions, as well as the conditions for being a clustering from Equation (9.3). That is, we solve the integer optimization problem

$$\begin{aligned} \max_{Y} \quad & \langle S, (\mathbf{1} - Y) \rangle + \lambda \langle \mathbf{1}, Y \rangle, \\ \text{subject to} \quad & b_l \leq V \cdot \mathrm{vec}(Y) \leq b_u, \end{aligned} \tag{9.6}$$

where the tensors Y and S are as above, and the vector $\mathrm{vec}(Y)$ is the tensor Y having been vectorized, a vector of length $n_1^2 n_2^2$. The coefficient λ is a regularization term introduced to control the number of clusters and $\mathbf{1}$ is the tensor of ones of format $n_1 \times n_2 \times n_1 \times n_2$. The

notation $\langle \cdot, \cdot \rangle$ denotes the entry-wise inner product, and $\cdot$ represents matrix multiplication. The matrix V encodes the constraints on Y given in Equations (9.3) and (9.5), The kth row of V is the kth constraint on Y: the entry is the coefficient (which can be 0, 1, or -1) with which each entry of Y appears in the constraint. Hence at most three entries in each row of V are non-zero. The kth entry of b_l and b_u (which can be 0, 1, or 2) gives the lower and upper bounds respectively of each linear inequality.

The resulting rectangular clusters are a sparse, low-rank representation of the data. The tensor $\mathbf{1} - Y$ of format $n_1 \times n_2 \times n_1 \times n_2$ gives a binary measure of the distance between any two experiments. This tensor has sparse block structure: it consists of m cuboids of 1s along the diagonal, where m is the number of clusters, and has zeros everywhere else. The tensor Y has flattening ranks bounded above by (m, m, m, m).

The following Matlab code generates the arrays in the optimization problem in Equation (9.6), for an array of format $n_1 \times n_2 \times p_1 \times \cdots \times p_h$. First, flatten the data into a three-dimensional array, `Tensr`, of format $n_1 \times n_2 \times p$, where $p = \prod_{k=1}^{h} p_h$. The following code makes the similarity tensor S with respect to cosine dissimilarity.

```
[n1 n2 p] = size(Tensr);
C = zeros(n1,n1,n2,n2);
for i = 1:n1; for j = 1:n1; for k = 1:n2; for l = 1:n2;
                v1 = zeros(p,1); v2 = zeros(p,1);
                for ii = 1:p;
                    v1(ii) = Tensr(i,k,ii); v2(ii) = Tensr(j,l,ii);
                end
                C(i,j,k,l) = 1 - dot(v1,v2)/(norm(v1,2)*norm(v2,2));
end end end end
```

Next we encode the constraints on the tensor Y, via the matrix V. It is constructed as a sparse array, by specifying row and column indices `Vrow` and `Vcol`, and values `Vval` of all non-zero entries. The lower and upper bounds for each linear constraints are organized into the vectors `lb` and `ub` respectively.

```
t = 1;
for i = 1:n; for j = 1:n; for k = 1:m; for l = 1:m;
                if (i ~= j) || (k ~= l);
                    W = vectorize(n,i,j,k,l);
                    Vrow(w) = t; Vcol(w) = W;Vval(w) = 1; w = w+1;
                    W = vectorize(n,j,i,l,k);
                    Vrow(w) = t; Vcol(w) = W; Vval(w) = -1; w = w+1;
                    lb(t) = 0; ub(t) = 0; t = t+1;
                end
                if (k ~= l);
                    W = vectorize(n,i,j,k,l);
```

```
                    Vrow(w) = t; Vcol(w) = W; Vval(w) = 1; w = w+1;
                    W = vectorize(n,i,j,l,k);
                    Vrow(w) = t; Vcol(w) = W; Vval(w) = -1; w = w+1;
                    lb(t) = 0; ub(t) = 0; t = t+1;
                % opposite diagonals must be same
                end
                if (i ~= j);
                    W = vectorize(n,i,j,k,l);
                    Vrow(w) = t; Vcol(w) = W; Vval(w) = 1; w = w+1;
                    W = vectorize(n,j,i,k,l);
                    Vrow(w) = t; Vcol(w) = W; Vval(w) = -1; w = w+1;
                    lb(t) = 0; ub(t) = 0; t = t+1;
                end
                if (k ~= l);
                    W = vectorize(n,i,j,k,l);
                    Vrow(w) = t; Vcol(w) = W; Vval(w) = 1; w = w+1;
                    W = vectorize(n,j,i,k,k);
                    Vrow(w) = t; Vcol(w) = W; Vval(w) = -1; w = w+1;
                    lb(t) = 0; ub(t) = 1; t = t+1;
                % vertical conditions
                end
                if (i ~= j);
                    W = vectorize(n,i,j,k,l);
                    Vrow(w) = t; Vcol(w) = W; Vval(w) = 1; w = w+1;
                    W = vectorize(n,i,i,k,l);
                    Vrow(w) = t; Vcol(w) = W; Vval(w) = -1; w = w+1;
                    lb(t) = 0; ub(t) = 1; t = t+1;
                % horizontal conditions
                end
end end end end
for i = 1:n; for k = 1:m;
        W = vectorize(n,i,i,k,k);
        Vrow(w) = t; Vcol(w) = W; Vval(w) = 1; w = w+1;
        lb(t) = 0; ub(t) = 0; t = t+1;
end end
for i1 = 1:n; for i2 = 1:n; for i3 = 1:n;
for k1 = 1:m; for k2 = 1:m; for k3 = 1:m;
if (((i1 ~= i2) || (k1 ~= k2)) && ((i2 ~= i3) || (k2 ~= k3))
&& ((i1 ~= i3) || (k1 ~= k3)));
            W = vectorize(n,i1,i3,k1,k3);
            Vrow(w) = t; Vcol(w) = W; Vval(w) = -1; w = w+1;
            W = vectorize(n,i1,i2,k1,k2);
```

```
            Vrow(w) = t; Vcol(w) = W; Vval(w) = 1; w = w+1;
            W = vectorize(n,i2,i3,k2,k3);
            Vrow(w) = t; Vcol(w) = W; Vval(w) = 1; w = w+1;
            lb(t) = 0; ub(t) = 2; t = t+1;
end end end end end end end
maxt = t-1;
V = sparse(Vrow,Vcol,Vval,maxt,n*n*m*m);
```

The function `vectorize` combines the multi-index into a single index:

```
function t = vectorize(n,m,i,j,k,l);
i1 = n*m*m*(i-1); j1 = m*m*(j-1); k1 = m*(k-1); l1 = l;
t = i1+j1+k1+l1;
```

Pre-existing clusters

Assume we have a partitioning of the multi-indexed data X that is not rectangular. In this subsection, I describe how to find the nearest rectangular clusters. The input is an initial partition of the data points into m clusters. We then modify as few clustering assignments as possible, to reach the closest rectangular clustering of the data.

The initial clustering is encoded by a partition tensor T of format $n_1 \times n_2 \times m$, with entries

$$t_{\mathbf{i}k} = \begin{cases} 1, & \mathbf{i} \text{ is in cluster } k, \\ 0, & \text{otherwise,} \end{cases}$$

where $\mathbf{i} = (i_1, i_2)$ indexes an experiment. The new clusters are encoded by a tensor Z of the same format, defined according to Equation (9.4). In order to have rectangular clusters, the entries of Z must satisfy the linear inequalities

$$\sum_{r=1}^{m} z_{ijr} = 1, \quad \text{(unique cluster assignment)}$$

$$z_{ikr} + z_{jlr} - z_{ilr} \leq 1. \quad \text{(interpretability condition)}$$

As before, we use the branch and cut algorithm to obtain the global optimum of the optimization problem

$$\max_{Z} \quad \langle T, Z \rangle.$$

The entrywise inner product $\langle \cdot, \cdot \rangle$ sums the number of clustering of assignments unchanged by the optimization. Similarly to the tensor Y arising from the standalone clustering algorithm, the tensor Z also has sparse and low-rank structure. The two-dimensional slices of format $n_1 \times n_2$ consist of a rectangle of 1s and all other values equal to 0.

We generate the constraints on the tensor Z using the following Matlab code. Assume that there are m pre-existing clusters. As before, the constraints are organized into the matrix V and the lower and upper bounds in the constraints are the vectors b_l and b_u.

```
w = 1; t = 1;
for i = 1:n1; for j = 1:n2; or k = 1:m;
            W = m*n2*(i-1) + m*(j-1) + k;
            Vrow(w) = t; Vcol(w) = W; Vval(w) = 1; w = w+1;
        end
        lb(t) = 1; ub(t) = 1; t = t+1;
end end
for i = 1:n1; for j = 1:n1; for k = 1:n2; for l = 1:n2;
                if (i ~= j) && (k ~= l) ;
                for r = 1:m;
                    W = m*n2*(i-1) + m*(k-1) + r;
                    Vrow(w) = t; Vcol(w) = W; Vval(w) = 1; w = w+1;
                    W = m*n2*(j-1) + m*(l-1) + r;
                    Vrow(w) = t; Vcol(w) = W; Vval(w) = 1; w = w+1;
                    W = m*n2*(i-1) + m*(l-1) + r;
                    Vrow(w) = t; Vcol(w) = W; Vval(w) = -1; w = w+1;
                    lb(t) = -1; ub(t) = 1;  t = t+1;
                end end
end end end end
maxt = t-1;
Vnew = sparse(Vrow,Vcol,Vval,maxt,n1*n2*m);
```

9.3 Application to biological data

In this section, I describe how the clustering method is applied to a biological dataset in [166]. I describe the biological dataset, motivate the rectangular condition on the clusters, and give sample output of the method.

In terms of broader applicability, the algorithm can be used to impose any shape constraints that take the form of linear inequalities, including the must-link and cannot-link conditions from usual structured clustering. For example, the algorithm could be used to construct optimal portfolios that comply with rules about their composition [121], to help the formation of teams that maximize members' preferences and are compliant with skill requirements [57], or to find communities in networks with quotas. Depending on the application, there may not be a measurement for every combination of indices, so the tensor may be incomplete. The method can be adapted for dealing with tensors of incomplete entries, by only optimizing over the data entries that are known.

We examine an experimental dataset detailing the temporal phosphorylation response of genetically diverse breast cancer cell lines to different ligands, a dataset originally introduced in [131]. See Example 1.4 for an explanatory picture of the dataset. The output measurements are the activation levels of the mitogen-activated protein kinase (MAPK) and phosphoinositide 3-kinase (PI3K) pathways, which are involved in cellular decisions and fates [49, 97, 119, 144], and are known to dysfunction in cancer [21, 80, 120, 167, 188]. The dataset is complete: there is a measurement for each combination of cell line, ligand, dose, time point, protein. It can be represented by a tensor of order 5 and format $36 \times 14 \times 2 \times 4 \times 2$ whose dimensions correspond to 36 cell lines, 14 ligands, 2 doses, 4 time points, and 2 proteins (ERK/MAPK or AKT/PI3K). Each experiment $x^{(c_h,l_i)} \in \mathbb{R}^{2\times4\times2}$ consists of adding the ligand l_i to the cell line c_h. It is labeled by a (cell line, ligand) pair, hence we have $36 \cdot 14 = 504$ experiments. Our aim is to find sets of experiments with a similar temporal response and, specifically, to find similarities consistent with mechanisms for signal transduction.

In a clinical setting, prognosis and treatment decisions for breast cancer are guided by tumor grade, stage and clinical subtype [130], which is based on the presence of cellular receptors:

- HER2^{amp} cells are characterized by amplification of the HER2 gene, leading to overexpression of the ErbB2 receptor tyrosine kinase;
- HR^{+} cells are characterized by the expression of the estrogen receptor (ER) or progesterone receptor (PR);
- Triple negative breast cancer (TNBC) cells are negative for HER2 amplification, and express ER and PR at low levels.

The key signaling proteins and subtype responses in breast cancer cells are known; however, among genetically diverse cell lines the specific dysfunction mechanisms vary and are not well understood [84, 91, 131]. A better understanding would lead to improved personalized treatments for breast cancer.

A high similarity between experiments suggests the possibility of a common underlying biological mechanism. But in our clustering method, high similarity alone is not enough to cluster two experiments together. We also require that the observations are compatible with the same mechanistic interpretation.

If a similarity between experiments (c_h, l_i) and (c_k, l_j) arises for a mechanistic reason, the cell lines must share some property (e.g. a mutation) that causes them to respond in the same way to the ligands. If we swap the ligand, i.e., we look at the experiment (c_k, l_i) or (c_h, l_j), this experiment should share a similar temporal response. For this reason, we constrain the clusters to match a subset of cell lines with a subset of ligands: we seek clusters which are rectangular. See the left columns of Figure 9.1 for examples of rectangular clusters. Conversely, if (c_h, l_i) and (c_k, l_j) are clustered together but (c_k, l_i) and (c_h, l_j) are not (see the right column of Figure 9.1), it is more difficult to assign mechanistic interpretation to the

cluster. In this way, imposing the rectangular shape constraint helps to rule out similarities between measurements that are spurious, incompatible with a mechanistic interpretation.

I give sample output of the method, for a partition of the data into three clusters based on pre-existing non-rectangular clusters, in Figure 9.2. The clustering is visualized by colour-coding the grid of experiments according to their cluster assignment. Each entry of the grid represents the cluster assignment of the tensor from that experiment. On the left hand side, we see the pre-existing clustering of the data into non-rectangular clusters. On the right hand side we see the output of the algorithm: the closest partition of the experiments into rectangular clusters. The ligands have been divided into two subsets, and within one of these subsets the cell lines have been divided into two subsets. The labels on the cell lines show how the three clinical subtypes disperse among the clusters, while the labels on the ligands show that the cluster of ligands consists of the ERB4 and FGFR subtypes, as well as the HGF ligand. I refer the reader to the paper [166] for more output from the method, and more details on the biological and mechanistic interpretation of the clustering assignments.

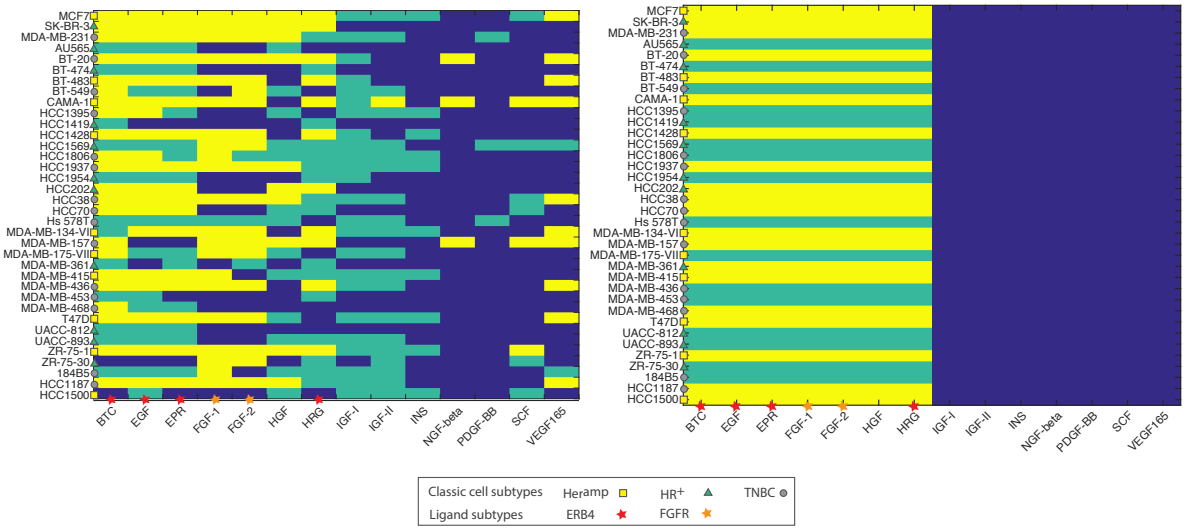


Figure 9.2: A non-rectangular clustering (left) and its nearest rectangular clustering (right). The clustering assignments are represented by yellow, green and blue squares.

Bibliography

[1] H. Abo, G. Ottaviani, and C. Peterson. "Induction for secant varieties of Segre varieties". In: *Transactions of the American Mathematical Society* 361.2 (2008), pp. 767–792.

[2] H. Abo, A. Seigal, and B. Sturmfels. "Eigenconfigurations of tensors". In: *Algebraic and Geometric Methods in Discrete Mathematics*. Ed. by H. Harrington, M. Omar, and M. Wright. Vol. 685. Contemporary Mathematics, Amer. Math. Soc, 2017, pp. 1–25.

[3] E. Acar et al. "Tensor-based fusion of EEG and fMRI to understand neurological changes in schizophrenia". In: *IEEE International Symposium on Circuits and Systems* (2017), p. 17208572.

[4] J. Alexander and A. Hirschowitz. "Polynomial interpolation in several variables". In: *J. Algebraic Geom.* 4 (1995), pp. 201–222.

[5] E. Allman et al. "Maximum likelihood estimation of the latent class model through model boundary decomposition". In: *Journal of Algebraic Statistics* 10.1 (2019), pp. 51–84.

[6] E. Allman et al. "Tensors of nonnegative rank two". In: *Linear Algebra and its Applications* 473 (2015), pp. 37–53.

[7] C. Améndola, P. Friz, and B. Sturmfels. "Varieties of signature tensors". In: *Forum of Mathematics, Sigma* 7.e10 (2019).

[8] A. Anandkumar et al. "Tensor decompositions for learning latent variable models". In: *Journal of Machine Learning Research* 15 (2014), pp. 2773–2832.

[9] E. M. Andreev and V. L. Popov. "Stationary subgroups of points of general position in the representation space of a semisimple Lie group". In: *Funct. Anal. Appl.* 5 (1971), pp. 265–271.

[10] I. Arad et al. "Rigorous RG algorithms and area laws for low energy eigenstates in 1D". In: *Communications in Mathematical Physics* 356.1 (2017), pp. 65–105.

[11] W. Austin, G. Ballard, and T. Kolda. "Parallel tensor compression for large-scale scientific data". In: *Proceedings of the 30th IEEE International Parallel and Distributed Processing Symposium* (2016), pp. 912–922.

[12] P. Austrin, P. Kaski, and K. Kubjas. "Tensor network complexity of multilinear maps". In: *10th Innovations in Theoretical Computer Science Conference* 7.1-21 (2019).

[13] N. Ay and A. Knauf. "Maximizing multi–information". In: *Kybernetika* 42.5 (2006), pp. 517–538.

[14] M. Bachmayr, R. Schneider, and A. Uschmajew. "Tensor networks and hierarchical tensors for the solution of high-dimensional partial differential equations". In: *Foundations of Computational Mathematics* 16.6 (2016), pp. 1423–1472.

[15] B. Bader and T. G. Kolda. *MATLAB tensor toolbox https://www.tensortoolbox.org.*

[16] R. Bailly, F. Denis, and G. Rabusseau. "Recognizable series on hypergraphs". In: *Language and Automata Theory and Applications, Lecture Notes in Comput. Sci., 8977, Springer, Cham* (2015), pp. 639–651.

[17] E. Ballico et al. "Bounds on the tensor rank". In: *Annali di Matematica Pura ed Applicata* 197.6 (2018), pp. 1771–1785.

[18] M. Banchi. "Rank and border rank of real ternary cubics". In: *Boll. Unione Mat. Ital.* 8.1 (2015), pp. 65–80.

[19] A. S. Bandeira et al. "Estimation under group actions: recovering orbits from invariants". In: *preprint arXiv:1712:10163* (2017).

[20] A. Banerjee, A. Char, and B. Mondal. "Spectra of general hypergraphs". In: *Linear Algebra and its Applications* 518 (2017), pp. 14–30.

[21] J. Baselga. "Targeting tyrosine kinases in cancer: the second wave". In: *Science* 312.5777 (2006), pp. 1175–1178.

[22] S. Basu, I. Davidson, and K. Wagstaff. *Constrained clustering: Advances in algorithms, theory, and applications.* CRC Press, 2008.

[23] D. Bates and L. Oeding. "Toward a salmon conjecture". In: *Exp. Math.* 20.3 (2011), pp. 358–370.

[24] D. Bates et al. *Bertini: Software for Numerical Algebraic Geometry, available at bertini.nd.edu.*

[25] N. Beerenwinkel, L. Pachter, and B. Sturmfels. "Epistasis and shapes of fitness landscapes". In: *Statistica Sinica* 17.4 (2007), pp. 1317–1342.

[26] Y. Bengio. "Learning deep architectures for AI". In: *Found. Trends Mach. Learn.* 2.1 (2009), pp. 1–127.

[27] C. Berge. *Hypergraphs.* Ed. by N.-H. M. Library. Vol. 45. Combinatorics of finite sets. Amsterdam: North-Holland Publishing Co., 1989.

[28] A. Bernardi, G. Blekherman, and G. Ottaviani. "On real typical ranks". In: *Boll. Unione Mat. Ital.* 11.3 (2018), pp. 293–307.

[29] J. Biamonte and V. Bergholm. "Quantum tensor networks in a nutshell". In: *preprint arXiv:1708.00006* (2017).

[30] C. M. Bishop. *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer-Verlag New York, Inc., 2006.

[31] G. Blekherman and R. Sinn. "Real rank with respect to varieties". In: *Linear Algebra and its Applications* 505 (2016), pp. 344–360.

[32] J. Bochnak, M. Coste, and M. F. Roy. *Real algebraic geometry*. A Series of Modern Surveys in Mathematics. Springer-Verlag, 1998.

[33] A. Boralevi et al. "Orthogonal and unitary tensor decomposition from an algebraic perspective". In: *Israel Journal of Mathematics* 222.1 (2017), pp. 223–260.

[34] K. Borsuk. "On the imbedding of systems of compacta in simplicial complexes". In: *Fund. Math.* 35 (1948), pp. 217–234.

[35] W. Bosma, J. Cannon, and C. Playoust. "The Magma algebra system. I. The user language". In: *Journal of Symbolic Computation* 24 (1997), pp. 235–265.

[36] N. Boumal et al. "Manopt, a Matlab toolbox for optimization on manifolds". In: *J. Machine Learning Research* 15 (2014), pp. 1455–1459.

[37] J. Brachat et al. "Symmetric tensor decomposition". In: *Linear Algebra and its Applications* 433 (2010), pp. 1851–1872.

[38] P. Breiding and N. Vannieuwenhoven. "Convergence analysis of Riemannian Gauss-Newton methods and its connection with the geometric condition number". In: *Appl. Math. Lett.* 78 (2018), pp. 42–50.

[39] J. W. Bruce and C. T. C. Wall. "On the classification of cubic surfaces". In: *J. London Math. Soc.* 19.2 (1979), pp. 245–256.

[40] P. Bürgisser and F. Cucker. *Condition. The Geometry of Numerical Algorithms*. Vol. 349. Fundamental Principles of Mathematical Sciences. Heidelberg: Springer-Verlag, 2013.

[41] E. Carlini, C. Guo, and E. Ventura. "Real and complex Waring rank of reducible cubic forms". In: *J. Pure Appl. Algebra* 220.11 (2016), pp. 3692–3701.

[42] A. Cayley. "On the theory of linear transformations". In: *Cambridge Mathematical Journal* 4 (1845), pp. 193–209.

[43] M. E. Celebi. *Partitional clustering algorithms*. Springer, 2014.

[44] K. Chang, L. Qi, and G. Zhou. "Singular values of a real rectangular tensor". In: *Journal of Mathematical Analysis and Applications* 370 (2010), pp. 284–294.

[45] J. C. Chen. "The nonnegative rank factorizationsn of nonnegative matrices". In: *Linear Algebra and its Applications* 62.207-217 (1984).

[46] J. Chen et al. "On the equivalence of restricted Boltzmann machines and tensor network states". In: *Phys. Rev. B* 97.8 (2018), p. 085104.

[47] K. T. Chen. "Integration of paths - a faithful representation of paths by noncommutative formal power series". In: *Transactions of the American Mathematical Society* 89 (1958), pp. 395–407.

[48] K. T. Chen. "Integration of paths, geometric invariants and a generalized Baker-Hausdorff formula". In: *Annals of Mathematics* 65 (1957), pp. 163–178.

[49] W. W. Chen et al. "Input–output behavior of ErbB signaling pathways as revealed by a mass action model trained against dynamic data". In: *Molecular Systems Biology* 5.1 (2009).

[50] P. Comon et al. "Symmetric tensors and symmetric tensor rank". In: *SIAM J. Matrix Anal. Appl.* 30.3 (2008), pp. 1254–1279.

[51] A. Critch and J. Morton. "Algebraic geometry of matrix product states". In: *Symmetry Integrability Geom. Methods Appl.* 10 (2014), p. 095.

[52] I. Csiszár and P. C. Shields. "Information theory and statistics: a tutorial". In: *Foundations and Trends in Communications and Information Theory* 1.4 (2004), pp. 417–528.

[53] M. A. Cueto, E. A. Tobis, and J. Yu. "An implicitization challenge for binary factor analysis". In: *Journal of Symbolic Computation* 45.12 (2010), pp. 1296–1315.

[54] T.-B.-H. Dao, K.-C. Duong, and C. Vrain. "Constrained clustering by constraint programming". In: *Artificial Intelligence* 244 (2017), pp. 70–94.

[55] J. H. Davenport and J. Heintz. "Real quantifier elimination is doubly exponential". In: *Journal of Symbolic Computation* 5 (1988), pp. 29–35.

[56] I. Davidson and S. Basu. "A survey of clustering with instance level constraints". In: *ACM Transactions on Knowledge Discovery from Data* (2007), pp. 1–41.

[57] E. W. Davis and G. E. Heidorn. "An algorithm for optimal project scheduling under multiple resource constraints". In: *Management Science* 17.12 (1971).

[58] J. Demmel. *Applied Numerical Linear Algebra*. Other Titles in Applied Mathematics. SIAM, 1997.

[59] J. Demmel. "On condition numbers and the distance to the nearest ill-posed problem". In: *Numerische Mathematik* 51.3 (1987), pp. 251–289.

[60] J. Diehl and J. Reizenstein. "Invariants of multidimensional time series based on their iterated-integral signature". In: *Acta Applicandae Mathematicae* (2018), pp. 1–40.

[61] I. Domanov, A. Stegeman, and L. D. Lathauwer. "On the largest multilinear singular values of higher-order tensors". In: *SIAM J. Matrix Anal. Appl.* 38.4 (2017), pp. 1434–1453.

[62] J. Draisma and G. Regts. "Tensor invariants for certain subgroups of the orthogonal group". In: *J Algebr Comb* 38 (2013), pp. 393–405.

[63] C. Eckart and G. Young. "The approximation of one matrix by another of lower rank". In: *Psychometrika* 1.3 (1936), pp. 211–218.

[64] D. Eisenbud. *Commutative Algebra with a View Toward Algebaric Geometry*. Vol. Graduate Texts in Mathematics. Springer-Verlag, 1995.

[65] D. Eisenbud and B. Sturmfels. "Binomial ideals". In: *Duke Mathematical Journal* 84 (1996), pp. 1–45.

[66] D. B. Ennis and G. Kindlmann. "Orthogonal tensor invariants and the analysis of diffusion tensor magnetic resonance images". In: *Magnetic Resonance in Medicine* 55.1 (2006), pp. 136–146.

[67] S. Friedland. "Best rank one approximation of real symmetric tensors can be chosen symmetric". In: *Front. Math. China* 8.1 (2013), pp. 19–40.

[68] S. Friedland. "Remarks on the symmetric rank of symmetric tensors". In: *SIAM J. Matrix Anal. Appl.* 37.1 (2016), pp. 320–337.

[69] S. Friedland and E. Gross. "A proof of the set-theoretic version of the salmon conjecture". In: *Journal of Algebra* 356.1 (2012), pp. 374–379.

[70] S. Friedland and G. Ottaviani. "The number of singular vector tuples and uniqueness of best rank one approximation of tensors". In: *Found. Comput. Math.* 14 (2014), pp. 1209–1242.

[71] P. Friz and M. Hairer. *A Course on Rough Paths. With an introduction to regularity structures*. Vol. Universitext. Springer, 2014.

[72] E. Gawrilow and M. Joswig. "Polymake: a framework for analyzing convex polytopes". In: *Polytopes—combinatorics and computation* (1997), pp. 43–73.

[73] D. Geiger, C. Meek, and B. Sturmfels. "On the toric algebra of graphical models". In: *The Annals of Statistics* 34.3 (2006), pp. 1463–1492.

[74] I. Gelfand, M. Kapranov, and A. Zelevinsky. *Discriminants, Resultants, and Multidimensional Determinants*. Birkhäuser, 1994.

[75] R. Goodman and N. R. Wallach. *Symmetry, Representations, and Invariants*. Vol. 255. Graduate Texts in Mathematics. Springer, 2009.

[76] D. Grayson and M. Stillman. *Macaulay2, a software system for research in algebraic geometry*.

[77] W. Hackbusch. *Tensor Spaces and Numerical Tensor Calculus*. Vol. 42. Springer Series in Computational Mathematics, 2012.

[78] W. Hackbusch and A. Uschmajew. "On the interconnection between the higher-order singular values of real tensors". In: *Numerische Mathematik* 135.3 (2017), pp. 875–894.

[79] B. Hambly and T. Lyons. "Uniqueness for the signature of a path of bounded variation and the reduced path group". In: *Annals of Mathematics* 171 (2010), pp. 109–167.

[80] D. Hanahan and R. Weinberg. "Hallmarks of cancer: the next generation". In: *Cell* 144.5 (2011), pp. 646–674.

[81] M. Hastings. "An area law for one-dimensional quantum systems". In: *Journal of Statistical Mechanics: Theory and Experiment* 8 (2007), P08024.

[82] A. Hatcher. *Algebraic Topology*. Cambridge: Cambridge University Press, 2002.

[83] J. Hauenstein et al. "Homotopy techniques for tensor decomposition and Homotopy techniques for tensor decomposition and perfect identifiability". In: *J. Reine Angew. M* (2016).

[84] L. M. Heiser et al. "Subtype and pathway specific responses to anticancer compounds in breast cancer." In: *Proc Natl Acad Sci* 109.8 (Feb. 2012), pp. 2724–2729.

[85] C. Hillar and L.-H. Lim. "Most tensor problems are NP hard". In: *Journal of the ACM* 60.6 (2013), article 45.

[86] *http://cubics.wikidot.com/*.

[87] *http://gs.statcounter.com/search-engine-market-share*.

[88] *http://www.internetlivestats.com/one-second/*.

[89] IBM. *IBM ILOG CPLEX Optimization Studio CPLEX User's Manual*. 2011.

[90] A. Knutson and T. Tao. "Honeycombs and sums of Hermitian matrices". In: *Notices Amer. Math. Soc.* 48.2 (2001), pp. 175–186.

[91] W. Kolch et al. "The dynamic control of signal transduction networks in cancer cells." In: *Nat Rev Cancer* 15.9 (2015), pp. 515–527.

[92] T. Kolda. "Orthogonal tensor decompositions". In: *SIAM J. Matrix Anal. Appl.* 23.1 (2001), pp. 243–255.

[93] T. Kolda and B. Bader. "Tensor decompositions and applications". In: *SIAM Review* 51 (2009), pp. 455–500.

[94] A. Kormilitzin et al. "Application of the signature method to pattern recognition in the CEQUEL clinical trial". In: *preprint arXiv:1606.02074* (2016).

[95] A. Kormilitzin et al. "Detecting early signs of depressive and manic episodes in patients with bipolar disorder using the signature-based model". In: *preprint arXiv:1708.01206* (2017).

[96] S. Krämer. "The geometrical description of feasible singular values in the tensor train format." In: *preprint arXiv:1701.08437* (2017).

[97] A. von Kriegsheim et al. "Cell fate decisions are specified by the dynamic ERK interactome." In: *Nature* 11.12 (Dec. 2009), pp. 1458–1464.

[98] J. Kruskal. "Three-way arrays: rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics". In: *Linear Algebra Appl.* 18 (1977), pp. 95–138.

[99] F. Kschischang, B. J. Frey, and H.-A. Loeliger. "Factor graphs and the sum-product algorithm". In: *IEEE Trans. Inform. Theory* 47.2 (2001), pp. 498–519.

[100] K. Kubjas, P. Parrilo, and B. Sturmfels. "How to flatten a soccer ball". In: *Homological and Computational Methods in Commutative Algebra*. Ed. by A. Conca, J. Gubeladze, and T. Römer. Vol. 20. Springer INdAM series, 2017, pp. 141–162.

[101] J. M. Landsberg and L. Manivel. "On the ideals of secant varieties of Segre varieties". In: *Foundations of Computational Mathematics* 4.4 (2004), pp. 397–422.

[102] J. M. Landsberg, Y. Qi, and K. Ye. "On the geomety of tensor network states". In: *Journal of Quantum Information and Computation* 12.3-4 (2012), pp. 346–354.

[103] J. Landsberg. *Geometry and Complexity Theory*. Cambridge University Press, 2017.

[104] J. Landsberg. *Tensors and their uses in Approximation Theory, Quantum Information Theory and Geometry*. draft notes, 2017.

[105] J. Landsberg. *Tensors: Geometry and Applications*. Providence RI: Graduate Studies in Mathematics, American Mathematical Society, 2012.

[106] L. D. Lathauwer, B. D. Moor, and J. Vandewalle. "A multilinear singular value decomposition". In: *SIAM J. Matrix Anal. Appl.* 21.4 (2000), pp. 1253–1278.

[107] S. Lauritzen. *Graphical Models*. Oxford Statistical Sci. Ser. Clarendon Press, 1996.

[108] F. Li, S. Li, and T. Denœux. "k-CEVCLUS: Constrained evidential clustering of large dissimilarity data". In: *Knowledge-Based Systems* 142 (2018), pp. 29–44.

[109] L.-H. Lim. "Singular values and eigenvalues of tensors: a variational approach". In: *Proceedings of IEEE Workshop on Computational Advances in Multisensor Adaptive Processing* (2005), pp. 129–132.

[110] L.-H. Lim and P. Comon. "Nonnegative approximations of nonnegative tensors". In: *J Chemometr.* 23 (2009), pp. 432–441.

[111] H.-A. Loeliger and P. Vontobel. "A factor-graph representation of probabilities in quantum mechanics". In: *IEEE Int. Symp. Inf. Theory, Cambridge, MA, USA* (2012), pp. 656–660.

[112] H. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos. "Multilinear principal component analysis of tensor objects". In: *IEEE Transactions on Neural Networks* 19.1 (2008), p. 9742962.

[113] T. Lyons and Z. Qian. *System Control and Rough Paths*. Oxford University Press, 2002.

[114] T. Lyons and W. Xu. "Hyperbolic development and inversion of signature". In: *J. Funct. Anal.* 272 (2017), pp. 2933–2955.

[115] T. Lyons and W. Xu. "Inverting the signature of a path". In: *Journal of the European Mathematical Society* 20.7 (2014), pp. 1655–1687.

[116] S. C. Madeira and A. L. Oliveira. "Biclustering algorithms for biological data analysis: a survey". In: *IEEE/ACM Trans. Comput. Biol. Bioinformatics* 1.1 (2004), pp. 24–45.

[117] *Maple 2017. Maplesoft, a division of Waterloo Maple Inc., Waterloo, Ontario.*

[118] I. Markov and Y. Shi. "Simulating quantum computation by contracting tensor networks". In: *SIAM J. Comput.* 38.3 (2008), pp. 963–981.

[119] C. J. Marshall. "Specificity of receptor tyrosine kinase signaling: transient versus sustained extracellular signal-regulated kinase activation". In: *Cell* (1995), pp. 179–185.

[120] J. A. McCubrey et al. "Therapeutic resistance resulting from mutations in Raf/MEK/ERK and PI3K/PTEN/Akt/mTOR signaling pathways". In: *Journal of Cellular Physiology* 226.11 (2011), pp. 2762–2781.

[121] A. McNeil, R. Frey, and P. Embrechts. *Quantitative Risk Management: Concepts, Techniques and Tools*. Princeton Series in Finance. Princeton University Press, 2015.

[122] M. Michalek and H. Moon. "Spaces of sums of powers and real rank boundaries". In: *Beiträge zur Algebra und Geometrie / Contributions to Algebra and Geometry* (2018).

[123] M. Michałek, L. Oeding, and P. Zwiernik. "Secant cumulants and toric geometry". In: *International Mathematics Research Notices* 12 (2015), pp. 4019–4063.

[124] J. E. Mitchell. "Branch-and-cut algorithms for combinatorial optimization problems". In: *Handbook of Applied Optimization* (2002), pp. 65–77.

[125] G. Montúfar and J. Morton. "When does a mixture of products contain a product of mixtures?" In: *SIAM Journal on Discrete Mathematics* 29.1 (2015), pp. 321–347.

[126] G. Montúfar. "Mixture decompositions of exponential families using a decomposition of their sample spaces". In: *Kybernetika* 49.1 (2013), pp. 23–39.

[127] G. Montúfar and J. Rauh. "Hierarchical models as marginals of hierarchical models". In: *International Journal of Approximate Reasoning* 88 (2017), pp. 531–546.

[128] G. Montúfar, J. Rauh, and N. Ay. "Expressive power and approximation errors of restricted Boltzmann machines". In: *Advances in Neural Information Processing Systems 24*. Curran Associates, Inc., 2011, pp. 415–423.

[129] G. Montúfar, J. Rauh, and N. Ay. "Maximal information divergence from statistical models defined by neural networks". In: *Geometric Science of Information: First International Conference, GSI 2013, Paris, France, August 28-30, 2013. Proceedings*. Ed. by F. Nielsen and F. Barbaresco. Berlin, Heidelberg: Springer, 2013, pp. 759–766.

[130] *National Cancer Institute, http://www.cancer.gov/*. Aug. 2016.

[131] M. Niepel et al. "Analysis of growth factor signaling in genetically diverse breast cancer lines." In: *BMC Biology* 12.20 (2014).

[132] T. Nishino and K. Okunishi. "Corner transfer matrix renormalization group". In: *J. Phys. Soc. Jpn.* 65 (1996), pp. 891–894.

[133] J. Nocedal and S. J. Wright. *Numerical Optimization*. Springer Verlag, 2006.

[134] J. Novembre et al. "Genes mirror geography within Europe". In: *Nature* 456 (2008), pp. 98–101.

[135] L. Oeding. "Hyperdeterminants of polynomials". In: *Advances in Math.* 231 (2012), pp. 1308–1326.

[136] L. Oeding and C. Raicu. "Tangential varieties of Segre-Veronese varieties". In: *Collectanea Mathematica* 65 (2014), pp. 303–330.

[137] R. Orús. "A practical introduction to tensor networks: matrix product states and projected entangled pair state". In: *Ann. Physics* 349 (2014), pp. 117–158.

[138] I. Oseledets. "Tensor-train decomposition". In: *SIAM Journal on Scientific Computing* 33.5 (2011), pp. 2295–2317.

[139] J. Pan and M. Ng. "Symmetric orthogonal approximation to symmetric tensors with applications to image construction". In: *Numerical Linear Algebra with Applications* 25.5 (2018), e2180.

[140] M. Pejic. "Quantum Bayesian networks with application to games displaying Parrondo's paradox". PhD thesis. University of California, Berkeley, 2014.

[141] M. Pfeffer, A. Seigal, and B. Sturmfels. "Learning paths from signature tensors". In: *SIAM Journal on Matrix Analysis and Applications* 40.2 (2019), pp. 394–416.

[142] A. M. Popov. "Finite stationary subgroups in general position of simple linear Lie groups". In: *Trans. Mosc. Math. Soc.* 48 (1988), pp. 3–63.

[143] D. Porras, F. Verstraete, and J. I. Cirac. "Density matrix renormalization group and periodic boundary conditions: a quantum information perspective". In: *Phys. Rev. Lett.* 93.227205 (2004).

[144] J. E. Purvis and G. Lahav. "Encoding and Decoding Cellular Information through Signaling Dynamics." In: *Cell* 152.5 (Feb. 2013), pp. 945–956.

[145] L. Qi and Z. Luo. *Tensor analysis: Spectral theory and special tensors*. Philadelphia, PA: Society for Industrial and Applied Mathematics, 2017.

[146] C. Raicu. "Secant varieties of Segre-Veronese varieties". In: *Algebra and Number Theory* 6 (2012), pp. 1817–1868.

[147] K. Ranestad and B. Sturmfels. "On the convex hull of a space curve". In: *Advances in Geometry* 12 (2012), pp. 157–178.

[148] K. Ranestad and B. Sturmfels. "The convex hull of a variety". In: *Notions of Positivity and the Geometry of Polynomials*. Ed. by P. Bränden, M. Passare, and M. Putinar. Springer Verlag, Basel, 2011, pp. 331–344.

[149] J. Renegar. "Incorporating condition measures into the complexity theory of linear programming". In: *SIAM Journal on Optimization* 5.3 (1995), pp. 506–524.

[150] N. Robertson and P. Seymour. "Graph minors. III. planar tree-width". In: *J. Combin. Theory Ser. B* 36.1 (1984), pp. 49–64.

[151] E. Robeva. "Orthogonal decomposition of symmetric tensors". In: *SIAM J. Matrix Anal. Appl.* 37 (2016), pp. 86–102.

[152] E. Robeva and A. Seigal. "Duality of graphical models and tensor networks". In: *Information and Inference: a Journal of the IMA* iay009 (2018).

[153] E. Robeva and A. Seigal. "Singular vectors of orthogonally decomposable tensors". In: *Linear and Multilinear Algebra* 65.12 (2017), pp. 2457–2471.

[154] *Sage Mathematics Software Version 7.4.*

[155] P. Sankaranarayanan et al. "Tensor GSVD of patient- and platform-matched tumor and normal DNA copy-number profiles uncovers chromosome arm-wide patterns of tumor-exclusive platform-consistent alterations encoding for cell transformation and predicting ovarian cancer survival". In: *PLoS One* (2015).

[156] A. Schmitt. "Quaternary cubic forms and projective algebraic threefolds". In: *Enseign. Math.* 43 (1997), pp. 253–270.

[157] U. Schollwöck. "The density-matrix renormalization group". In: *Rev. Mod. Phys.* 77.259 (2005).

[158] A. Schrijver. "Tensor subalgebras and first fundamental theorems in invariant theory". In: *Journal of Algebra* 319 (2008), pp. 1305–1319.

[159] B. Segre. *The Non-singular Cubic Surfaces.* Oxford: Oxford University Press, 1942.

[160] A. Seigal. "Gram determinants of real binary tensors". In: *Linear Algebra and its Applications* 544 (2018), pp. 350–369.

[161] A. Seigal. "Ranks and symmetric ranks of cubic surfaces". In: *preprint arXiv:1801.05377* (2018).

[162] A. Seigal. "The algebraic statistics of an Oberwolfach workshop". In: *Snapshots of modern mathematics from Oberwolfach* 1 (2018).

[163] A. Seigal and G. Montúfar. "Mixtures and products in two graphical models". In: *Journal of Algebraic Statistics* 9.1 (2018), pp. 1–20.

[164] A. Seigal and B. Sturmfels. "Real rank two geometry". In: *Journal of Algebra* 484 (2017), pp. 310–333.

[165] A. Seigal et al. "Does antibiotic resistance evolve in hospitals?" In: *Bulletin of Mathematical Biology* 79 (2017), pp. 191–208.

[166] A. Seigal et al. "Tensor clustering with algebraic constraints gives interpretable groups of crosstalk mechanisms in breast cancer". In: *Journal of the Royal Society Interface* 16.151 (2019), p. 20180661.

[167] V. Serra et al. "PI3K inhibition results in enhanced HER signaling and acquired ERK dependency in HER2-overexpressing breast cancer". In: *Oncogene* 30.22 (2011), pp. 2547–2557.

[168] Y. Shitov. "A Counterexample to Comon's Conjecture". In: *SIAM J. Appl. Algebra Geometry* 2.3 (2018), pp. 428–443.

[169] N. D. Sidiropoulos, G. B. Giannakis, and R. Bro. "Blind PARAFAC receivers for DS-CDMA systems". In: *IEEE Transactions on Signal Processing* 48.3 (2000), pp. 810–823.

[170] V. de Silva and L.-H. Lim. "Tensor rank and the ill-posedness of the best low-rank approximation problem". In: *SIAM J. Matrix Anal. Appl.* 30 (2008), pp. 1084–1127.

[171] *singsurf.org/parade/Cubics*.

[172] R. Sinn. "Algebraic boundaries of $SO(2)$-orbitopes". In: *Discrete Comput. Geom.* 50 (2013), pp. 219–235.

[173] R. P. Stanley. *An Introduction to Hyperplane Arrangements*. AMS, 2007.

[174] A. Stegeman and P. Comon. "Subtracting a best rank-1 approximation may increase tensor rank". In: *Linear Algebra and its Applications* 433.1276-1399 (2010).

[175] A. Stegeman and S. Friedland. "On best rank-2 and rank-(2,2,2) approximations of order-3 tensors". In: *Linear and Multilinear Algebra* 65 (2017), pp. 1289–1310.

[176] G. W. Stewart. "On the early history of the singular value decomposition". In: *SIAM Review* 35.4 (1993), pp. 551–566.

[177] V. Strassen. "Relative bilinear complexity and matrix multiplication". In: *J. Reine Angew. M* (1987), pp. 406–443.

[178] B. Sturmfels. "The Hurwitz form of a projective variety". In: *Journal of Symbolic Computation* 79 (2017), pp. 186–196.

[179] S. Sullivant. *Algebraic Statistics*. Vol. 194. American Mathematical Society, Graduate Studies in Mathematics, 2018.

[180] *Tensorflow https://www.tensorflow.org/*.

[181] R. Vakil. *The Rising Sea: Foundations of Algebraic Geometry*. online, 2017.

[182] N. Vervliet et al. *Tensorlab 3.0 https://www.tensorlab.net/*.

[183] K. Wagstaff et al. "Constrained K-means clustering with background knowledge". In: *Proceedings of the Eighteenth International Conference on Machine Learning*. 2001, pp. 577–584.

[184] M. Wainwright and M. Jordan. *Graphical Models, Exponential Families, and Variational Inference*. Vol. 1. Foundation and Trends in Machine Learning 1-2. Now Publishers Inc, 2008.

[185] X. Wang, B. Qian, and I. Davidson. "On constrained spectral clustering and its applications". In: *Data Mining and Knowledge Discovery* 28.1 (2014), pp. 1–30.

[186] A. H. Williams et al. "Unsupervised discovery of demixed, low-dimensional neural dynamics across multiple timescales through tensor components analysis". In: *Neuron* 98.6 (2018), pp. 1099–1115.

[187] L. K. Williams. "Invariant polynomials on tensors under the action of a product of orthogonal groups". In: *Transactions of the American Mathematical Society* 368 (2016), pp. 1411–1433.

[188] J.-K. Won et al. "The crossregulation between ERK and PI3K signaling pathways determines the tumoricidal efficacy of MEK inhibitor". In: *Journal of Molecular Cell Biology* 4.3 (2012), pp. 153–163.

[189] Z. Y. Xie et al. "Coarse-graining renormalization by higher-order singular value decomposition". In: *Phys. Rev. B* 86.045139 (2012).

[190] K. Ye and L.-H. Lim. "Tensor network ranks". In: *preprint arXiv:1801.02662* (2018).

[191] F. Zak. *Tangents and Secants of Algebraic Varieties*. Vol. 127. Translations of Mathematical Monographs. American Mathematical Society, Providence, RI, 1993.

[192] T. Zhang and G. Golub. "Rank-one approximation to high order tensors". In: *SIAM Journal on Matrix Analysis and Applications* 23.2 (2001), pp. 534–550.

[193] X. Zhang, Z.-H. Huang, and L. Qi. "Comon's conjecture, rank decomposition, and symmetric rank decomposition of symmetric tensors". In: *SIAM J. Matrix Anal. Appl.* 37.4 (2016), pp. 1719–1728.

[194] P. Zwiernik. *Semialgebraic statistics and latent tree models*. Vol. 146. Chapman & Hall/CRC, Boca Raton, FL: Monographs on Statistics and Applied Probability, 2016.