UC Berkeley
UC Berkeley Electronic Theses and Dissertations

Title
Applications of non-linear algebra to biology

Permalink
https://escholarship.org/uc/item/2cv4q5b3

Author
Cartwright, Dustin Alexander

Publication Date
2010

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

Applications of non-linear algebra to biology

by

Dustin Alexander Cartwright

A dissertation submitted in partial satisfaction of the
requirements for the degree of
Doctor of Philosophy

in

Mathematics

and the Designated Emphasis

in

Computational and Genomic Biology

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Bernd Sturmfels, Chair
Professor Lior Pachter
Professor Yun S. Song

Fall 2010

Applications of non-linear algebra to biology

Copyright 2010
by
Dustin Alexander Cartwright

Abstract

Applications of non-linear algebra to biology

by

Dustin Alexander Cartwright

Doctor of Philosophy in Mathematics
and the Designated Emphasis in Computational and Genomic Biology

University of California, Berkeley

Professor Bernd Sturmfels, Chair

We present two applications of non-linear algebra to biology. Our first application is to the analysis of gene expression data from *Arabidoposis* roots. In Chapter 2, we present a method for computing non-negative roots to certain systems of polynomials. This algorithm is based on a generalization of the Expectation-Maximization and Iterative Proportional Fitting from statistics. In Chapter 3, this method is applied to a model for gene expression coming from roots of the *Arabidopsis* plant. Variation in gene expression is one method in which different tissue types develop different functional characteristics. Our model for gene expression in these roots is non-linear and so we apply the method from Chapter 2.

Our second application is the use of secant varieties to study mixture models for the distribution of single-nucleotide polymorphisms in genes. In Chapter 4, we give generators for the defining ideal of the first 5 secant varieties of $\mathbb{P}^2 \times \mathbb{P}^{n-1}$ embedded by $\mathcal{O}(1,2)$. Our equations come from a generalization of the flattening of a tensor, which we call an exterior flattening. The study of secant varieties is a classical subject in algebraic geometry, which has recently been connected to applications in algebraic statistics. In Chapter 5, this connection is used in order to understand how single-nucleotide polymorphisms occur within human genes. Because genes perform important functions, there is selective pressure against mutations which affect the gene's behavior. As we show, these selective pressures are closely tied to the nature by which genetic sequence codes for the constituent amino acids of a protein.

Contents

List of Figures

List of Tables

Acknowledgments

I'd like to acknowledge my mother, father, and sister for their support during my time in graduate school. In addition, Philip Benfey, Siobhan Brady, Daniel Erman, Luke Oeding, and David Orlando were my co-authors on papers which served as a basis for this dissertation. I'd like to thank them for their collaboration on these papers and for the many things they've taught me. Lastly, I'd like to thank Bernd Sturmfels, for his tireless guidance, without which this document wouldn't have been possible.

Chapter 1

Introduction

In this dissertation, we present two applications of non-linear algebra to biology. Our use of non-linear algebra includes both polynomial algebra, in which linear equations are replaced by polynomials, and multi-linear algebra, in which a single linear dependence, represented by a matrix, is replaced by multiple linear dependences, represented by a tensor. While linear algebra is widely applied across many fields, the applications of non-linear algebra are not as prevalent. Non-linearity presents additional computational and interpretational challenges, some of which are addressed in this dissertation.

Because of the non-linearity, algebraic geometry becomes a key tool in our analysis. Algebraic geometry is the study of solutions to polynomial equations. Traditionally, the solutions are taken to be in an algebraically closed field, such as the complex numbers, and the strongest results still only hold in that context. For example, in Chapter 2, we study the problem of finding only non-negative real solutions to polynomial equations. A number of methods exist for finding all complex solutions, from which the non-negative real solutions can be selected. However, in cases when the number of complex solutions is vastly larger, it is more efficient to directly find just the non-negative real solutions, but few methods exist in this situation. Our method for doing so is specifically built for real non-negative solutions.

Our first biological application is inferring expression patterns in *Arabidoposis* roots, which is presented in Chapter 3. Variation in gene expression level is a key factor in the differentiation of cell functionality, and thus essential to understanding the developmental process. We describe a model for expression levels in the root of the *Arabidopsis* plant. Our model is parametrized by bilinear polynomials. In order to solve these equations, we adapt the methods of Expectation-Maximization and Iterative Proportional Fitting from statistics. These adaptations are described in Chapter 2.

In our second application, we look at the distribution of single-nucleotide polymorphisms (SNPs) in genes. Since genes code for proteins, whose functions are essential to the growth and reproduction of the organism in which they occur, there are selective pressures on the mutations which can occur within a gene. In Chapter 5, we organize the occurrence of SNPs within human genes into a $3 \times 4 \times 4$ semi-symmetric contingency table. This contingency

table is tested against a mixture model with various numbers of hidden states in order to understand the interaction between phase within a codon and selective pressure. Our method for testing mixture models is based on having a determinantal representation for the secant varieties of $\mathbb{P}^2 \times \mathbb{P}^3$ embedded by the ample line bundle $\mathcal{O}(1,2)$.

Chapter 4 is devoted to the secant varieties of $\mathbb{P}^2 \times \mathbb{P}^{n-1}$ embedded by $\mathcal{O}(1,2)$. In particular, we give equations for the first 5 secant varieties of this Segre-Veronese variety, in a determinantal representation. This determinantal representation, together with the use of singular value decomposition, is what allows us to robustly test these conditions allowing them to be the basis of a statistical test in Chapter 5.

1.1 Algebraic statistics

Algebraic statistics involves the application of algebraic geometry techniques to statistical problems [19]. We start with the definition of probabilistic model of an event on a discrete state space of n events as a function $f\colon \Omega \to \Delta_{n-1}$, where Δ_{n-1} is the $(n-1)$-dimensional simplex, considered as the set of points in $\mathbb{R}^n_{\geq 0}$ whose sum is 1, and Ω is a parameter space. In algebraic statistics, we define the function f by polynomials. Likewise, Ω is a semi-algebraic set in $\mathbb{R}^n$, such as a product of simplices. One of the standard problems in algebraic statistics is to find the ideal I of all polynomials vanishing on the image of f. One case of this problem is taken up in Chapter 4.

One application of knowing these equations is to provide a way to test whether a probability distribution belongs to the image of the statistical model. Given data which comes from the model f, the counts of the different events can be normalized to give an empirical distribution which is close to being in the image of f. Consequently, the values polynomials in I will be close to zero when evaluated at this point. However, the converse isn't true: the set of probability distributions for which all polynomials in I vanish is, by definition the Zariski closure of the image of f, and the Zariski closure may be larger than the image. Over the complex numbers, the image of a polynomial map contains a dense open subset of its Zariski closure, which is thus equal to the closure in the usual Euclidean topology. However, over the real numbers, the image of f may not be dense in its Zariski closure in the Euclidean topology. In summary, the vanishing of the polynomials in I on a particular distribution means that there exist complex parameters for which this distribution can be approximated arbitrarily closely, but it may not be possible to choose these parameters to be real.

Another difficulty with using the vanishing of polynomials to test membership in a statistical model is the necessity of being robust to the presence of noise. The empirical distribution calculated from a series of observations is unlikely to exactly reflect the actual probability distribution. Therefore, using polynomial equations to test membership in a statistical model requires some way of doing it robustly. In Chapter 5, instead of using the equations of Theorem 4.4.3 directly, we leverage their origins as determinants and Pfaffians

of certain matrices. The robust estimation of the rank of a matrix can be performed using singular value decomposition, as will be described in Section 5.1. In essence, we have reduced a multi-linear problem to a linear one.

In Chapter 2, the need for robustness comes from a non-statistical model, but statistical methods are used to find an approximate solution. More specifically, we have a system of real algebraic equations and we want to find parameters for which these equations are approximately satisfied as closely as possible. Usual tools for solving polynomial equations, such as Gröbner bases or numerical homotopies, work over the complex numbers. While these methods have been developed to be numerically robust, they do not allow robustness in the sense of relaxing equality constraints to approximate equality. Thus, in Chapter 2, we develop a method in polynomial algebra which builds on statistical techniques for maximum likelihood estimation.

1.2 Maximum likelihood estimation

Maximum likelihood estimation is a method of approximately inverting the parametric model f in a statistically rigorous way. In order to do so, we suppose that we have a number of observations of the event. We can summarize these as a vector of event counts $u \in \mathbb{Z}^n_{\geq 0}$. The likelihood of such a vector is a function of the parameter $x \in \Omega$ defined by

$$L(x) = \prod_{i=1}^{n} f_i(x)^{u_i}.$$

The maximum likelihood parameter is, by definition, the point $x \in \Omega$ at which $L(x)$ achieves its maximum. Equivalent to maximizing the likelihood is maximizing the log-likelihood:

$$\ell(x) = \log L(x) = \sum_{i=1}^{n} u_i \log f_i(x).$$

This formulation is the basis for the notion of Kullback-Leibler divergence, which is generalized in (2.2). Two computational methods for maximum likelihood estimation are Expectation-Maximization (EM) and Iterative Proportional Fitting (IPF), which will together be generalized in Chapter 2 to the situation when the probability distributions are replaced by non-negative real numbers and the f_i are polynomials with non-negative coefficients.

Expectation-Maximization is an algorithm for computing maximum likelihood parameters for models which can be written in terms of some hidden state. We assume that the model f can be described in terms of first producing, as a pair, the observed state $1 \leq i \leq n$ and a hidden state $1 \leq j \leq m$. In other words, f the composition of a model $g\colon \Omega \to \mathbb{R}^{n\times m}$, followed by a summation, so that $f_i(x) = g_{i1} + \cdots + g_{im}$. The idea of EM is to alternate between using a guess of the parameters x to compute the expected distribution over the

hidden states (the E step), and using this distribution to compute the maximum likelihood parameters for the hidden model (the M step). Thus, the M step requires a method for solving the maximum likelihood problem for the hidden model. This algorithm is explicitly spelled out in Section 2.1, where IPF is used for the M step. The EM algorithm is also applied in Section 5.1, in which the hidden model is an independence model, and likelihood maximization can be trivially done in closed form. Note that EM is essentially a local search, it will converge to a local maximum, but not necessarily a global maximum.

Iterative Proportional Fitting finds maximum likelihood parameters for toric statistical models, also called log-linear models. These are where the probabilities $f_i(x)$ are products of powers of the parameters, i.e. $f_i(x) = x_1^{\alpha_1} \cdots x_k^{\alpha_k}$. Thus, the logarithm of the probability is linear in the logarithms of the parameters, explaining the name "log-linear." As its name implies, IPF is an iterative algorithm for finding maximum likelihood parameters for such a model. Note that a toric model is guaranteed to have a unique maximum likelihood solution, to which IPF will always converge [45, Thm. 1.10].

Chapter 2

Non-negative solutions to positive systems of polynomial equations

This chapter is based on the paper "An algorithm for finding positive solutions to polynomial equations" [10]. We present an iterative numerical method for finding non-negative solutions and approximate solutions to systems of polynomial equations. We require two assumptions about our system of equations. First, for each equation, all the coefficients other than the constant term must be non-negative. Second, there is a technical assumption on the exponents, described at the beginning of Section 2.1, which, for example, is satisfied if all non-constant terms have the same total degree. In Section 2.3, there is a discussion of the range of possible systems which can arise under these hypotheses.

Because of the assumption on signs, we can write our system of equations as

$$\sum_{\alpha \in S} a_{i\alpha} x^{\alpha} = b_i \quad \text{for } i = 1, \ldots, \ell, \tag{2.1}$$

where the coefficients $a_{i\alpha}$ are non-negative and the b_i are positive, and $S \subset \mathbb{R}^n_{\geq 0}$ is a finite set of possibly non-integer multi-indices. Our algorithm works by iteratively decreasing the generalized Kullback-Leibler divergence of the left-hand side and right-hand side of (2.1). The generalized Kullback-Leibler divergence of two positive vectors a and b is defined to be

$$D\left(a \,\|b\right) := \sum_i \left(a_i \log\left(\frac{a_i}{b_i}\right) - a_i + b_i \right). \tag{2.2}$$

The standard Kullback-Leibler consists only of the first term and is defined only for probability distributions, i.e. when the sum of each vector is 1. The last two terms are necessary so that the generalized divergence has, for arbitrary positive vectors a and b, the property of being non-negative and zero exactly when a and b are equal (Proposition 2.1.4).

Our algorithm converges to local minima of the Kullback-Leibler divergence, including exact solutions to the system (2.1). We will refer to local minima which are not exact

solutions as a *approximate solutions*, because in these cases, the equations (2.1) hold only approximately. In order to find multiple local minima, we can repeat the algorithm for randomly chosen starting points. For finding approximate solutions, this may be sufficient. However, there are no guarantees of completeness for the exact solutions obtained in this way. Nonetheless, we hope that in certain situations, our algorithm will find applications both for finding exact and approximate solutions.

Lee and Seung applied the EM algorithm to the problem of non-negative matrix factorization [38]. They introduced the generalized Kullback-Leibler divergence in (2.2) and used it to find approximate non-negative matrix factorizations. Since the product of two matrices can be expressed by polynomials in the entries of the matrices, matrix-factorization is a special case of the equations in (2.1).

For finding exact solutions to arbitrary systems of polynomials, there are a variety of approaches which find all complex or all real solutions. Homotopy continuation methods find all complex roots of a system of equations [51]. Even to find only positive roots, these two methods finds all complex or all real solutions, respectively. Lasserre, Laurent, and Rostalski have applied semidefinite programming to find all real solutions to a system of equations and a slight modification of their algorithm will find all positive real solutions [36, 37]. Nonetheless, neither of these methods has any notion of approximate solutions.

For directly finding only positive real solutions, Bates and Sottile have proposed an algorithm based on fewnomials bounds on the number of real solutions [2]. However, their method is only effective when the number of monomials (the set S in our notation) is slightly more than the number of variables. Our method only makes weak assumptions on the set of monomials, but stronger assumptions on the coefficients.

Our inspiration comes from tools for maximum likelihood estimation in statistics. Parameters which maximize the likelihood are exactly the parameters such that the model probabilities are closest to the empirical probabilities, in the sense of minimizing Kullback-Leibler divergence. Expectation-Maximization [45, Sec. 1.3] and Iterative Proportional Fitting [17] are well-known iterative methods for maximum likelihood estimation. We re-interpret these algorithms in the somewhat more general setting of finding approximate solutions to polynomial equations.

The impetus behind the work in this chapter was the need to find approximate positive solutions to systems of bilinear equations which will appear in Chapter 3. In that application the variables will represent expression levels, which only made sense as positive parameters. Moreover, in order to accommodate noise in the data, there are more equations than variables, so it is necessary to find approximate solutions. Thus, the algorithm described in this chapter incorporates both the positivity and the robustness.

An implementation of our algorithm in the C programming language is freely available at `http://math.berkeley.edu/~dustin/pos/`.

In Section 2.1, we describe the algorithm and the connection to maximum likelihood estimation. In Section 2.2, we prove the necessary convergence for our algorithm. Finally, in Section 2.3, we show that even with our restrictions on the form of the equations, there

can be exponentially many positive real solutions.

2.1 Algorithm

In addition to the assumption on the non-negativity of the coefficients, we make a further assumption on the exponent set. We assume that we have an $s \times n$ non-negative matrix g, with no column identically zero, and positive real numbers d_j for $1 \leq j \leq s$ such that for each $\alpha \in S$ and each $j \leq s$, the sum $\sum_{i=1}^{n} g_{ji}\alpha_i$ is either 0 or d_j. For example, if all the monomials x^α have the same total degree d_1, we can take $s = 1$ and $g_{1i} = 1$ for all i. Also of particular interest are multilinear systems of equations, in which each α_i is at most one. In these systems, the variables can be partitioned into sets such that the equations are linear in each set of variables, so we can take $d_j = 1$ for all j. Note that because d_j is in the denominator in (2.4), convergence is fastest when the d_j are small, such as in the multilinear case. We also note that, for an arbitrary set of exponents S, there may not exist such a matrix g.

The algorithm begins with a randomly chosen starting vector and iteratively improves it through two nested iterations:

- Initialize $x = (x_1, x_2, \ldots, x_n)$ with n randomly chosen positive real numbers.
- Loop until the vector x stabilizes:
  - For all $\alpha \in S$, compute
$$w_\alpha := \sum_{i=1}^{\ell} b_i \frac{a_{i\alpha}x^\alpha}{\sum_\beta a_{i\beta}x^\beta}. \tag{2.3}$$
  - Loop until the vector x stabilizes:
    - Loop for j from 1 to s:
      - Simultaneously update all entries of x:
$$x_i \leftarrow x_i \left(\frac{\sum_\alpha \alpha_i g_{ji} w_\alpha}{\sum_\alpha \alpha_i g_{ji} a_\alpha x^\alpha} \right)^{g_{ji}/d_j} \quad \text{where } a_\alpha = \sum_{i=1}^{\ell} a_{i\alpha}. \tag{2.4}$$

Because there is no subtraction, it is clear that the entries of x remain positive throughout this algorithm.

Since our algorithm is iterative, it is dependent on stopping criteria to determine when to stop the iterations. For the inner loop, we terminate when the relative change in the components of the vector x is less than the square of the relative change in the last iteration of the outer loop. In this way, the threshold becomes more stringent as the algorithm converges. For the outer loop, we use the quadratic Taylor expansion of the divergence (2.2) to estimate the distance between the current value of x and the local minimum of the divergence. When

Inner loop iterations	x	y
5	0.6944381615	0.9327435254
4	0.6339861529	0.9862879982
3	0.6217014340	0.9968624369
5	0.6187745905	0.9993672667
6	0.6181817273	0.9998738089
8	0.6180632943	0.9999749698
9	0.6180397990	0.9999950375
11	0.6180351404	0.9999990163

Table 2.1: Eight iterations of the outer loop applied to the system of equations in (2.5).

the estimated distance is less than a given threshold, the loop terminates. This criterion was chosen over more straightforward ones, which proved to be unreliable in practice.

Note that each iteration is quite fast. The computation in (2.3) is equivalent to a single evaluation of the system of polynomials, plus a division for each term. On the other hand, each update (2.4) is potentially faster than a single evaluation of the system of polynomials, especially if $d_j = 1$, because the iteration is only over S, the set of exponent vectors, regardless of the number of equations in which a term with that exponent occurs. Finally to gain additional speed, in our implementation, when the algorithm gets close to an actual solution, we use Newton's algorithm to quickly refine it.

Our method is inspired by interpreting the equations in (2.1) as a maximum likelihood problem for a statistical model and applying the well-known methods of Expectation-Maximization (EM) and Iterative Proportional Fitting (IPF). For simplicity, let us assume that all the monomials x^α have the same total degree. Our statistical model is that a hidden process generates an integer $i \leq \ell$ and an exponent vector α with joint probability $a_{i\alpha}x^\alpha$. The vector x contains n positive parameters for the model, restricted such that the total probability $\sum_{i,\alpha} a_{i\alpha}x^\alpha$ is 1. The empirical data consists of repeated observations of the integer i, but not the exponent α, and b_i is the proportion of observations of i. In this situation, the vector x which minimizes the divergence of (2.1) is the maximum likelihood parameter for the empirical distribution b_i. The inner loop of the algorithm consists of using IPF to solve the log-linear hidden model and the outer loop consists of using EM to estimate the distribution on the hidden states.

We give a simple example of our algorithm in action.

Example 2.1.1. The (reciprocal of the) golden ratio is the unique positive root of the equation

$$x^2 + x - 1 = 0.$$

The coefficients have the correct signs, but doesn't satisfy the condition on the exponent vectors. We can fix this by adding a dummy variable y, which will be forced to be equal

to 1:

$$\begin{aligned} x^2 + xy &= 1 \\ y^2 &= 1. \end{aligned} \tag{2.5}$$

A sample run took 8 iterations of the outer loop and a total of 51 iterations of the inner loop to converge to 5 digits of accuracy, as detailed in Table 2.1. □

Example 2.1.2. Next we give an example of a system with only approximate solutions. One can check that the system of equations

$$\begin{aligned} x^2 + y^2 &= 1 \\ x^2 + 2xy + y^2 &= 4 \end{aligned}$$

has no real solutions. On this system, our algorithm converges to the approximate solution $(x, y) = (\sqrt{5/6}, \sqrt{5/6}) \approx (0.91287, 0.91287)$. With these values, $x^2 + y^2 = 5/3$, which is somewhat greater than 1, and $x^2 + 2xy + y^2 = 10/3$, which is somewhat less than 4. □

Although the Kullback-Leibler divergence is rarely used outside probability and statistics, it can be approximated as a weighted L^2-distance. In order to make this statement precise, we define the function $C(t)$ to be 1 if $t \leq 1$ and to be $\log(t)/(t-1)$ if $t > 1$. Note that $C(t)$ is approximately equal to 1 in a neighborhood of $t = 1$, i.e., when comparing vectors which are close to each other.

Proposition 2.1.3. *With a and b two positive real vectors, and $C(t)$ as defined above, then $D(a\,\|b)$ is bounded below by the square of the weighted L^2-distance*

$$\sum_{i=1}^{n} \frac{C(a_i/b_i)}{2b_i}(a_i - b_i)^2 \tag{2.6}$$

Proposition 2.1.3 implies that, at least for nearby vectors, the major difference between divergence and Euclidean distance is that divergence places more weight on components whose values are closer to zero.

Proof of Proposition 2.1.3. Note that $D(a\,\|b) = \sum_{i=1}^{n} D(a_i\,\|b_i)$ and thus, it suffices to prove the statement when a and b are scalars. We let $t = a/b$, and then,

$$D(a\,\|b) = a\log\left(\frac{a}{b}\right) - a + b = b(t\log t - t + 1) = b\int_1^t \log s\, ds.$$

Now we bound the integral using a linear approximation of the logarithm. For $s \leq 1$, $\log s \leq s - 1$. On the other hand, for $1 \leq s \leq t$, $\log s \geq (s-1)\log(t)/(t-1)$, since log is a convex function. Therefore, regardless of whether $t \leq 1$ or $t \geq 1$, we have

$$\int_1^t \log s\, ds \geq C(t)\int_1^t (s-1)\, ds.$$

Combining this inequality with (2.1),

$$D\left(a\,\|b\right) \geq C(t)b\int_1^t (s-1)\,ds = \frac{C(a/b)b}{2}\left(\frac{a}{b}-1\right)^2 = \frac{C(a/b)}{2b}(a-b)^2,$$

which is the desired result. □

As a corollary, we get:

Proposition 2.1.4. *For vectors a and b of positive real numbers, the divergence $D\left(a\,\|b\right)$ is always non-negative with $D\left(a\,\|b\right) = 0$ if and only if $a = b$.*

Proof. Since the quantity $C(a_i/b_i)$ from Proposition 2.1.3 is always positive, each term of the summation in (2.6) is non-negative and it is zero if and only $a_i = b_i$. □

2.2 Proof of convergence

In this section we prove our main theorem:

Theorem 2.2.1. *The Kullback-Leibler divergence*

$$\sum_{i=1}^{\ell} D\left(b_i \left\| \sum_{\alpha\in S} a_{i\alpha}x^\alpha\right.\right). \tag{2.7}$$

is weakly decreasing during the algorithm in Section 2.1. Moreover, assuming that the set S contains a multiple of each unit vector e_i, i.e. some power of each x_i appears in the system of equations, then the vector x converges to a critical point of the function (2.7) or the boundary of the positive orthant.

Remark 2.2.2. *The condition that S contains a pure power of each variable x_i is in order to ensure that the vector x remains bounded during the algorithm.*

We begin by establishing several basic properties of the generalized Kullback-Leibler divergence in Lemmas 2.2.3 and 2.2.4. The proof of Theorem 2.2.1 itself is divided into two parts, corresponding to the two nested iterative loops. The first step is to prove that the updates (2.4) in the inner loop converge a local minimum of the divergence $D\left(w_\alpha\,\|a_\alpha x^\alpha\right)$. The second step is to show that this implies that the outer loop strictly decreases the divergence function (2.7), with equality only at a critical point.

Lemma 2.2.3. *Suppose that a and b are vectors of n positive real numbers. Let t be any positive real number, and then*

$$D\left(a\,\|tb\right) = D\left(a\,\|b\right) + (t-1)\sum_{i=1}^{m} b_i - \sum_{i=1}^{m} a_i \log t$$

Proof. As in the proof of Proposition 2.1.3, we can assume that a and b are scalars. In this case, it becomes a straightforward computation. □

Lemma 2.2.4. *If a and b are vectors of m positive real numbers, then we can relate their divergence to the divergence of their sums by*

$$D\left(a \,\|b\right) = D\left(\textstyle\sum_{i=1}^m a_i \,\|\sum_{i=1}^m b_i\right) + D\left(a \,\middle\|\, \frac{\sum_{i=1}^m a_i}{\sum_{i=1}^m b_i} b\right).$$

Proof. We let $A = \sum_{i=1}^m a_i$ and $B = \sum_{i=1}^m b_i$, and apply Lemma 2.2.3 to the last term:

$$\begin{aligned} D\left(a \,\middle\|\, \frac{A}{B} b\right) &= D\left(a \,\|b\right) + \left(\frac{A}{B} - 1\right) B - A \log \frac{A}{B} \\ &= D\left(a \,\|b\right) - A \log \frac{A}{B} + A - B = D\left(a \,\|b\right) - D\left(A \,\|B\right). \end{aligned}$$

After rearranging, we get the desired expression. □

Lemma 2.2.5. *The update rule (2.4) weakly decreases the divergence. If*

$$x_i' = x_i \left(\frac{\sum_\alpha \alpha_i g_{ji} w_\alpha}{\sum_\alpha \alpha_i g_{ji} a_\alpha x^\alpha}\right)^{g_{ji}/d_j},$$

for any $j \le s$, then

$$\sum_\alpha D\left(w_\alpha \,\|a_\alpha (x')^\alpha\right) \le \sum_\alpha D\left(w_\alpha \,\|a_\alpha x^\alpha\right) - \frac{1}{d_j} \sum_{i=1}^n D\left(\sum_\alpha \alpha_i g_{ji} w_\alpha \,\middle\|\, \sum_\alpha \alpha_i g_{ji} a_\alpha x^\alpha\right). \quad (2.8)$$

Proof. First, since the statement only depends on the jth row of the matrix g, we can assume that g is a row vector and we drop j from future subscripts. Second, we can assume that $d = 1$ by replacing g_i with g_i/d.

Third, we reduce to the case when $g_i = 1$ for all i. We define a new set of exponents $\tilde{\alpha}$ and coefficients $\tilde{a}_{\tilde{\alpha}}$ by $\tilde{\alpha}_i = g_i \alpha_i$ and $\tilde{a}_\alpha = a_\alpha \prod x_i$, where the product is taken over all indices i such that $g_i = 0$. We take $\tilde{x}$ to be a vector indexed by those i such that $g_i \neq 0$. Then, under the change of coordinates $\tilde{x}_i = x_i^{1/g_i}$, we have $a_\alpha x^\alpha = \tilde{a}_\alpha \tilde{x}^{\tilde{\alpha}}$ and the update rule in (2.4) is the same for the new system with coefficients $\tilde{a}_{\tilde{\alpha}}$ and exponents $\tilde{\alpha}$. Furthermore, if all entries of $\tilde{\alpha}$ are zero, then $\tilde{x}^{\tilde{\alpha}} = 1$ for all vectors x and so we can drop $\tilde{\alpha}$ from our exponent set. Therefore, for the rest of the proof, we drop the tildes, and assume that $\sum_i \alpha_i = 1$ for all $\alpha \in S$ and $g_i = 1$ for all i, in which case g drops out of the equations.

To prove the desired inequality, we substitute the updated assignment x' into the definition of Kullback-Leibler divergence:

$$\begin{aligned} D\left(w_\alpha \,\|a_\alpha (x')^\alpha\right) &= w_\alpha \log \left(\frac{w_\alpha}{a_\alpha x^\alpha \prod_{i=1}^n \left(\frac{\sum_\beta w_\beta}{\sum_\beta \beta_i a_\beta x^\beta}\right)^{\alpha_i}} \right) - w_\alpha + a_\alpha (x')^\alpha \\ &= w_\alpha \log \frac{w_\alpha}{a_\alpha x^\alpha} - \sum_{i=1}^n \alpha_i w_\alpha \log \frac{\sum_\beta \beta_i w_\beta}{\sum_\beta \beta_i a_\beta x^\beta} - w_\alpha + a_\alpha (x')^\alpha \\ &= D\left(w_\alpha \,\|a_\alpha x^\alpha\right) - \sum_{i=1}^n \alpha_i w_\alpha \log \frac{\sum_\beta \beta_i w_\beta}{\sum_\beta \beta_i a_\beta x^\beta} - a_\alpha x^\alpha + a_\alpha (x')^\alpha. \end{aligned} \tag{2.9}$$

On the other hand, let C denote the last term of (2.8), which we can expand as,

$$\begin{aligned} C &= \sum_{i=1}^n D\left(\sum_\alpha \alpha_i w_\alpha \middle\| \sum_\alpha \alpha_i a_\alpha x^\alpha \right) \\ &= \sum_{i=1}^n \left(\left(\sum_\alpha \alpha_i w_\alpha \right) \log \frac{\sum_\alpha \alpha_i w_\alpha}{\sum_\alpha \alpha_i a_\alpha x^\alpha} - \sum_\alpha \alpha_i w_\alpha + \sum_\alpha \alpha_i a_\alpha x^\alpha \right) \\ &= \sum_{i=1}^n \sum_\alpha \left(\alpha_i w_\alpha \log \frac{\sum_\beta \beta_i w_\beta}{\sum_\beta \beta_i a_\beta x^\beta} - \alpha_i w_\alpha + \alpha_i a_\alpha x^\alpha \right) \\ &= \sum_\alpha \left(\sum_{i=1}^n \alpha_i w_\alpha \log \frac{\sum_\beta \beta_i w_\beta}{\sum_\beta \beta_i a_\beta x^\beta} \right) - w_\alpha + a_\alpha x^\alpha, \end{aligned} \tag{2.10}$$

where the last step follows from the assumption that that $\sum_i \alpha_i = 1$ for all $\alpha \in S$. We take the sum of (2.9) over all $\alpha \in S$ and add it to (2.10) to get,

$$\sum_\alpha D\left(w_\alpha \,\|a_\alpha (x')^\alpha\right) + C = \sum_\alpha D\left(w_\alpha \,\|a_\alpha x^\alpha\right) - \sum_\alpha b_\alpha + \sum_\alpha a_\alpha (x')^\alpha. \tag{2.11}$$

Finally, we expand the last term of (2.10) using the definition of x' and apply the

arithmetic-geometric mean inequality,

$$\begin{aligned}\sum_\alpha a_\alpha (x')^\alpha &= \sum_\alpha a_\alpha x^\alpha \prod_{i=1}^n \left(\frac{\sum_\beta \beta_i b_\beta}{\sum_\beta \beta_i a_\beta x^\beta} \right)^{\alpha_i} \\ &\leq \sum_\alpha a_\alpha x^\alpha \sum_{i=1}^n \alpha_i \frac{\sum_\beta \beta_i b_\beta}{\sum_\beta \beta_i a_\beta x^\beta} \\ &= \sum_{i=1}^n \left(\sum_\alpha \alpha_i a_\alpha x^\alpha \right) \frac{\sum_\beta \beta_i b_\beta}{\sum_\beta \beta_i a_\beta x^\beta} \\ &= \sum_{i=1}^n \sum_\beta \beta_i b_\beta = \sum_\beta b_\beta.\end{aligned}$$

Together with (2.11), this gives the desired inequality. □

Proposition 2.2.6. *A positive vector x is a fixed point of the update rule (2.4) for all $1 \leq j \leq s$ if and only if x is a critical point of the divergence function $\sum_\alpha D\left(w_\alpha \| a_\alpha x^\alpha\right)$.*

Proof. For the update rule to be constant means that the numerator and denominator in (2.4) are equal, i.e.

$$\sum_\alpha \alpha_i g_{ji} a_\alpha x^\alpha = \sum_\alpha \alpha_i g_{ji} w_\alpha \quad \text{for all } i \text{ and } j. \tag{2.12}$$

By our assumption on g, for each i, some g_{ji} is non-zero, so (2.12) is equivalent to

$$\sum_\alpha \alpha_i a_\alpha x^\alpha = \sum_\alpha \alpha_i w_\alpha \quad \text{for all } i. \tag{2.13}$$

On the other hand, we compute the partial derivative

$$\frac{\partial}{\partial x_i} \sum_\alpha D\left(w_\alpha \| a_\alpha x^\alpha\right) = \sum_\alpha -w_\alpha \frac{\alpha_i}{x_i} + \alpha_i a_\alpha \frac{x^\alpha}{x_i}.$$

Since each x_i is assumed to be non-zero, it is clear that all partial derivatives being zero is equivalent to (2.13). □

Lemma 2.2.7. *If we define w_α as in (2.3), then*

$$\begin{aligned}\sum_{i=1}^n D\left(b_i \middle\| \sum_\alpha a_{i\alpha} (x')^\alpha\right) - \sum_{i=1}^n D\left(b_i \middle\| \sum_\alpha a_{i\alpha} x^\alpha\right) \\ \leq \sum_\alpha D\left(w_\alpha \| a_\alpha (x')^\alpha\right) - \sum_\alpha D\left(w_\alpha \| a_\alpha x^\alpha\right).\end{aligned}$$

Moreover, a positive vector x is a fixed point of the update rule if and only if x is a critical point for the divergence function.

Proof. We consider

$$\sum_{i,\alpha} D\left(w_{i\alpha} \,\|a_{i\alpha}(x')^\alpha\right) \qquad \text{where } w_{i\alpha} = \frac{b_i a_{i\alpha} x^\alpha}{\sum_\beta a_{i\beta} x^\beta}, \tag{2.14}$$

and apply Lemma 2.2.4 in two different ways. First, by applying Lemma 2.2.4 to each group of (2.14) with fixed α, we get

$$\sum_{i,\alpha} D\left(w_{i\alpha} \,\|a_{i\alpha}(x')^\alpha\right) = \sum_\alpha D\left(w_\alpha \,\|a_\alpha(x')^\alpha\right) + \sum_{i,\alpha} D\left(w_{i\alpha} \left\| \frac{\sum_j w_{j\alpha}}{\sum_j a_\alpha(x')^\alpha} a_{i\alpha}(x')^\alpha \right.\right).$$

In the last term, the monomials $(x')^\alpha$ cancel and so it is a constant independent of x' which we denote E. On the other hand, we can apply Lemma 2.2.4 to each group in (2.14) with fixed i,

$$\sum_{i,\alpha} D\left(w_{i\alpha} \,\|a_{i\alpha}(x')^\alpha\right) = \sum_i D\left(b_i \left\| \sum_\alpha a_{i\alpha}(x')^\alpha \right.\right) + \sum_{i,\alpha} D\left(w_{i\alpha} \left\| \frac{b_i a_{i\alpha}(x')^\alpha}{\sum_\beta a_{i\beta}(x')^\beta} \right.\right).$$

We can combine these equations to get

$$\sum_i D\left(b_i \left\| \sum_\alpha a_{i\alpha}(x')^\alpha \right.\right) = \sum_\alpha D\left(w_{i\alpha} \,\|a_\alpha(x')^\alpha\right) + E - \sum_{i,\alpha} D\left(w_{i\alpha} \left\| \frac{b_i a_{i\alpha}(x')^\alpha}{\sum_\beta a_{i\beta}(x')^\beta} \right.\right). \tag{2.15}$$

By Proposition 2.1.4, the last term of (2.15) is non-negative, and by the definition of $w_{i\alpha}$, it is zero for $x' = x$. Therefore, any value of x' which decreases the first term compared to x will also decrease the left hand side by at least as much, which is the desired inequality.

In order to prove the statement about the derivative, we consider the derivative of (2.15) at $x' = x$. Because the last term is minimized at $x' = x$, its derivative is zero, so

$$\left.\frac{\partial}{\partial x'_j}\right|_{x'=x} \sum_i D\left(b_i \left\| \sum_\alpha a_{i\alpha}(x')^\alpha \right.\right) = \left.\frac{\partial}{\partial x'_j}\right|_{x'=x} \sum_{i,\alpha} D\left(w_{i\alpha} \,\|a_{i\alpha}(x')^\alpha\right).$$

By Proposition 2.2.6, a positive vector x is a fixed point of the inner loop if and only if these partial derivatives on the right are zero for all indices j, which is the definition of a critical point. □

Proof of Theorem 2.2.1. The Kullback-Leibler divergence $\sum_\alpha D\left(w_\alpha \,\|a_\alpha x^\alpha\right)$ decreases at each step of the inner loop by Lemma 2.2.5. Thus, by Lemma 2.2.7, the divergence

$$\sum_{i=1}^n D\left(b_i \left\| \sum_\alpha a_{i\alpha} x^\alpha \right.\right) \tag{2.16}$$

decreases at least as much. However, the divergence (2.16) is non-negative according to Proposition 2.1.4. Therefore, the magnitude of the decreases in divergence must approach zero over the course of the algorithm. By Lemma 2.2.5, this means that the quantity C in that theorem must approach zero. By Proposition 2.1.4, this means that the quantities in that divergence approach each other. However, up to a power, these are the numerator and denominator of the factor in the update rule (2.4), so the difference between consecutive vectors x approaches zero.

Thus, we just need to show that x remains bounded. However, since some power of each variable x_i occurs in some equation, as x_i gets large, the divergence for that equation also gets arbitrarily large. Therefore, each x_i must remain bounded, so the vector x must have a limit as the algorithm is iterated. If this limit is in the interior of the positive orthant, then it must be a fixed point. By Lemma 2.2.7 and Proposition 2.2.6, this fixed point must be a critical point of the divergence (2.7). □

2.3 Universality

Although the restriction on the exponents and especially the positivity of the coefficients seem like strong conditions, such systems can nonetheless be quite complex. In this section, we investigate the breadth of such equations.

Proposition 2.3.1. *For any system of ℓ real polynomial equations in n variables, there exists a system of $\ell + 1$ polynomial equations in $n + 1$ variables, in the form (2.1), such that the positive solutions $(x_1, \ldots, x_n)$ to the former system are in bijection with the positive solutions $(x'_1, \ldots, x'_{n+1})$ of the latter, with $x'_i = x_i/x_{n+1}$.*

Proof. We write our system of equations as $\sum_{\alpha \in S} a_{i\alpha} x^\alpha = 0$ for $1 \leq i \leq \ell$, where $S \subset \mathbb{N}^n$ is an arbitrary finite set of exponents and $a_{i\alpha}$ are any real numbers. We let d be the maximum degree of any monomial x^α for $\alpha \in S$. We homogenize the equations with a new variable x_{n+1}. Explicitly, define $S' \subset \mathbb{N}^{n+1}$ to consist of $\alpha' = (\alpha, d - \sum_i \alpha_i)$ for all α in S and we also set $a_{i\alpha'} = a_{i\alpha}$. We add a new equation with coefficients $a_{\ell+1,\alpha} = 1$ for all $\alpha \in S'$ and $b_{\ell+1} = 1$. For this system, we can clearly take $g_{1i} = 1$ and $d_1 = d$ to satisfy the condition on exponents. Furthermore, for any positive solution $(x_1, \ldots, x_n)$ to the original system of equations, $(x'_1, \ldots, x'_{n+1})$ with $x'_i = x_i / \big(\sum_\alpha x^\alpha\big)^{1/d}$ and $x'_{n+1} = 1/\big(\sum_\alpha x^\alpha\big)^{1/d}$ is a solution to the homogenized system of equations.

Next, we add a multiple of the last equation to each of the others in order to make all the coefficients positive. For each $1 \leq i \leq \ell$, choose a positive $b_i > -\min_\alpha\{a_{i\alpha} \mid \alpha \in S'\}$, and define $a'_{i\alpha} = a_{i\alpha} + b_i$. By construction, the resulting system has all positive coefficients, and since the equations are formed from the previous equations by elementary linear transformations, the set of solutions are the same. □

The practical use of the construction in the proof of Proposition 2.3.1 is mixed. The first step, of homogenizing to deal with arbitrary sets of exponents, is a straightforward way of

guaranteeing the existence of the matrix g. However, for large systems, the second step tends to produce an ill-conditioned coefficient matrix. In these cases, our algorithm converges very slowly. Nonetheless, Proposition 2.3.1 shows that, in the worst case, systems satisfying our hypotheses can be as complicated as arbitrary polynomial systems.

Proposition 2.3.2. *There exist bilinear equations in $2m$ variables with $\binom{2m-2}{m-1}$ positive real solutions.*

Proof. We use a variation on the technique used to prove Proposition 2.3.1.

First, we pick $2m-2$ generic homogeneous linear functions $b_1, \ldots, b_{2m-2}$ on m variables. By generic, we mean for any m of the b_k, the only simultaneous solution of all m linear equations is the trivial one. This genericity implies that any $m-1$ of the b_k define a point in $\mathbb{P}^{m-1}$ By taking a linear changes of coordinates in each set of variables, we can assume that all of these points are positive, i.e. have a representative consisting of all positive real numbers.

Then we consider the system of equations

$$b_k(x_1, \ldots, x_m) \cdot b_k(x_{m+1}, \ldots, x_n) = 0, \text{ for } 1 \leq k \leq 2m-2 \tag{2.17}$$

$$(x_1 + \ldots + x_m)(x_{m+1} + \ldots + x_{2m}) = 1 \tag{2.18}$$

$$x_1 + \ldots + x_m = 1. \tag{2.19}$$

The equations (2.17) are bihomogeneous and so we can think of their solutions in $\mathbb{P}^{m-1} \times \mathbb{P}^{m-1}$. There are exactly $\binom{2m-2}{m-1}$ positive real solutions, corresponding to the subsets $A \subset [2m-2]$ of size $m-1$. For any such A, there is a unique, distinct solution satisfying $b_k(x_1, \ldots, x_m) = 0$ for all k in A and $b_k(x_{m+1}, \ldots, x_{2m}) = 0$ for all k not in A. By assumption, for each solution, all the coordinates can be chosen to be positive. The last two equations (2.18) and (2.19) dehomogenize the system in a way such that there are $\binom{2m-2}{m-1}$ positive real solutions. Finally, as in the last paragraph of the proof of Proposition 2.3.1, we can add multiples of (2.18) to the equations (2.17) in order to make all the coefficients positive. □

Example 2.3.3. We illustrate the construction in Proposition 2.3.2 in the simplest case, when $m = 2$, and $n = 2m = 4$. We take the two homogenous functions $b_1 = x_1 - x_2$ and $b_2 = x_1 - 2x_2$, which each have a positive real solution, as desired. Following (2.17), we have the bilinear equations equations:

$$(x_1 - x_2)(x_3 - x_4) = x_1x_3 - x_1x_4 - x_2x_3 + x_2x_4 = 0$$
$$(x_1 - 2x_2)(x_3 - 2x_4) = x_1x_3 - 2x_1x_4 - 2x_2x_3 + 4x_2x_4 = 0$$

We add appropriate multiples of $(x_1 + x_2)(x_3 + x_4) = 1$ to these two equations, to get the

system with non-negative coefficients:

$$\begin{aligned} 2x_1x_3 + 2x_2x_4 &= 1 \\ 3x_1x_3 + 6x_2x_4 &= 2 \\ x_1x_3 + x_1x_4 + x_2x_3 + x_2x_4 &= 1 \\ x_1 + x_2 &= 1. \end{aligned}$$

The two solutions to these equations are $(1/2, 1/2, 2/3, 1/3)$ and $(1/3, 2/3, 1/2, 1/2)$. Because of the symmetry, our algorithm will converge to each half of the time. □

Chapter 3

Gene expression in Arabidopsis roots

This chapter is based on the paper "Reconstructing spatiotemporal gene expression data from partial observations," which was jointly authored with Siobhan M. Brady, David A. Orlando, Bernd Sturmfels, and Philip N. Benfey [12].

Transcriptional regulation plays an important role in orchestrating a host of biological processes, particularly during development (reviewed in [27, 40]). Advances in microarray and sequencing technologies have allowed biologists to capture genome-wide gene expression data; the output of this transcriptional regulation. This expression data can then be used to identify genes whose expression is correlated with a particular biological process, and to identify transcriptional regulators that coordinate the expression of groups of genes that are important for the same biological process.

The identification of such genes and transcriptional regulators is complicated by the complex heterogeneous mixture of cell types and developmental stages that comprise each organ of an organism. Expression patterns that are found only in a subset of cell types within an organ will be diluted and may not be detectable in the collection of expression patterns obtained from RNA isolated from samples of an entire organ. Therefore techniques have been developed to enrich samples for specific cell types or developmental stages, especially for studies in plants [9]. In the model plant, *Arabidopsis thaliana*, several features of the root organ reduce its developmental complexity and facilitate analysis. Specifically, most root cell types are found within concentric cylinders moving from the outside of the root to the inside of the root (Figure 3.1). These cell type layers display rotational symmetry thus simplifying the spatial features of development. This feature has been exploited in the development of a cell type enrichment method. This enrichment method uses green fluorescent protein (GFP)-marked transgenic lines and fluorescently-activated cell sorting (FACS) to collect cell type enriched samples and has allowed for the identification of cell type-specific expression patterns [4, 5]. Using this technique, high resolution expression data have been obtained for nearly all cell types in the *Arabidopsis* root (herein called the *marker-line dataset*) [7, 28].

Another feature that makes the *Arabidopsis* root a tractable developmental model is that cell types are constrained in files along the root's longitudinal axis, and most of these cells

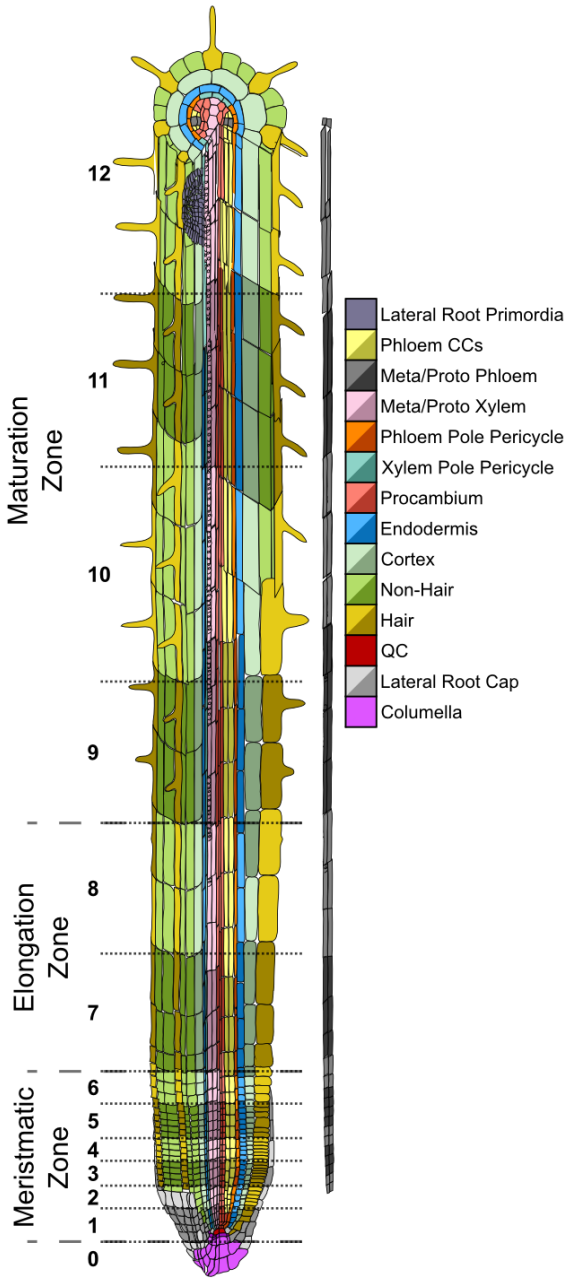

Figure 3.1: Regions and cell types in the structure of the Arabidopsis root.

are produced from a stem cell population found at the apex of the root. This feature allows a cell's developmental timeline to be represented by its position along the length of the root. To obtain a developmental time-series expression dataset individual *Arabidopsis* roots were sectioned into thirteen pieces, each piece representing a developmental time point (herein called the *longitudinal dataset*) [7]. Each of these sections, however, contains a mixture of cell types, and the microarray expression values obtained are therefore the average of the expression levels over multiple cell types present at these specific developmental time points.

While the 19 fluorescently marked lines in [7] cover expression in nearly all cell types, they do not comprehensively mark all developmental stages of these cell types. Also, the procambium cell type was not measured, as a fluorescent marker-line that marks that cell type did not exist at the time. However, expression from the longitudinal dataset, does contain averaged expression of all cell types, and may be used to infer the missing cell type data.

Previous studies have looked at separating expression data from the heterogeneous cell populations that make up tumors into the contributions of their constituent cell types [25, 50]. However, in that context, the difficulty comes from the fact that the mixture of cell types in each sample is unknown, whereas within our experimental context, the cell type mixture of each sample is known. Two computational methods have been developed to combine the *Arabidopsis* longitudinal and marker-line datasets as experimentally resolving this expression with marker lines is nearly impossible [7, 14]. However, neither method takes all data into account when reconstructing expression. In [7], only high relative gene expression is considered, and in [14], no attempt is made to infer expression for cells not covered by any marker-line.

In this work we formulate a model for expression levels in *Arabidopsis* roots in which cell type and developmental stage are independent sources of variation. The microarray data specifying overall expression levels for certain mixtures of cells lead to an overconstrained system of bilinear equations. Moreover, due to the nature of the problem, we are exclusively interested in positive real solutions. We present a new method for finding non-negative real approximate solutions to bilinear equations, based on the techniques of expectation maximization (EM) [45, Sec. 1.3] and iterative proportional fitting (IPF) [17] from likelihood maximization in statistics. Earlier work has used expectation maximization to find non-negative matrix factorizations [38], and our method is a generalization of that work.

We applied our method to estimate spatiotemporal subregion expression patterns for 20,872 *Arabidopsis* transcripts. These patterns have identified gene expression in cell types and developmental stages which were previously unknown. Visualizations of these patterns on a schematic *Arabidopsis* root are available at http://www.arexdb.org/.

Cell type	Marker-lines
Quiescent center	AGL42, RM1000, SCR5
Columella	PET111
Lateral root cap	LRC
Hair cell	COBL9 (8-13)
Non-hair cell	GL2
Cortex	J0571, CORTEX (7-13)
Endodermis	J0571, SCR5
Xylem pole pericycle	WOL (2-9), JO121 (9-13), J2661 (13)
Phloem pole pericycle	WOL (2-9), S17 (8-13), J2661 (13)
Phloem	S32, WOL (2-9)
Phloem companion cells	SUC2 (10-19), WOL (2-9)
Xylem	S4 (2-7), S18 (8-13), WOL (2-9)
Lateral root primordia	RM1000
Procambium	WOL (2-9)

Table 3.1: The 14 cell types in the *Arabidopsis* root and the 17 marker-lines which mark them [7]. For markers that only mark the cell type in some of the sections, these sections are indicated by the range in parenthesis.

3.1 Methods

Expression data

Our method uses the normalized expression data collected in [7]. Expression levels were measured across 13 longitudinal sections in a single root (*longitudinal dataset*) and across 19 different markers (*marker-line dataset*). For simplicity, the J2501 line was removed from further analysis as it is redundant with the WOODEN-LEG marker-line. The APL marker-line was also removed, as it contains domains of expression marked by both the S32 and the SUC2 marker-lines and adds no extra information. The remaining 17 markers covering 14 cell types are listed in the second column of Table 3.1.

Due to computational constraints, the original normalization of this data was performed for the longitudinal and the marker-line datasets independently [7]. In order to account for differences caused by these separate normalization procedures, we adjusted the marker-line data by a global factor of 0.92. This factor was calculated by comparing the expression values of ubiquitous, evenly expressed probe sets between the two datasets. We assume that by comparing these probe sets, any true expression differences due to cell type and longitudinal section specificity should be minimal and thus any differences in expression

level is a byproduct of the separate normalization procedures. A set of 43 probesets were identified which were expressed ubiquitously (above a normalized value of 1.0 in all samples) and whose expression did not vary significantly among samples within a dataset (ratio of min/max expression within a dataset is at most 0.5). The scaling factor necessary to make the mean expression within the marker-line dataset equal to the mean expression within the longitudinal dataset was calculated for each probe set in this set. The median value of these 43 scaling factors was 0.92, which was used as the global adjustment factor.

Model

To model the transcript expression level of an individual cell we assume that the effects of its cell type and its section on its expression level are independent of each other. More precisely, we assume that the transcript expression level of a cell of type j in section i is equal to the product $x_i \cdot y_j$, where x_i depends only on the section and y_j depends only on the cell type. In other words, for each transcript, there is an idealized profile of expression over different cell types, and an idealized profile of expression over different sections. Within a given section, our assumption is that the transcript expression level varies proportionally to its cell type profile, and within a given cell type, proportionally to its longitudinal profile.

Although our model's assumption of independence between cell type and longitudinal section is a simplification, we believe it is appropriate for two reasons. First, we expect that in most cases, a transcript controlling development will have a single temporal pattern in the cell type or cell types in which it is active and have negligible expression elsewhere. This profile is consistent with our model by taking y_j to be either high or almost zero depending on whether or not the transcript is expressed in that cell type. Second, the input longitudinal and marker-line data sets correspond roughly to independent measurements of the temporal and cell type profiles of expression level. Thus, fitting independent temporal and cell type profiles is less speculative than a more complicated model would be in the absence of more detailed data.

Each microarray sample in the two datasets (described in the above section titled "Expression data") is composed of a distinct mixture of cell types and sections. Within each sample, the measured transcript expression level is a convex linear combination of the expression levels of its constituent cells. Under the above independence assumption, the longitudinal measurements give us a system of 13 equations,

$$\sum_{j=1}^{14} a'_{kj} x_k y_j = b_k \quad \text{for } k = 1, \ldots, 13, \tag{3.1}$$

where x_k and y_j are the model parameters for the 13 sections and 14 cell types respectively, a'_{kj} is the proportion of the jth cell type in the kth section, and b_k is the measured expression level in the kth longitudinal section. The coefficients a'_{kj} form a 13×14 matrix, where the

$$\begin{pmatrix}
0 & 24 & 51 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
4 & 12 & 152 & 24 & 48 & 12 & 12 & 12 & 22 & 0 & 0 & 12 & 0 & 28 \\
0 & 0 & 280 & 40 & 80 & 40 & 40 & 20 & 45 & 20 & 20 & 25 & 0 & 80 \\
0 & 0 & 210 & 40 & 80 & 40 & 40 & 20 & 45 & 20 & 20 & 25 & 0 & 80 \\
0 & 0 & 210 & 40 & 80 & 40 & 40 & 20 & 45 & 20 & 20 & 25 & 0 & 80 \\
0 & 0 & 210 & 40 & 80 & 40 & 40 & 20 & 45 & 20 & 20 & 25 & 0 & 80 \\
0 & 0 & 0 & 40 & 80 & 40 & 40 & 20 & 45 & 20 & 20 & 25 & 0 & 80 \\
0 & 0 & 0 & 40 & 80 & 40 & 40 & 20 & 45 & 20 & 20 & 25 & 0 & 80 \\
0 & 0 & 0 & 40 & 80 & 40 & 40 & 20 & 45 & 20 & 20 & 25 & 0 & 80 \\
0 & 0 & 0 & 40 & 80 & 40 & 40 & 20 & 45 & 20 & 20 & 25 & 0 & 80 \\
0 & 0 & 0 & 40 & 80 & 40 & 40 & 20 & 45 & 20 & 20 & 25 & 0 & 80 \\
4 & 0 & 0 & 40 & 80 & 40 & 40 & 20 & 45 & 20 & 20 & 25 & 130 & 80 \\
0 & 0 & 0 & 40 & 80 & 40 & 40 & 20 & 45 & 20 & 20 & 25 & 0 & 80
\end{pmatrix}$$

Table 3.2: The cell count matrix gives the number of cells in each spatiotemporal subregion. The 13 rows correspond to longitudinal sections 1 through 13. From left to right, the 14 columns correspond to the following spatiotemporal subregions: quiescent center, columella, lateral root cap, hair cell, non-hair cell, cortex, endodermis, xylem pole pericycle, phloem pole pericycle, phloem, phloem companion cells, xylem, lateral root primordia, and procambium.

kth row is proportional to the kth row of the cell-count matrix (Table 3.2), but rescaled to sum to 1.

In addition, the 17 marker-line measurements give the additional equations,

$$\sum_{i=1}^{13}\sum_{j=1}^{14} a_{ijk}x_i y_j = b_k \quad \text{for } k = 14, \ldots, 30, \tag{3.2}$$

where the b_k are the measured expression level of the 17 markers, indexed with k from 14 through 30, and a_{ijk} is the proportion of the cells in the kth marker-line which come from the ith section and the jth cell type. For each marker-line k, we denote by a_{**k} the matrix which has a_{ijk} in the ith row and jth column. The matrix a_{**k} is zero except for those spatiotemporal subregions marked by that marker as indicated in Table 3.1. The non-zero entries of a_{**k} are proportional to the corresponding entries of the cell matrix, but rescaled to sum to 1.

In order to combine (3.1) and (3.2) in a uniform system of equations, we define we define $a_{kjk} = a'_{kj}$ for $k \leq 13$ and $a_{ijk} = 0$ for $i \neq k \leq 13$. Then (3.1) and (3.2) together give

$$\sum_{i=1}^{13}\sum_{j=1}^{14} a_{ijk}x_i y_j = b_k \quad \text{for } k = 1, \ldots, 30 \tag{3.3}$$

Cell matrix

As described in the previous section, the coefficients a_{ijk} in our model depend on the number of cells in each spatiotemporal subregion. These cell number estimates were generated by visual inspection of successive optical cross-sections of *Arabidopsis* roots along the longitudinal axis using confocal laser scanning microscopy. For the xylem, phloem and procambium cell types, cell counts were obtained from earlier experiments [6, 42]. What follows is a detailed description of this visual and literature analysis. These results are also summarized in Table 3.2.

Longitudinal section 1 encompasses two tiers of 12 columella cells, and three tiers of lateral root cap cells (15, 18 and 18 moving up from the tip).

Longitudinal section 2 contains one tier of 12 columella cells and six tiers of lateral root cap cells (20, 20, 28, 28, 28 and 28 moving up from the tip). For all other cell types in longitudinal section 2, three tiers of cells are present. Eight trichoblast (hair cell precursor) cells and 16 atrichoblast (non-hair cell precursor) cells are present circumferentially throughout the root, resulting in 24 and 48 cells respectively in the hair cell and non-hair cell precursor files in longitudinal section 2. Throughout the root, eight cortex and eight endodermis cells are present circumferentially. However in longitudinal section 2, the cortex/endodermis initial is undergoing asymmetric periclinal divisions to produce the cortex and endodermis cell files, so we consider there to be approximately 0.5 cells of the cortex and endodermis type, resulting in 12 cells of each type in longitudinal section 2. When the *Arabidopsis* root is seven days old, each longitudinal section from 3–13 contains approximately five cells of each type along the root's longitudinal axis.

In longitudinal section 2, the tangential and periclinal divisions that give rise to phloem cell files do not occur, but do occur in longitudinal section 3 [6]. Three cells are present in the main xylem axis in the first tier of cells, four cells in the second tier, and five cells in the third tier [42]. Eight procambial cells are present in the first cell tier, 12 procambial cells in the second tier, and 18 cells in the third tier resulting in 28 procambial cells in longitudinal section 2 [42]. For all sections xylem pole pericycle cells are the two cells that flank the xylem axis on either end, and phloem pole pericycle cells are considered the intervening cells. Four pericycle cells can be identified as flanking xylem cells in all three tiers of cells present in longitudinal section 2 [42]. Seven intervening phloem pole pericycle cells can be found in tier one, and eight intervening cells can be identified in the third tier [42], resulting in 22 procambial cells in longitudinal section 2.

In a seven day old root, each of the longitudinal sections 3–13 contains approximately five tiers of cells. In longitudinal section 3, columella cells can no longer be identified, and 10 tiers of lateral root cap cells exist containing 28 cells each. In sections 4–6, a lateral root cap cell is twice the length and half the width of an epidermal cell. Eighty-four cells were identified in each tier, and two and a half tiers of cells exist each for longitudinal sections 4–6 resulting in 210 cells for each longitudinal section. All other cell types have undergone the appropriate tangential and periclinal divisions to establish their respective

cell files by longitudinal section 3. Two protophloem cells, two metaphloem cells and four accompanying companion cells are present in the phloem tissue [6]. With the combination of protophloem and metaphloem cells, 20 phloem cells and 20 companion cells exist in each longitudinal section. Approximately 40 procambial cells exist in each longitudinal section. Secondary cell growth does not occur in the developmental stages sampled, therefore, this number remains fixed throughout all developmental stages. In longitudinal section 12, a non-emerged lateral root is hypothesized to be present based on microarray expression data [7]. This lateral root is estimated to be approximately 130 cells, or one tier of cells in longitudinal section 2.

In our modelling the distinct vasculature, protophloem and metaphloem cell types were treated as a single cell type, as no marker-line was specific enough to differentiate clearly between these cell types. Also, the metaxylem and protoxylem were considered as a single cell type by the same rationale.

Solving bilinear equations

In this section, we show how the bilinear equations given by (3.3) can be solved using the algorithm in Section 2.1. The equations (3.3) have the form of equation (2.1) with $n = 13 + 14 = 27$ and $\ell = 30 + 1 = 31$. The extra equation beyond the 30 in (2.1) comes because we add a normalization condition, that $\sum_{i=1}^{13} x_i = 1$. We can take the matrix g to have $s = 2$ rows and have columns $\begin{pmatrix}1 & 0\end{pmatrix}^T$ for the x variables and $\begin{pmatrix}0 & 1\end{pmatrix}^T$ for the y variables. Then, the system satisfies the condition the beginning of Section 2.1 with $d_1 = d_2 = 1$.

For completeness, we write out the iteration steps (2.3) and (2.4) in this situation. At each iteration, the expectation step (2.3) computes the quantities:

$$w_{ijk}^{(s)} := b_k \frac{a_{ijk} x_i^{(s)} y_j^{(s)}}{\sum_{i'=1}^{n} \sum_{j'=1}^{m} a_{i'j'k} x_{i'}^{(s)} y_{j'}^{(s)}} \tag{3.4}$$

for all i, j, and k. This quantity $w_{ijk}^{(s)}$ is an estimate of the contribution of the (i, j) term in the kth equation in (3.3). We break the maximization step into first computing the analogues of the sufficient statistics:

$$X_i^{(s)} = \sum_{j=1}^{m} \sum_{k=1}^{\ell} w_{ijk}^{(s)}$$

$$Y_j^{(s)} = \sum_{i=1}^{n} \sum_{k=1}^{\ell} w_{ijk}^{(s)}.$$

Then we perform an iteration beginning with $x_i^{(s,0)} = x_i^{(s)}$ and $y_j^{(s,0)} = y_j^{(s)}$ and the update

rules

$$x_i^{(s,t+1)} := x_i^{(s,t)} \frac{X_i^{(s)}}{\sum_{j=1}^{m} \sum_{k=1}^{\ell} a_{ijk} x_i^{(s,t)} y_j^{(s,t)}}$$
$$y_j^{(s,t+1)} := y_j^{(s,t)} \frac{Y_j^{(s)}}{\sum_{i=1}^{n} \sum_{k=1}^{\ell} a_{ijk} x_i^{(s,t+1)} y_j^{(s,t)}}$$

until the parameters converge. Note that we are ignoring the normalization equation $\sum x_i = 1$. Instead, we incorporate that condition right before the next EM step by re-normalizing:

$$x_i^{(s+1)} := \frac{x_i^{(s,t)}}{\sum_{i'=1}^{n} x_{i'}^{(s,t)}}$$
$$y_j^{(s+1)} := y_j^{(s,t)} \sum_{i'=1}^{m} x_{i'}^{(s,t)}.$$

Since these are just local searches which may converge only to local minima, for each transcript, we ran our algorithm 20 different times starting from 20 different randomly chosen starting points. For every transcript in our data, all 20 runs of the algorithm converged to the same solution, up to a small tolerance. We therefore believe that in almost all cases we have found a global, and not merely local, minimum to the modified Kullback-Leibler divergence. Most likely, this consistency is a consequence of the particular coefficients of our equations, and in general there may be multiple local minima.

Computational validation methodology

In order to validate our method, we simulated expression profiles according to various models and tested our method's ability to reconstruct the underlying parameters. First, we simulated data according to the same independence model defined in the Model section. The underlying spatiotemporal subregion expression levels were sampled from a log-normal distribution with standard deviation 0.5. The simulated measurements b_k were computed from these subregion levels according to our model of the *Arabidopsis* root in (3.3). Finally, multiplicative error was added, distributed according to a log-normal distribution with standard deviation 0.03 to simulate measurement noise. This procedure created expression data with varying but comparable expression levels, which we will call the "uniform" dataset. However, since we are particularly interested in genes for which the expression levels are not uniform, we also produced simulations with the expression level for a given section or cell type raised by a factor of 10, which we will call the "elevated" dataset. In this dataset, we only measured the error for the same section or cell type which was elevated. These simulations measure our ability to detect a dominant expression pattern.

In addition, we designed simulations that test the robustness of the algorithm to failures of the bilinear model for root expression levels. For each section and cell type, we simulated data in which the expression levels for cells in that section or cell type did not follow the bilinear model, and call these the "section" and "cell type" datasets respectively. Instead, the expression levels in the given section or cell type were chosen independently according to a log-normal distribution with standard deviation $0.5\sqrt{2}$. The factor of $\sqrt{2}$ was introduced because the product of two log-normally distributed numbers with standard deviation 0.5 is distributed log-normally with standard deviation $0.5\sqrt{2}$.

The predictions were compared to the true expression levels across the spatiotemporal subregions within each section and each cell type. For each section and each cell type, the expression levels in its spatiotemporal subregions were averaged, ignoring those combinations which are not physically present in the root, (i.e. those whose entry in Table 3.2 is 0). The difference between the predicted and true average expressions was computed as a proportion of the true average expression. We then computed the root mean square of the proportional error over 500 simulations.

Visualization of predicted expression patterns

Predicted expression values were colored according to an *Arabidopsis* root template (Figure 3.1). The green channel of each cell was set according to a linear mapping between the expression range shown in the template $[1, 10]$ or $[1, 5]$ to the range $[0, 255]$. Expression values above or below that range are given values of 255 or 0 respectively. The mapping is also shown to the right of the false color image in the form of a gradient key. Phloem cells by longitudinal section are visualized separately on the right hand side of the root as they are physically occluded by other cells in the left hand side representation. The minimum and maximum range of expression value visualized can also be adjusted by the user.

In vivo validation methodology

To validate predicted expression values, we used transgenic *Arabidopsis thaliana* lines containing transcriptional GFP fusions in the Columbia ecotype [39]. For each gene being validated, six plants from at least two insertion lines previously described as expressing GFP were characterized. All plants were grown vertically on 1X Murashige and Skoog salt mixture, 1% sucrose and 2.3 mM 2-(N-morpholino)ethanesulfonic acid (pH 5.7) in 1% agar. Seedlings were prepared for microscopy at 5 days of age. Confocal images were obtained using a 25x water-immersion lens on a Zeiss LSM-510 confocal laser-scanning microscope using the 488-nm laser for excitation. Roots were stained with 10 μg/mL propidium iodide for 0.5 to 2 minutes and mounted in water. GFP was rendered in green and propidium iodide in red. Images were saved in TIFF format. Images were manually stitched together in Adobe Photoshop CS2 using the Photomerge command. The black background surrounding the

root was modified to ensure uniformity across figures. No other image enhancement was performed.

3.2 Results

Computational validation

The root mean square percentage errors in the reconstruction of each parameter are shown in Table 3.3. In the first two columns, where the data were generated according to the bilinear model, the error rate is generally no greater than the simulated measurement error. In most cases, elevated expression led to a lower error rate. In particular, reconstruction of expression in procambium was much more accurate in the elevated dataset.

The last two columns show that the algorithm is robust to violations of the bilinear model. Also, the predicted expression level in each cell type is generally not greatly affected by the failure of the model in other cell types, and similarly with sections.

In vivo validation

To determine whether our algorithm can accurately resolve spatiotemporal subregion-level transcript expression values, it would be ideal to compare the predictions to measured microarray expression values of the same spatiotemporal subregion. However, due to technical constraints, it is not possible to measure mRNA expression to such a degree of specificity and thus we cannot validate the estimates directly. Instead, we validated the method by visually comparing the predicted pattern of expression to patterns obtained from transcriptional GFP fusions using laser scanning confocal microscopy, as described in [39].

For each gene validated, a false-colored root image was generated by coloring each spatiotemporal subregion of an annotated *Arabidopsis* root template (Figure 3.1) according to the expression level in that subregion as predicted by our method. This false-colored image was then visually compared against the actual pattern of fluorescence observed in plants expressing a transcriptional GFP fusion specific for the promoter of that gene. These transcriptional GFP fusions contain up to 3 kb of regulatory sequence upstream of the translational start site of the respective gene. In many cases, this sequence is sufficient to recapitulate endogenous mRNA expression patterns as defined by cell type resolution microarray data [39]. This comparative method of validation allows us to assess the accuracy of spatiotemporal subregion expression predictions in an efficient and technically feasible way.

As a benchmark validation test, a set of three transcriptional fusions which were used to obtain some of the marker-line dataset were examined: S18(*AT5G12870*), S4(*AT3G25710*), and S32(*AT2G18380*). These fusions were originally selected for use in profiling because they exhibited enriched cell type expression as observed by laser scanning confocal microscopy and subsequently confirmed in the microarray expression data. The expression predictions

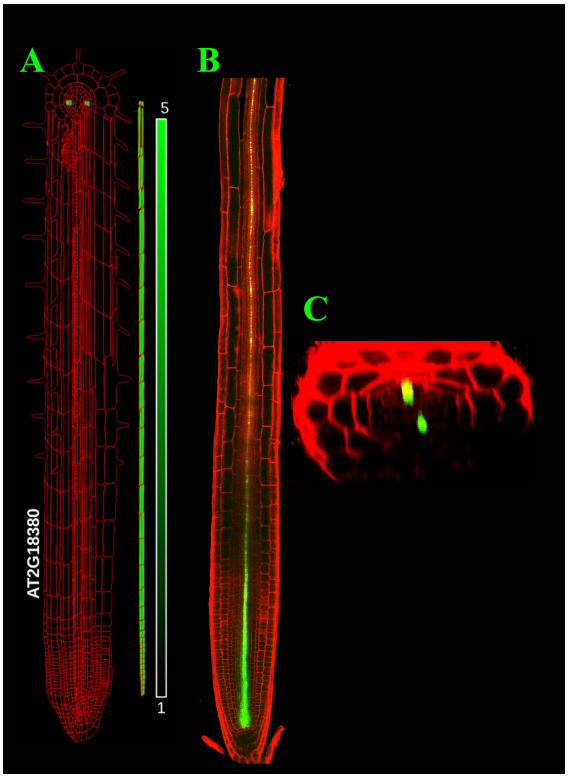

Figure 3.2: (A) Expression of *AT2G18380* in all developmental stages of the phloem was predicted by our method and visualized in a representation of the *Arabidopsis* root. Phloem cells are shown external to the root. (B) GFP expression in the longitudinal axis and (C) expression in cross-section of expression driven by the *AT2G18380* promoter validate the prediction.

	Error rate			
Variable	uniform	elevated	cell type	section
Section 1	2.7	2.4	3.3	3.6
Section 2	3.4	3.0	5.7	7.5
Section 3	3.3	2.7	5.8	7.2
Section 4	3.2	2.8	5.3	6.5
Section 5	3.1	2.7	5.3	6.5
Section 6	3.3	2.7	5.3	6.5
Section 7	3.1	2.5	3.7	5.0
Section 8	3.0	2.3	3.6	4.9
Section 9	3.0	2.2	3.6	4.8
Section 10	2.7	2.1	3.5	4.5
Section 11	2.9	2.2	3.4	4.6
Section 12	3.3	2.2	4.4	5.3
Section 13	2.4	2.1	3.6	5.3
Quiescent center	3.0	3.1	3.0	3.1
Columella	3.1	3.8	4.9	4.1
Lateral root cap	2.6	1.6	3.6	3.1
Hair cell	3.4	2.8	9.1	4.3
Non-hair cell	3.0	2.1	3.1	3.0
Cortex	2.9	2.1	6.9	3.6
Endodermis	2.8	2.2	3.5	3.2
Xylem pole pericycle	3.3	3.1	10.8	4.9
Phloem pole pericycle	3.0	2.9	9.4	4.9
Phloem	3.0	2.9	3.0	3.0
Phloem ccs	3.3	3.4	11.7	4.9
Xylem	2.2	2.1	2.5	2.2
Lateral root primordia	3.5	3.0	3.4	3.3
Procambium	8.3	1.8	12.7	12.7

Table 3.3: Root mean square percentage error rates in the reconstruction of simulated data. The first column is under a model of comparable but varying expression levels across all sections and cell types. The second type is the error rate when that section or cell type has its expression level raised by a factor of 10. The third and fourth columns show models in which the bilinear assumption is violated in one of the sections or one of the cell types respectively. In all cases, 3% measurement error has been added to the expression levels.

from our method accurately recapitulated the observed pattern of all three benchmark genes (Figure 3.2 and data not shown).

To assess the novel predictive ability of our method to reconstruct *in vivo* expression pat-

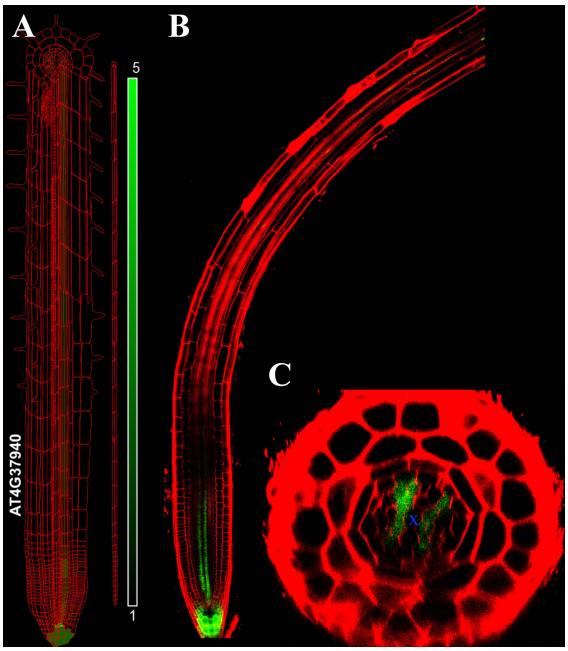

Figure 3.3: (A) Our method correctly predicts specific expression of *AT4G37940* in a cell type, procambium, that is only covered by a general tissue marker, WOL. Expression conferred by the *AT4G37940* promoter fused to GFP as a reporter was visualized in the columella (B) and in the procambium by a longitudinal section (B) and a cross section (C). The label X indicates the xylem axis. The expression also validates a maximal peak in the meristematic zone.

terns given missing data, we selected transcriptional fusions for genes for which our method predicts expression in cell types or in spatiotemporal subregions that were not marked by fluorescent marker-lines in the original dataset. At least two lines per transcriptional fusion were monitored. With respect to an unmarked cell type, our method predicted that *AT4G37940* was highly expressed in the columella and developing procambium. Imaging of a transcriptional fusion of this gene confirmed this expression (Figure 3.3).

To determine if our method could correctly differentiate expression in a specific developmental stage of a cell type, we selected *AT5G43040* for further analysis. The collection of marker-lines used to generate the original dataset included a marker for all developmental stages of non-hair cells, composed of their precursors (atrichoblasts) and fully developed non-hair cells. However, the marker-line used for hair cells only marks mature hair cells, and not their precursors (trichoblasts). Our method predicts *AT5G43040* expression throughout the epidermis—in mature hair cell, trichoblast, mature non-hair cell and atrichoblast cell files—with higher expression predicted in non-hair cells than in hair cells. This differential expression was validated using the *AT5G43040* transcriptional fusion (Supplementary Figure 2) demonstrating that our method is not only able to identify expression in a developmental stage of a cell type not marked by the marker-line data, but also to accurately differentiate relative levels of a transcript. However, it should be noted that expression in the transcriptional fusion did not fully corroborate the expression predicted by our algorithm—specifically, expression was found in the lateral root cap which was not predicted by our algorithm.

Examination of the raw microarray expression data revealed that expression was not elevated in the lateral root cap in the input microarray data. Most likely, the presence of GFP is not indicative of erroneous reconstruction of *AT5G43040* expression in this case. Instead, the transcriptional fusion does not contain sufficient regulatory elements to direct the appropriate expression as described in [39], perhaps within downstream sequences. For this reason, a comparison of the ratio between raw marker line and section expression data can be obtained as a link for each gene so that the user can simultaneously assess raw expression data with the reconstructed expression patterns.

3.3 Discussion

We have shown that spatiotemporal patterns of gene expression in the *Arabidopsis* root can be reconstructed using information from the marker-line and longitudinal datasets. Current experimental techniques are limited in their ability to rapidly and accurately microdissect organs into all component cell types at all developmental stages. Our computational technique helps to overcome these limitations. We fully integrate the marker-line and longitudinal data sets into a comprehensive expression pattern, across both space and time. In particular, this method has enabled the identification of *Arabidopsis* root procambium and trichoblast-specific genes, which have been previously experimentally intractable cell types.

Our high-resolution expression patterns will allow us to better understand the regulatory logic that controls developmental processes of the *Arabidopsis* root. These transcriptional regulatory networks are key to understanding developmental processes and environmental responses. With only a portion of these genes and fewer cell types, high-resolution spatiotemporal data has been used to identify transcriptional regulatory modules [7]. Our more accurate and complete dataset will allow a more comprehensive discovery of regulatory networks across additional cell types.

Moreover, we expect that our algorithm and the model which underlies it are applicable to time course experiments on other heterogeneous cell mixtures. Measurements in multicellular organisms are taken from complex cell mixtures of organs, tissues, heterogeneous cell lines, or cancerous samples. When precise histological characterization of these samples can estimate underlying cell type composition, our method can be used to reconstruct the underlying cell type-specific gene expression patterns or any other type of quantitative data, such as high-throughput protein abundance measurements. Theoretically, this algorithm can be applied to identify missing data in any experimental system that captures data in two or more dimensions which are assumed to be independent of one another.

Chapter 4

Semi-symmetric tensor ranks

This chapter is based on this paper "Secant varieties of $\mathbb{P}^2 \times \mathbb{P}^n$ embedded by $\mathcal{O}(1,2)$," which was jointly authored with Daniel Erman and Luke Oeding [11].

Let U, V, and W be complex vector spaces of dimension m, n, and k respectively, and let x be an element in the tensor product of of their duals, $U^* \otimes V^* \otimes W^*$. The *border rank* of x is the minimal r such that the corresponding point $[x] \in \mathbb{P}(U^* \otimes V^* \otimes W^*)$ lies in the rth secant variety of the Segre variety of $\mathbb{P}(U^*) \times \mathbb{P}(V^*) \times \mathbb{P}(W^*)$. Similarly, for a symmetric tensor $x \in S^3U^*$ or a partially symmetric tensor $x \in U^* \otimes S^2V^*$, the *symmetric border rank* and the *partially symmetric border rank* are the smallest r such that $[x]$ is in the rth secant variety of the Veronese or the Segre-Veronese variety, respectively. Developing effective techniques for computing and characterizing the border rank of tensors is an active area of research which spans classical algebraic geometry and representation theory [30, 31, 32, 33, 34, 43].

In the partially symmetric case, the secant varieties of $\mathbb{P}^1 \times \mathbb{P}^{n-1}$ embedded by the line bundle $\mathcal{O}(1,2)$ are closely related to standard results about pencils of symmetric matrices. Moreover, the non-symmetric analogue is $\mathbb{P}^1 \times \mathbb{P}^{n-1} \times \mathbb{P}^{k-1}$ embedded by $\mathcal{O}(1,1,1)$, and the defining equations of all of its secant varieties are known by work of Landsberg and Weyman [34, Thm. 1.1]. We record the partially symmetric analogue in Proposition 4.4.2.

The main result of this chapter is Theorem 4.4.3, which focuses on the next case: secant varieties of $\mathbb{P}^2 \times \mathbb{P}^{n-1}$ embedded by $\mathcal{O}(1,2)$. We give two explicit matrices, and we prove that, when $r \leq 5$, their minors and Pfaffians, respectively, generate the defining ideal for these secant varieties. To illustrate, fix a basis $\{e_1, e_2, e_3\}$ of U^*. We may then express any point $x \in \mathbb{P}(U^* \otimes S^2V^*)$ as $x = e_1 \otimes A_1 + e_2 \otimes A_2 + e_3 \otimes A_3$ where each $A_i \in S^2V^*$ can be represented by an $n \times n$ symmetric matrix. With the ordered triplet of matrices (A_1, A_2, A_3) serving as coordinates on $\mathbb{P}(U^* \otimes S^2V^*)$, our main result is the following, which is a restatement of Theorem 4.4.3:

Theorem 4.0.1. *Let Y be the image of $\mathbb{P}^2 \times \mathbb{P}^{n-1}$ in $\mathbb{P}^{3\binom{n+1}{2}-1} \cong \mathbb{P}(U^* \otimes S^2V^*)$ embedded by $\mathcal{O}(1,2)$. For any $r \leq 5$, the rth secant variety of Y is defined by the prime ideal generated*

by the $(r+1)\times(r+1)$ *minors of the* $n \times 3n$ *block matrix*

$$\begin{pmatrix} A_1 & A_2 & A_3 \end{pmatrix} \tag{4.1}$$

and by the $(2r+2)\times(2r+2)$ *principal Pfaffians of the* $3n \times 3n$ *block matrix*

$$\begin{pmatrix} 0 & A_3 & -A_2 \\ -A_3 & 0 & A_1 \\ A_2 & -A_1 & 0 \end{pmatrix}. \tag{4.2}$$

The matrices which appear in the statement of the above theorem are examples of what we call the "exterior flattenings" of a 3-tensor (see §4.1), and the construction of these matrices is motivated by the κ-invariant of a 3-tensor, as introduced in [22, §1.1]. The minors of the exterior flattenings of a 3-tensor impose necessary equations on a wide array of secant varieties of Segre-Veronese embeddings of products of projective spaces. The minors of these exterior flattenings simultaneously generalize both the minors obtained from flattenings of a 3-tensor and the determinantal equations of [47, Lem. 4.4] and [43, Thm. 3.2].

Under the hypotheses of Theorem 4.0.1, the minors and Pfaffians of these exterior flattenings are insufficient to generate the ideal of the rth secant variety for $r \geq 7$. In other words, Theorem 4.0.1 is false if $r \geq 7$, and we do not know if Theorem 4.0.1 holds when $r = 6$. See Example 4.4.12 for more details. Note that by [1, Cor. 1.4(ii)], these secant varieties have the expected dimension except when n is odd and $r = n + (n+1)/2$.

The proof of our main result uses a mix of representation theory and geometric techniques for studying determinantal varieties. We first introduce the relevant determinantal ideals and we use their equivariance properties to relate these ideals as the size of the tensor varies. Next, we apply this relation in order to understand the defining ideals of certain auxiliary varieties known as the *subspace varieties* $\mathrm{Sub}_{m',n'}$ (see Definition 4.3.1). We then prove our main result in the special case that $n = r$, by relating the secant variety with the variety of commuting symmetric $n \times n$ matrices. A similar idea has appeared in several instances previously [3, 43, 47]. This step requires $r \leq 5$. Finally, we prove our main result by blending our results about subspace varieties with our knowledge about the case $n = r$.

Partially symmetric 3-tensors are closely related to the study of vector spaces of quadrics, which arise naturally in algebraic geometry. For instance, in the study of Hilbert schemes of points, border rank is connected to the smoothability of zero dimensional schemes [13, 22]. As another example, [43, Prop. 6.3] relates the border rank of a partially symmetric tensor $x \in \mathbb{C}^3 \otimes S^2(\mathbb{C}^n)$ with properties of the corresponding degree n determinantal curve in $\mathbb{P}^2$.

Questions about the border rank of partially symmetric tensors also arise in algebraic statistics [24, §7]. For instance, the situation of Theorem 4.4.3 corresponds to a mixture of random processes, each independently sampling from a distribution with 3 states and sampling twice from a distribution with n states. The border rank of the observed distribution corresponds to the number of processes in the mixture. This connection will be taken up in Chapter 5

In signal processing, a partially symmetric tensor in $U^* \otimes S^2V^*$ can be constructed as the second derivative of the cumulant generating function taken at m points [46]. The matrix equations in Theorem 4.4.3 can be used to study small border ranks of such tensors when $m = 3$.

The defining ideal of the rth secant variety of $\mathbb{P}^2 \times \mathbb{P}^{n-1}$ was previously known in the case when this secant variety is a hypersurface. This occurs when $n \geq 4$ is even, and $r = \frac{3n-2}{2}$, and this result follows from an analogue to Strassen's argument [47, § 4], as shown by Ottaviani in the remark following Theorem 4.1 in [43]. For historical interest, we note that the hypersurface case $n = 4$ and $r = 5$ dates to Emil Toeplitz [48].

Theorem 4.4.3 thus provides a new family of examples where we can effectively compute the border rank of a partially symmetric tensor. Our main result also provides evidence for a partially symmetric analogue of Comon's Conjecture, which posits that the symmetric rank of a tensor equals the rank [15, §5], as discussed in Remark 4.4.5 below.

This chapter is organized as follows. In §4.1, we define a vector κ as an invariant of a 3-tensor. We use this κ-invariant to produce explicit matrix equations which vanish on the secant varieties of Segre-Veronese embeddings of projective spaces. To provide a more invariant perspective, and to connect with previous literature [30, 31, 34, 35], we also provide Schur module decompositions for our matrix equations. In §4.2, we restrict to the κ-invariant of a partially symmetric tensor. Here we also provide Schur module decompositions in the partially symmetric case. In §4.3, we show that the κ_0 equations define subspace varieties. We prove our main result, Theorem 4.4.3, in §4.4.

Remark 4.0.2. The results giving equations vanishing on the Segre and Segre-Veronese varieties (Propositions 4.1.5 and 4.2.3) hold in arbitrary characteristic. However, our proof of Theorem 4.4.3 does not extend to arbitrary characteristic because it relies on Lemmas 4.3.4 and 4.4.9 and [8, Thm. 3.1], all of which require characteristic 0.

4.1 The κ-invariant of a 3-tensor

From a tensor in $U^* \otimes V^* \otimes W^*$, we will construct a series of linear maps, whose ranks we define to be the *κ-invariants* of the tensor. The κ-invariants give inequalities on the rank of the tensor, and thus, determinantal equations which vanish on the secant variety.

There is a natural map $U^* \otimes \bigwedge^j U^* \to \bigwedge^{j+1} U^*$ defined by sending $u \otimes u' \mapsto u \wedge u'$ for any $0 \leq j \leq m-1$. This induces an inclusion $U^* \subseteq \bigwedge^j U \otimes \bigwedge^{j+1} U^*$. By tensoring on both sides by $V^* \otimes W^*$, we get an inclusion $U^* \otimes V^* \otimes W^* \subseteq (V \otimes \bigwedge^j U^*)^* \otimes (W^* \otimes \bigwedge^{j+1} U^*)$. An element of the tensor product on the right-hand side may be interpreted as a linear homomorphism, meaning that for any $x \in U^* \otimes V^* \otimes W^*$ we have a homomorphism

$$\psi_{j,x} \colon V \otimes \bigwedge^j U^* \to W^* \otimes \bigwedge^{j+1} U^*,$$

and $\psi_{j,x}$ depends linearly on x. We call $\psi_{j,x}$ an *exterior flattening* of x, as it generalizes the flattening of a tensor, as discussed below.

Definition 4.1.1. *Following [22, Defn. 1.1], we define $\kappa_j(x)$ to be the rank of $\psi_{j,x}$, and we let $\kappa(x)$ denote the vector of κ-invariants $(\kappa_0(x), \ldots, \kappa_{m-1}(x))$.*

More concretely, by choosing bases for the vector spaces, we can represent $\psi_{j,x}$ as a matrix. If $e_1, \ldots, e_m$ is a basis for U^*, then a basis for $\bigwedge^j U^*$ is given by the set of all $e_{i_1} \wedge \ldots \wedge e_{i_j}$ for $1 \leq i_1 < \cdots < i_j \leq m$, and analogously for $\bigwedge^{j+1} U^*$. For a fixed $u = \sum_{k=1}^m u_k e_k$ in U^*, the map $\bigwedge^j U^* \to \bigwedge^{j+1} U^*$ defined by $u' \mapsto u \wedge u'$ will send $e_{i_1} \wedge \ldots \wedge e_{i_j}$ to $\sum_k u_k\, e_k \wedge e_{i_1} \wedge \ldots \wedge e_{i_j}$. Thus, this map will be represented in the above bases by a matrix whose entries are either 0 or $\pm u_i$. The matrix for $\psi_{j,x}$ is the block matrix formed by replacing the scalar u_i with the matrix A_i where $A_i \in V^* \otimes W^*$ are the matrices such that $x = \sum_{i=1}^m e_i \otimes A_i$.

For example, if $m = 4$, then $\psi_{j,x}$ are represented by the following matrices (in the coordinates described above):

$$\psi_{0,x} \colon V \otimes \textstyle\bigwedge^0 U^* \xrightarrow{\begin{pmatrix} A_1 \\ A_2 \\ A_3 \\ A_4 \end{pmatrix}} W^* \otimes \bigwedge^1 U^*,$$

$$\psi_{1,x} \colon V \otimes \textstyle\bigwedge^1 U^* \xrightarrow{\begin{pmatrix} 0 & A_3 & -A_2 & 0 \\ -A_3 & 0 & A_1 & 0 \\ A_2 & -A_1 & 0 & 0 \\ A_4 & 0 & 0 & -A_1 \\ 0 & A_4 & 0 & -A_2 \\ 0 & 0 & A_4 & -A_3 \end{pmatrix}} W^* \otimes \bigwedge^2 U^*,$$

$$\psi_{2,x} \colon V \otimes \textstyle\bigwedge^2 U^* \xrightarrow{\begin{pmatrix} -A_4 & 0 & 0 & 0 & A_3 & -A_2 \\ 0 & -A_4 & 0 & -A_3 & 0 & A_1 \\ 0 & 0 & -A_4 & A_2 & -A_1 & 0 \\ A_1 & A_2 & A_3 & 0 & 0 & 0 \end{pmatrix}} W^* \otimes \bigwedge^3 U^*$$

$$\psi_{3,x} \colon V \otimes \textstyle\bigwedge^3 U^* \xrightarrow{(\,A_1\ A_2\ A_3\ A_4\,)} W^* \otimes \bigwedge^4 U^*.$$

The entries of these matrices are linear forms on $\mathbb{P}(U^* \otimes V^* \otimes W^*)$, and the minors of these matrices $\psi_{j,x}$ are the "explicit matrix equations" alluded to in the introduction.

Note that for $j = 0$, the map $\psi_0 \colon V \otimes \bigwedge^0 U^* \cong V \to W^* \otimes U^*$ is the homomorphism corresponding to x in the identification $U^* \otimes V^* \otimes W^* \cong \operatorname{Hom}(V, U^* \otimes W^*)$. In the literature, the matrix for ψ_0 is referred to as a "flattening" of x by grouping W^* and U^*. Similarly, $\psi_{m-1,x} \colon V \otimes \bigwedge^{m-1} U^* \to W^*$ is the flattening formed by grouping U^* and V^*, because $\bigwedge^{m-1} U^* \cong U$.

If $[x] \in \mathbb{P}(U^* \otimes V^* \otimes W^*)$ is a tensor, then the *rank* of $[x]$ is the number r in a minimal expression $x = u_1 \otimes v_1 \otimes w_1 + \cdots + u_r \otimes v_r \otimes w_r$, where $u_i \in U$, $v_i \in V$, and $w_i \in W$ for $1 \leq i \leq r$. The set of rank-one tensors is closed and equals the Segre variety $\operatorname{Seg}(\mathbb{P}U^* \times \mathbb{P}V^* \times \mathbb{P}W^*)$. More generally, the Zariski closure of the set of tensors $[x] \in \mathbb{P}(U^* \otimes V^* \otimes W^*)$ having rank at most r is the rth secant variety of the Segre product, denoted $\sigma_r(\operatorname{Seg}(\mathbb{P}U^* \times \mathbb{P}V^* \times \mathbb{P}W^*))$.

Definition 4.1.2. *The* border rank *of $[x] \in \mathbb{P}(U^* \otimes V^* \otimes W^*)$ is the minimal r such that $[x]$ is in $\sigma_r(\operatorname{Seg}(\mathbb{P}U^* \times \mathbb{P}V^* \times \mathbb{P}W^*))$.*

The following lemma generalizes the well-known fact that the ranks of the flattenings are bounded above by the tensor rank, and extends a result of Ottaviani [43, Thm. 3.2(i)].

Lemma 4.1.3. *If $x \in U^* \otimes V^* \otimes W^*$ has border rank at most r, then $\kappa_j(x) \leq r\binom{m-1}{j}$.*

Proof. Since κ_j is defined in terms of a matrix rank, an upper bound on κ_j is a closed condition on the set of tensors. It thus suffices to prove the statement with border rank replaced by rank. As observed above, $\psi_{j,x}$ depends linearly on x, so it is sufficient to assume that x is an indecomposable tensor, and then show that $\kappa_j(x) \leq \binom{m-1}{j}$. We can choose coordinates such that $x = e_1 \otimes A$, and A is a matrix with only one non-zero entry. The non-zero rows of the matrix for $\psi_{j,x}$ will correspond to those basis elements $e_{i_1} \wedge \cdots \wedge e_{i_j} \in \bigwedge^j U^*$ such that $2 \leq i_1 < \cdots < i_j \leq n$, each of which is sent to a multiple of $e_1 \wedge e_{i_1} \wedge \cdots \wedge e_{i_j}$. Since there are $\binom{m-1}{j}$ such basis elements, the rank of $\psi_{j,x}$ is equal to $\binom{m-1}{j}$. □

The above lemma illustrates that the minors of the exterior flattenings provide equations which vanish on the secant variety of a Segre triple product. We write $S^\bullet(U \otimes V \otimes W)$ to denote the polynomial ring on the affine space $U^* \otimes V^* \otimes W^*$.

Definition 4.1.4. *Let $c = (c_0, \ldots, c_{m-1})$ be a vector of positive integers. We define $I_{\kappa_i \leq c_i}$ to be the ideal generated by the $(c_i+1) \times (c_i+1)$-minors of $\psi_{j,x}$. Similarly, we use the notation $I_{\kappa \leq c}$ for the ideal generated by $I_{\kappa_i \leq c_i}$ for all $0 \leq i \leq m-1$. Finally, we define $\Sigma_{\kappa_i \leq c_i}$ and $\Sigma_{\kappa \leq c}$ to be the subschemes of $\mathbb{P}(U^* \otimes V^* \otimes W^*)$ given by the ideals $I_{\kappa_i \leq c_i}$ and $I_{\kappa \leq c}$ respectively.*

Proposition 4.1.5. *Fix $r \geq 1$. If c is the vector defined by $c_j = r\binom{m-1}{j}$ for $0 \leq j \leq m-1$, then*

$$\sigma_r(\mathrm{Seg}(\mathbb{P}U^* \times \mathbb{P}V^* \times \mathbb{P}W^*)) \subseteq \Sigma_{\kappa \leq c}.$$

Remark 4.1.6. Fundamental in the construction of the exterior flattening $\psi_{j,x}$ was the inclusion of U^* into $\bigwedge^j U \otimes \bigwedge^{j+1} U^*$. More generally, any natural inclusion of U^* into the tensor product of two representations would yield an analogue of $\psi_{j,x}$ as well as analogues of Lemma 4.1.3 and Proposition 4.1.5. For instance, from the inclusion $U^* \subseteq S_{(2,1)}(U) \otimes S_{(2,1,1)}(U^*)$, we may associate to a tensor x a homomorphism:

$$V \otimes S_{(2,1)}(U^*) \to W^* \otimes S_{(2,1,1)}(U^*),$$

whose rank is at most 5 times the border rank of x. We restrict our attention to the κ-invariants because these seem to provide particularly useful inequalities in our cases of interest. However, an example of this generalized construction was introduced and applied in [44, Thm. 1.1] and has been further developed under the name of Young flattening in [32].

Example 4.1.7. If $m = 2$, then as stated above, κ_0 and κ_1 are the ranks of the flattenings formed by grouping W^* with U^* and V^* with U^*, respectively. The ideal of the rth secant variety is $I_{\kappa \leq (r,r)}$ [34, Theorem 1.1]. □

Example 4.1.8. Let $m = 3$ and suppose that $n = k$ is odd. Denote by X the Segre product $\mathrm{Seg}(\mathbb{P}U^* \times \mathbb{P}V^* \times \mathbb{P}W^*)$ in $\mathbb{P}(U^* \otimes V^* \otimes W^*)$. Then $I_{\kappa_1 \leq 3n-1}$ is a principal ideal generated by the determinant of $\psi_{1,x}$, which defines the secant variety $\sigma_{\frac{3n-1}{2}}(X)$ [47, Lem. 4.4] (see also [43, Rmk. 3.3]). □

Example 4.1.9. The exterior flattening $\psi_{1,x}$ has also arisen in the study of totally symmetric tensors. For instance, when $n = 3$, the secant variety $\sigma_3(\nu_3(\mathbb{P}^2)) \subseteq \mathbb{P}^9$ is a hypersurface defined by the Aronhold invariant. Ottaviani has shown that this hypersurface is defined by any of the 8×8 Pfaffians of the matrix representing $\psi_{1,x}$ specialized to symmetric tensors [44, Thm. 1.2]. □

However, the ideals $I_{\kappa \leq c}$ do *not* equal the defining ideals of secant varieties even in relatively simple cases.

Example 4.1.10. Let $n = m = k = 3$ and let Y be the image of $\mathbb{P}^2 \times \mathbb{P}^2 \times \mathbb{P}^2 \subseteq \mathbb{P}^{26}$ embedded by $\mathcal{O}(1, 1, 1)$. By Proposition 4.1.5, we know that $I_{\kappa \leq (3,6,3)}$ vanishes on $\sigma_3(Y)$, but we claim that it is not the defining ideal. Observe that the conditions $\kappa_0 \leq 3$ and $\kappa_2 \leq 3$ are trivial, and hence $I_{\kappa \leq (3,6,3)} = I_{\kappa_1 \leq 6}$. By definition, the ideal $I_{\kappa_1 \leq 6}$ is generated by the 7×7 minors of $\psi_{1,x}$. However, [34, Thm. 1.3] produces degree 4 equations which vanish on $\sigma_3(Y)$, and since $I_{\kappa_1 \leq 6}$ is generated in degree 7, we see that it does not equal the defining ideal of $\sigma_3(Y)$. □

We now study our matrix equations from the perspective of representation theory, which connects them to previous work on secant varieties of Segre-Veronese varieties. The representation theory of our ideals $I_{\kappa_i \leq c_i}$ will also be necessary in the proof of Lemma 4.3.4.

Since the ideals $I_{\kappa_i \leq c_i}$ are invariant under the natural action of $\mathrm{GL}(U) \times \mathrm{GL}(V) \times \mathrm{GL}(W)$, their generators can be decomposed as direct sums of irreducible representations of that group. Each polynomial representation of $\mathrm{GL}(U) \times \mathrm{GL}(V) \times \mathrm{GL}(W)$ is of the form $S_\mu U \otimes S_\nu V \otimes S_\omega W$ where $S_\mu U$, $S_\nu V$, and $S_\omega W$ are the Schur modules indexed by partitions μ, ν, and ω with at most m, n, and k parts, respectively, where if $\pi = (\pi_1, \ldots, \pi_s)$ is a partition with $\pi_1 \geq \pi_2 \geq \cdots \geq \pi_s > 0$, then we say that s is the number of parts of π. For the summands of the degree d part of $I_{\kappa_i \leq c_i} \subset S^d(U \otimes V \otimes W)$, the partitions will always be partitions of d. For general background on Schur modules see [23].

Lemma 4.1.11. *For each j and c_j, there is a $\mathrm{GL}(U) \times \mathrm{GL}(V) \times \mathrm{GL}(W)$-equivariant map*

$$\Phi_j \colon \bigwedge^{c_j+1} \left(V \otimes \bigwedge^j U^* \right) \otimes \bigwedge^{c_j+1} \left(W \otimes \bigwedge^{j+1} U \right) \to S^{c_j+1} (U \otimes V \otimes W),$$

whose image equals the vector space of generators of $I_{\kappa_j \leq c_j}$. In particular, every irreducible representation arising in the Schur module decomposition of the generators of $I_{\kappa_j \leq c_j}$ must be a submodule of both the source and target of Φ_j.

Proof. Consider the map

$$\psi_{j,x} \colon V \otimes \bigwedge^j U^* \to W^* \otimes \bigwedge^{j+1} U^*.$$

After choosing bases of U, V, and W, we may think of $\psi_{j,x}$ as a matrix of linear forms in $S^\bullet(U \otimes V \otimes W)$. Taking the $(c_j+1) \times (c_j+1)$-minors of $\psi_{j,x}$ then determines the map Φ_j. More concretely, our choice of bases for U, V, and W determines a natural basis for the source of Φ_j consisting of indecomposable tensors; we define the map Φ_j by sending a basis element to the corresponding minor of the matrix $\psi_{j,x}$. Since the ideal $I_{\kappa_j \leq c_j}$ is defined as the ideal generated by the image of Φ_j, the lemma follows from Schur's Lemma. □

When $j = 0$, it is straightforward to compute the Schur module decomposition of $I_{\kappa_0 \leq c_0}$, as illustrated by the following example.

Example 4.1.12. The map

$$\psi_{0,x} \colon V \otimes \textstyle\bigwedge^0 U^* \to W^* \otimes \bigwedge^1 U^*$$

is a flattening of the tensor x by grouping U^* and W^*. As representations, the minors of $\psi_{0,x}$ decompose into irreducibles using the skew Cauchy formula [23, p. 80]

$$\textstyle\bigwedge^{c_0+1} V \otimes \bigwedge^{c_0+1}(W \otimes U) = \bigwedge^{c_0+1} V \otimes \left(\bigoplus_{|\lambda|=c_0+1} S_\lambda W \otimes S_{\lambda'} U \right),$$

where λ ranges over all partitions of $c_0 + 1$ and λ' is the conjugate partition.

For instance, let $n = m = k = 3$ and consider the generators of $I_{\kappa_0 \leq 2}$. This is a vector space of cubic polynomials, and by Lemma 4.1.11, it must be the module

$$\textstyle\bigwedge^3 V \otimes \big((S_3 W \otimes S_{1,1,1} U) \oplus (S_{2,1} W \otimes S_{2,1} U) \oplus (S_{1,1,1} \otimes S_3 U) \big).$$

After distributing, each irreducible module is the tensor product of three Schur functors applied to U, V, and W respectively, and we can thus drop the vector spaces and the tensor products from our notation, replacing $S_\mu U \otimes S_\nu V \otimes S_\omega W$ with $S_\mu S_\nu S_\omega$ without any ambiguity. Thus, we rewrite this module as:

$$S_{1,1,1} S_{1,1,1} S_3 \oplus S_{2,1} S_{1,1,1} S_{2,1} \oplus S_3 S_{1,1,1} S_{1,1,1}.$$

The dimension of this space is $10+64+10 = 84$, which equals the number of maximal minors of the 3×9 matrix $\psi_{0,x}$. □

When $j > 0$, the existence of the dual vector space U^* in the source of $\psi_{j,x}$ makes finding the Schur module decomposition of $I_{\kappa_i \leq c_i}$ more subtle. In Proposition 4.1.14 below, we provide an upper bound for the Schur module decomposition of $I_{\kappa_1 \leq c_1}$ in the case that $\dim U = 3$. To state the formula precisely, we first recall some notation.

For any vector space A, the Littlewood-Richardson formula is

$$S_\lambda A \otimes S_\mu A = \bigoplus_{|\pi|=|\lambda|+|\mu|} S_\pi A^{\oplus c^\pi_{\lambda,\mu}}, \tag{4.3}$$

where the multiplicities $c^{\pi}_{\lambda,\mu}$ are the *Littlewood-Richardson numbers*. For two vectors spaces A and B, we use the outer plethysm formula

$$S_\pi(A \otimes B) = \bigoplus_{|\lambda|+|\mu|=|\pi|} (S_\lambda A \otimes S_\mu B)^{\oplus K_{\pi,\lambda,\mu}} \tag{4.4}$$

to define the *Kronecker coefficients* $K_{\pi,\lambda,\mu}$.

Remark 4.1.13. In Propositions 4.1.14, 4.2.5 and 4.2.8, we will use the fact that as $\mathrm{GL}(U)$ modules, $S_\pi U^* \otimes (\bigwedge^m U)^l \cong S_{l^m - \pi} U$, where l^m denotes the partition $(l, \ldots, l)$. We caution that the entries in $l^m - \pi = (l - \pi_m, \ldots, l - \pi_1)$ are reversed.

Proposition 4.1.14. *Let* $\dim(U) = 3$. *For any Schur module* $S_\pi U \otimes S_\lambda V \otimes S_\mu W$, *let* λ' *and* μ' *denote the conjugate partitions of* λ *and* μ *respectively, and let* $(3)^{c_0+1} - \pi$ *be the difference as in Remark 4.1.13. If* $S_\pi U \otimes S_\lambda V \otimes S_\mu W$ *occurs in the decomposition of* $(I_{\kappa_1 \le c_1})_{c_1+1}$ *from Definition 4.1.4, then* π, λ', *and* μ' *have at most* 3 *parts, and the multiplicity of* $S_\pi U \otimes S_\lambda V \otimes S_\mu W$ *is at most the minimum of* $c^{(3)^{c_0+1}-\pi}_{\lambda',\mu'}$ *and* $K_{\pi,\lambda,\mu}$.

Computations with the software package LiE [49] suggest that the decomposition of $(I_{\kappa_1 \le c_1})_{c_1+1}$ may equal the upper bound of Proposition 4.1.14, as in Example 4.1.15.

Proof of Proposition 4.1.14. Using Lemma 4.1.11, the $(c_1+1)\times(c_1+1)$-minors of $\psi_{1,x}$ belong to the common submodules of the polynomials $S^{c_1+1}(U \otimes V \otimes W)$ and the domain of Φ_j, which we can rewrite using the Cauchy skew formula:

$$\left(\bigoplus_{|\lambda|=c_1+1} S_\lambda V \otimes S_{\lambda'} U^* \right) \otimes \left(\bigoplus_{|\mu|=c_1+1} S_\mu W \otimes S_{\mu'}\big(\textstyle\bigwedge^2 U\big) \right). \tag{4.5}$$

Here we note that λ' and μ' must have no more than 3 parts or else the summand is zero.

We focus on the U factor and compute

$$\begin{aligned}
S_{\lambda'}U^* \otimes S_{\mu'}(\textstyle\bigwedge^2 U) &\cong S_{\lambda'}U^* \otimes S_{\mu'}(U^*) \otimes (\textstyle\bigwedge^3 U)^{c_0+1} && \text{because } \dim U = 3 \\
&\cong \bigoplus_{|\nu|=2(c_0+1)} (S_\nu U^*)^{\oplus c^\nu_{\lambda',\mu'}} \otimes (\textstyle\bigwedge^3 U)^{c_0+1} && \text{by (4.3)} \\
&\cong \bigoplus_{|\nu|=2(c_0+1)} (S_{(c_0+1)^3-\nu} U)^{\oplus c^\nu_{\lambda',\mu'}} && \text{by Remark 4.1.13} \\
&\cong \bigoplus_{|\pi|=c_0+1} (S_\pi U)^{\oplus c^{(c_0+1)^3-\pi}_{\lambda',\mu'}} && \text{by taking } \pi = (c_0+1)^3 - \nu.
\end{aligned}$$

Therefore expression (4.5) becomes

$$\bigoplus_{|\lambda|=|\mu|=|\pi|=c_0+1} S_\pi U \otimes S_\lambda V \otimes S_\mu W^{\oplus c^{(c_0+1)^3-\pi}_{\lambda',\mu'}}.$$

$$
\begin{aligned}
(I_{\kappa_1\leq 1})_2 = \;& (\mathfrak{S}_3\cdot S_{1,1}S_{1,1}S_2)\oplus S_2S_2S_2\\
(I_{\kappa_1\leq 2})_3 = \;& (\mathfrak{S}_3\cdot S_{1,1,1}S_{2,1}S_{2,1})\oplus S_{2,1}S_{2,1}S_{2,1}\oplus(\mathfrak{S}_3\cdot S_{2,1}S_{2,1}S_3)\oplus S_3S_{1,1,1}S_{1,1,1}\\
(I_{\kappa_1\leq 3})_4 = \;& S_{2,2}S_{2,2}S_{2,2}\oplus(\mathfrak{S}_3\cdot S_{2,2}S_{2,1,1}S_{2,1,1})\oplus(\mathfrak{S}_3\cdot S_{2,2}S_{2,1,1}S_{3,1})\oplus(\mathfrak{S}_3\cdot S_{2,2}S_{3,1}S_{3,1})\\
&\oplus S_{2,1,1}S_{2,1,1}S_{2,1,1}\oplus(\mathfrak{S}_3\cdot S_{2,1,1}S_{3,1}S_{3,1})\oplus S_{3,1}S_{2,1,1}S_{2,1,1}\oplus S_4S_{2,2}S_{2,2}\\
&\oplus S_4S_{2,1,1}S_{2,1,1}\\
(I_{\kappa_1\leq 4})_5 = \;& S_{2,2,1}S_{2,2,1}S_{2,2,1}\oplus(\mathfrak{S}_3\cdot S_{2,2,1}S_{2,2,1}S_{3,1,1})\oplus(\mathfrak{S}_3\cdot S_{2,2,1}S_{3,2}S_{3,2})\oplus(\mathfrak{S}_3\cdot S_{2,2,1}S_{3,2}S_{3,1,1})\\
&\oplus(\mathfrak{S}_3\cdot S_{3,2}S_{3,2}S_{3,1,1})\oplus S_{3,2}S_{2,2,1}S_{2,2,1}\oplus S_{3,2}S_{3,1,1}S_{3,1,1}\oplus S_{3,1,1}(\mathfrak{S}_2\cdot S_{2,2,1}S_{3,1,1})\\
&\oplus S_{3,1,1}S_{3,1,1}S_{3,1,1}\oplus S_{4,1}S_{2,2,1}S_{2,2,1}\oplus S_{4,1}(\mathfrak{S}_2\cdot S_{2,2,1}S_{3,2})\oplus S_{4,1}(\mathfrak{S}_2\cdot S_{2,2,1}S_{3,1,1})\\
&\oplus S_5S_{2,2,1}S_{2,2,1}\\
(I_{\kappa_1\leq 5})_6 = \;& S_{2,2,2}S_{3,3}S_{3,3}\oplus(\mathfrak{S}_3\cdot S_{2,2,2}S_{3,2,1}S_{3,2,1})\oplus(\mathfrak{S}_3\cdot S_{3,3}S_{3,2,1}S_{3,2,1})\oplus(S_{3,2,1}S_{3,2,1}S_{3,2,1})^{\oplus 2}\\
&\oplus S_{4,2}S_{2,2,2}S_{2,2,2}\oplus S_{4,2}(\mathfrak{S}_2\cdot S_{2,2,2}S_{3,2,1})\oplus S_{4,2}S_{3,2,1}S_{3,2,1}\oplus S_{4,1,1}(\mathfrak{S}_2\cdot S_{2,2,2}S_{3,3})\\
&\oplus S_{4,1,1}(\mathfrak{S}_2\cdot S_{2,2,2}S_{3,2,1})\oplus S_{4,1,1}S_{3,2,1}S_{3,2,1}\oplus S_{5,1}(\mathfrak{S}_2\cdot S_{2,2,2}S_{3,2,1})\oplus S_6S_{2,2,2}S_{2,2,2}\\
(I_{\kappa_1\leq 6})_7 = \;& S_{3,3,1}S_{3,3,1}S_{3,3,1}\oplus(\mathfrak{S}_3\cdot S_{3,3,1}S_{3,2,2}S_{3,2,2})\oplus S_{3,2,2}S_{3,3,1}S_{3,3,1}\oplus S_{4,2,1}(\mathfrak{S}_2\cdot S_{3,3,1}S_{3,2,2})\\
&\oplus S_{4,2,1}S_{3,2,2}S_{3,2,2}\oplus S_{5,1,1}S_{3,2,2}S_{3,2,2}\\
(I_{\kappa_1\leq 7})_8 = \;& S_{3,3,2}S_{3,3,2}S_{3,3,2}\oplus S_{4,2,2}S_{3,3,2}S_{3,3,2}\\
(I_{\kappa_1\leq 8})_9 = \;& S_{3,3,3}S_{3,3,3}S_{3,3,3}
\end{aligned}
$$

Figure 4.1: The Schur module decompositions of $(I_{\kappa_1\leq c_1})_{c_1+1}$ in the $\mathcal{O}(1,1,1)$ case when $k=m=n=3$.

Now we must decide which irreducible modules occur as a submodule of $S^{c_1+1}(U\otimes V\otimes W)$. For this, we decompose the space of polynomials using the Cauchy formula and the outer plethysm formula (4.4):

$$S^{c_1+1}(U\otimes V\otimes W)\cong\bigoplus_{|\pi|=c_1+1}S_\pi U\otimes S_\pi(V\otimes W)\cong\bigoplus_{|\pi|=|\lambda|=|\mu|=c_0+1}(S_\pi U\otimes S_\lambda V\otimes S_\mu W)^{\oplus K_{\pi,\lambda,\mu}}.$$

The proposition statement follows by Lemma 4.1.11. □

Example 4.1.15. Let $n=m=k=3$. Using LiE [49], we computed every decomposition of $(I_{\kappa_1\leq c_1})_{c_1+1}$ using Proposition 4.1.14. These decompositions appear in Figure 4.1. To save space we omit the notation of vector spaces and tensor products, as in Example 4.1.12. Further, we use the notation $\mathfrak{S}_s$ to indicate the direct sum over the (non-redundant) permutations of the subsequent s Schur modules. The dimensions of the modules found in Figure 4.1 are Since these dimensions match the dimensions of the space of minors of $\psi_{1,x}$, as computed in Macaulay2 [26], all of the modules must be in the space of minors.

For $c_1=6$, we have $\dim\,(I_{\kappa_1\leq 6})_7=1296=\binom{9}{7}^2$, and hence we see that all 7×7-minors of $\psi_{1,x}$ are linearly independent. By contrast, if $c_1=5$, the fact that $\dim_{\mathbb{C}}(I_{\kappa_1\leq 5})_6=7011<$

$7056 = \binom{9}{6}^2$ tells us that the 6×6 minors of $\psi_{1,x}$ are not all linearly independent. For example, the upper right and lower left 6×6 minors of

$$\psi_{1,x} = \begin{pmatrix} 0 & A_3 & -A_2 \\ -A_3 & 0 & A_1 \\ A_2 & -A_1 & 0 \end{pmatrix}$$

are both equal to $\det(A_1) \cdot \det(A_3)$. □

In the next section we impose partial symmetry on our 3-tensors. We remark that we could impose other types of symmetry and this would lead to different investigations. For instance, we could restrict attention to 3-tensors in any of the following cases: $S^3(U^*)$, $U^* \otimes \bigwedge^2 V^*$, $\bigwedge^3 V^*$, or $S_{2,1}U^*$. In these cases, it would be straightforward to prove analogues of Lemma 4.1.3 and Proposition 4.1.5. However, if we hope to produce the ideal defining the appropriate secant varieties, then it is less obvious how to generalize Definition 4.1.4. It might be interesting to investigate the secant varieties of these other special types of 3-tensors.

4.2 The κ-invariant for partially symmetric 3-tensors

For the rest of the chapter, we take $W = V$ and focus on partially symmetric 3-tensors $x \in U^* \otimes S^2V^* \subset U^* \otimes V^* \otimes V^*$. By this latter inclusion, we may extend the definition of $\kappa_j(x)$ to partially symmetric tensors. Not only does the κ-invariant provide a bound for the rank of x, but also for the *partially symmetric rank*, which is defined as the minimal r such that $x = \sum_{i=1}^{r} u_i \otimes v_i \otimes v_i$, for some $u_i \in U^*$ and $v_i \in V^*$. The set of rank-one partially symmetric tensors is known as the Segre-Veronese variety of $\mathbb{P}U^* \times \mathbb{P}V^*$ embedded by $\mathcal{O}(1,2)$. Therefore, the Zariski closure of the set of partially symmetric tensors of rank at most r is the rth secant variety of the Segre-Veronese variety. We have the following analogue of Definition 4.1.2.

Definition 4.2.1. *The* partially symmetric border rank *of* $[x] \in \mathbb{P}(U^* \otimes S^2V^*)$ *is the minimal* r *such that* $[x]$ *is in* $\sigma_r(\mathrm{Seg}(\mathbb{P}U^* \times v_2(\mathbb{P}V^*)))$.

Providing an analogue of the equations from Definition 4.1.4 is a bit more subtle when restricted to partially symmetric 3-tensors. In fact, it is necessary to refine the equations if we hope to produce ideals which are radical. To see this, consider the case where x is a partially symmetric $3 \times n \times n$ tensor. For such an x, the matrix representing $\psi_{1,x}$ has the form

$$\psi_{1,x} = \begin{pmatrix} 0 & A_3 & -A_2 \\ -A_3 & 0 & A_1 \\ A_2 & -A_1 & 0 \end{pmatrix}$$

where the A_i are symmetric $n \times n$-matrices. Since $\psi_{1,x}$ is a skew-symmetric matrix, all of the principal minors in $I_{\kappa_1 \leq c_1}$ are squares.

More generally, if $m = 4j+3$, then $\psi_{2j+1,x}\colon V \otimes \bigwedge^{2j+1} U^* \to V^* \otimes \bigwedge^{2j+2} U^*$ is represented by a skew-symmetric matrix in appropriate coordinates. Thus, the condition that $\psi_{2j+1,x}$ has rank at most an even integer c_{2j+1} is defined algebraically by the principal $(c_{2j+1}+2)\times(c_{2j+1}+2)$-Pfaffians of $\psi_{2j+1,x}$. These Pfaffians have degree $c_{2j+1}/2+1$, whereas the $(c_{2j+1}+1)\times(c_{2j+1}+1)$-minors have degree $c_{2j+1}+1$.

To encode this skew-symmetry into our matrix equations in the case of partially symmetric tensors, we introduce the following analogue of Definition 4.1.4.

Definition 4.2.2. *Let $I_{\kappa_j \le c_j}$ in $S^\bullet(U \otimes S^2 V)$ denote the ideal generated by the $(c_j+2)\times(c_j+2)$-Pfaffians of $\psi_{j,x}$, if $j = (m-1)/2$, j is an odd integer, and c_j is even. Otherwise, $I_{\kappa_j \le c_j}$ denotes the specialization of the ideal in Definition 4.1.4. As in Definition 4.1.4, for a vector c, $I_{\kappa \le c}$ is defined to be the ideal generated by the $I_{\kappa_j \le c_j}$ for all j, and $\Sigma_{\kappa_i \le c_i}$ and $\Sigma_{\kappa \le c}$ are the subschemes of $\mathbb{P}(U^* \otimes S^2 V^*)$ defined by $I_{\kappa_i \le c_i}$ and $I_{\kappa \le c}$ respectively.*

Note that for partially symmetric tensors x, we have $\kappa_j(x) = \kappa_{m-1-j}(x)$ and likewise $I_{\kappa_j(x) \le c_j} = I_{\kappa_{m-1-j} \le c_j}$ for all j. With notation in Definition 4.2.2, we also obtain the following analogue of Proposition 4.1.5.

Proposition 4.2.3. *Fix $r \ge 1$. Let X be the Segre-Veronese variety of $\mathbb{P}(U^*) \times \mathbb{P}(V^*)$ in $\mathbb{P}(U^* \otimes S^2 V^*)$ and let c be the vector defined by $c_j = r\binom{m-1}{j}$ for $0 \le j \le m-1$. Then $\sigma_r(X) \subseteq \Sigma_{\kappa \le c}$.*

Remark 4.2.4. Although the rest of this chapter concerns partially symmetric $3\times n\times n$ tensors, we note that the equations given in Definition 4.2.2 would be insufficient to generate the ideal of the secant varieties for $m \ge 4$. For example, consider the space of partially symmetric $4 \times n \times n$ tensors. If x is such a tensor, then the matrix representing $\psi_{2,x}$ has the form

$$\psi_{2,x} = \begin{pmatrix} -A_4 & 0 & 0 & \mathbf{0} & \mathbf{A_3} & \mathbf{-A_2} \\ 0 & -A_4 & 0 & \mathbf{-A_3} & \mathbf{0} & \mathbf{A_1} \\ 0 & 0 & -A_4 & \mathbf{A_2} & \mathbf{-A_1} & \mathbf{0} \\ A_1 & A_2 & A_3 & 0 & 0 & 0 \end{pmatrix}$$

where each A_i is an $n \times n$ symmetric matrix. If x has border rank at most r, then $\psi_{1,x}$ will have rank at most $3r$ by Proposition 4.2.3. However, the bold submatrix in the upper right will have rank at most $2r$. Moreover, since the bold submatrix is skew-symmetric, the condition that it has rank at most $2r$ is given by the vanishing of its $(2r+2)\times(2r+2)$-principal Pfaffians. Thus, the defining ideal of the rth secant variety must contain these Pfaffians, as well as 3 other sets of Pfaffians which arise by symmetry. Since the Pfaffians have degree $r+1$, they can not be in the ideal of the $(3r+1)\times(3r+1)$-minors.

In effect, these Pfaffians amount to the generators of $I_{\kappa_1 \le 2r}$ applied to a $3 \times n \times n$ subtensor. In the literature on tensors, this process for producing equations on larger tensors by applying known equations to all subtensors is known as *inheritance*. See [31, §2.1] for a precise definition in the language of representation theory. The above analysis shows how the inheritance of κ-equations can produce new equations beyond the κ-equations themselves.

Proposition 4.2.5. *As a Schur module, we have the following decomposition of the generators $I_{\kappa_0 \leq r}$ into irreducible representations of* $\mathrm{GL}(U) \times \mathrm{GL}(V)$

$$(I_{\kappa_0 \leq r})_{r+1} = \bigoplus_{|\pi|=r+1} S_\pi U \otimes S_{\pi'+1^{r+1}} V,$$

where π' is the conjugate partition to π, and $1^{r+1} = (1, \ldots, 1)$ is the partition with $r+1$ parts.

In the proof, we will need the following observation.

Lemma 4.2.6. *Suppose π is a partition of d and suppose A is a vector space. If $S_\lambda A$ is a module occurring in the decomposition of $S_\pi(S^2 A)$ then λ has at most d parts.*

Proof. Since π is a partition of d, we have an inclusion $S_\pi(S^2 A) \subset (S^2 A)^{\otimes d}$. By inductively applying the Pieri formula to $(S^2 A)^{\otimes d} = (S^2 A)^{\otimes d-1} \otimes S^2 A$, we see that every module in the decomposition of $(S^2 A)^{\otimes d}$ can have at most d parts. □

Proof of Proposition 4.2.5. After choosing bases of U and V, we may view the map $\psi_{0,x}$ as a matrix of linear forms in $S^\bullet(U \otimes S^2 V)$. By Lemma 4.1.11, $(I_{\kappa_0 \leq r})_{r+1}$ is the image of the $\mathrm{GL}(U) \times \mathrm{GL}(V)$-equivariant morphism

$$\Phi_0 \colon \bigwedge^{r+1} V \otimes \bigwedge^{r+1} (U \otimes V) \to S^{r+1}(U \otimes S^2(V)),$$

which sends an indecomposable basis element in the source to the corresponding minor in the polynomial ring. The first step of our proof is to show that only those representations of the form $S_\pi U \otimes S_{\pi'+1^{r+1}} V$ appear in both the source and target of Φ_0. By Schur's Lemma, this will provide a necessary condition on the representations which can appear in the decomposition of $(I_{\kappa_0 \leq r})_{r+1}$. The second step of our proof is to show that each such representation actually arises; for this, we produce an explicit nonzero minor of $\psi_{0,x}$ that is in the image of Φ_0 restricted to $S_\pi U \otimes S_{\pi'+1^{r+1}} V$, so that Φ_0 restricted to $S_\pi U \otimes S_{\pi'+1^{r+1}} V$ is nonzero.

For the first step, suppose that $S_\pi U \otimes S_\lambda V$ is a module in $S^{r+1}(U \otimes S^2 V)$. If we apply the Cauchy decomposition formula to $S^d(U \otimes S^2 V)$, and consider the resulting modules as $\mathrm{GL}(U)$-representations, then we must have $S_\pi U \otimes S_\lambda V$ contained in the summand $S_\pi U \otimes S_\pi(S^2 V)$. In particular we must have $S_\lambda V \subset S_\pi(S^2 V)$. Therefore, by Lemma 4.2.6, λ has at most $r+1$ parts.

On the other hand, we can use the skew Cauchy formula to decompose

$$\bigwedge^{r+1} V \otimes \bigwedge^{r+1} (U \otimes V) = \bigwedge^{r+1} V \otimes \bigoplus_{|\pi|=r+1} S_\pi U \otimes S_{\pi'} V.$$

Applying the Pieri rule to $\bigwedge^{r+1} V \otimes S_{\pi'} V$, we see that all of the summands have more than $r+1$ parts except for $S_{\pi'+1^{r+1}} V$. Therefore, the decomposition of $I_{\kappa_0 \leq r}$ must consist only

of the modules $S_\pi U \otimes S_{\pi'+1^{r+1}} V$, where π is a partition of $r+1$, and each such module can occur with multiplicity at most one.

For the second step, fix a partition π of $r+1$. Suppose that $u_1, \ldots, u_m$ is our ordered basis of U and $v_1, \ldots, v_n$ is our ordered basis of V. Consider the indecomposable basis element

$$z_\pi = (v_1 \wedge \cdots \wedge v_{r+1}) \otimes \big((u_1 \otimes v_1) \wedge \cdots \wedge (u_1 \otimes v_{\pi_1}) \wedge (u_2 \otimes v_1) \wedge \cdots \wedge (u_2 \otimes v_{\pi_2}) \\ \wedge \cdots \wedge (u_m \otimes v_1) \wedge \cdots \wedge (u_m \otimes v_{\pi_m})\big)$$

in $\bigwedge^{r+1} V \otimes \bigwedge^{r+1} (U \otimes V)$. We claim that z_π is in $S_\pi U \otimes S_{\pi'+1^{r+1}} V$, and, in fact is a non-zero highest weight vector in that representation. The vector z_π is non-zero because z_π is the tensor product of two tensors, each constructed as an exterior product of linearly independent tensors and hence non-zero. It is clear that z_π has weights π and $\pi' + 1^{r+1}$ in U and V respectively, with respect to our chosen bases. Moreover, replacing v_i by v_j or u_i by u_j, with $j < i$ in either case, would result in a repeated term in the exterior product, and thus any raising operator would send z_π to zero, so z_π is a highest weight vector.

Let M_π be the submatrix of the block matrix $\psi_{0,x}^T = \begin{pmatrix} A_1 & \cdots & A_m \end{pmatrix}$ defined by selecting the first $r+1$ rows and the first π_i columns of the ith block for each $i \leq n$. Then the map Φ_0 sends z_π to the determinant of M_π. For appropriate choices for A_i, we can make M_π equal the identity matrix, and therefore $\Phi(z_\pi) = \det(M_\pi)$ is nonzero. □

When $\dim(U) = 3$, we similarly produce a formula for the decomposition of the modules generating $I_{\kappa_1 \leq 2r}$ in Proposition 4.2.8.

Remark 4.2.7. Taken together, Propositions 4.2.5 and 4.4.2 provide a complete Schur module description of the generators of the ideal of any secant variety of the $\mathbb{P}^1 \times \mathbb{P}^{n-1}$ embedded by $\mathcal{O}(1,2)$. Similarly, Propositions 4.2.5 and 4.2.8, together with Theorem 4.4.3, provide a complete Schur module description of the generators of the ideal of the rth secant variety of $\mathbb{P}^2 \times \mathbb{P}^{n-1}$ embedded by $\mathcal{O}(1,2)$ for r at most 5.

Proposition 4.2.8. *Suppose that* $\dim(U)$ *is* 3. *As a Schur module, we have the following decomposition of the generators* $I_{\kappa_1 \leq 2r}$ *into irreducible representations of* $\mathrm{GL}(U) \times \mathrm{GL}(V)$

$$(I_{\kappa_1 \leq 2r})_{r+1} = \bigoplus_{|\pi| = r+1} S_\pi U \otimes S_{(3)^{r+1} - \pi'} V,$$

where π' *is the conjugate partition to* π. *In order for the summand to be non-zero,* π *must have at most* 3 *parts, and* π_3 *must be at least* $m - r + 1$ *(if the latter is positive).*

Proof. We consider the Pfaffians of a matrix representing the map $\psi_{1,x} \colon V \otimes U^* \to V^* \otimes \bigwedge^2 U^*$. In order to view $\psi_{1,x}$ as a skew-symmetric transformation, we identify $\bigwedge^2 U^*$ in the target with $U \otimes \bigwedge^3 U^*$. Then, we can view $\psi_{1,x}$ as a skew-symmetric form on $V \otimes U^*$, taking values in $\bigwedge^3 U^*$. Equivalently, a choice of a nonzero element in $\bigwedge^3 U$ gives a $\mathbb{C}$-valued skew-symmetric form.

The remainder of our proof essentially follows the same two steps as the proof of Proposition 4.2.5. The space of $(2r+2)\times(2r+2)$-Pfaffians of a skew-symmetric form on $V\otimes U^*$ is isomorphic to $\bigwedge^{2r+2}(V\otimes U^*)$. Therefore, similar to Lemma 4.1.11, $(I_{\kappa_1\leq 2r})_{r+1}$ is the image of the map

$$\Phi_1\colon\ \bigwedge^{2r+2}(V\otimes U^*)\otimes(\bigwedge^3 U)^{r+1}\to S^{r+1}\left(U\otimes S^2(V)\right),$$

which sends an indecomposable basis element to the corresponding Pfaffian. Note that the exponent of $r+1$ in $(\bigwedge^3 U)^{r+1}$ corresponds to the Pfaffian having degree $r+1$. First, we show that only modules of the form $S_\pi U\otimes S_{(3)^{r+1}-\pi'}V$ can arise as a representation in both the source and target of Φ_1. Second, we consider Φ_1 restricted to $S_\pi U\otimes S_{(3)^{r+1}-\pi'}V$ and we produce a Pfaffian in the image and explicitly show that it is non-zero. By Schur's Lemma, this will show that every such representation actually arises in the decomposition of $(I_{\kappa_1\leq 2r})_{r+1}$.

For the first step, we use the skew Cauchy formula to decompose the source of Φ_1 as

$$\bigwedge^{2r+2}(U^*\otimes V)\otimes(\bigwedge^3 U)^{r+1}=\bigoplus_{|\lambda|=2r+2}S_\lambda U^*\otimes S_{\lambda'}V\otimes(\bigwedge^3 U)^{r+1}=\bigoplus_{|\lambda|=2r+2}S_{(r+1)^3-\lambda}U\otimes S_{\lambda'}V,$$

where we have used the duality formula from Remark 4.1.13 for the second equality. Every module in the source of Φ_1 is thus of the form $S_{(r+1)^3-\lambda}U\otimes S_{\lambda'}V$, where λ is a partition of $2r+2$. We make the substitution $\lambda=(r+1)^3-\pi$ to arrive at the expression in the statement of the proposition.

For the second step, we explicitly produce a non-zero Pfaffian of $\psi_{1,x}$ in the image of Φ_1 restricted to $S_{\pi'}U\otimes S_{(r+1)^3-\pi}V$, and thus confirm that every module of the form $S_{\pi'}U\otimes S_{(r+1)^3-\pi}V$ (for appropriate π) occurs in the decomposition of $(I_{\kappa_1\leq 2r})_{r+1}$. Suppose that u_1,u_2,u_3 is our ordered basis for U and $v_1,\ldots,v_n$ is our ordered basis for V. Let $\pi=(\pi_1,\pi_2,\pi_3)$ be a a partition of $r+1$ with no more than three parts, and let $\lambda=(r+1)^3-\pi=(r+1-\pi_3,r+1-\pi_2,r+1-\pi_1)$, as before. Consider the element

$$z_\pi=\big((u_1^*\otimes v_1)\wedge\ldots\wedge(u_1^*\otimes v_{\lambda_3})\wedge(u_2^*\otimes v_1)\wedge\ldots\wedge(u_2^*\otimes v_{\lambda_2})\wedge(u_3^*\otimes v_1)\wedge\ldots\wedge(u_3^*\otimes v_{\lambda_1})\big)\otimes(u_1\wedge u_2\wedge u_3)^{\otimes r+1}$$

in $\bigwedge^{2r+2}(U^*\otimes V)\otimes(\bigwedge^3 U)^{r+1}$, where the u_i^* form the dual basis to the u_i. Note that z_π is non-zero since the vectors in each exterior product are linearly independent. Now, we will show that z_π is a highest weight vector in $S_\pi U\otimes S_{\lambda'}V$.

First, we claim that z_π has weight (π,λ'). By counting the occurrences of v_i in z_π, it is clear that the weight in the V-factor is λ'. For the U factor, we note that the weight of u_i^* is the negative of that of u_i, so that z_π has weight $(r+1-\lambda_3,r+1-\lambda_2,r+1-\lambda_1)$, which is equal to π.

Second, we must show that any raising operator will send z_π to zero, which will imply that z_π is a highest weight vector. In the V factor, sending v_i to v_j with $j<i$ would force a repeated vector in the exterior product. Likewise, a raising operator applied to the U factor

would send u_i^* to u_j^* with $j > i$, which would, again, create a repeated factor in the exterior product.

Finally, we check that $\Phi_1(z_\pi) \neq 0$. Let M_π be the principal submatrix obtained from $\psi_{1,x}$ by selecting the rows and columns with indices $\{1, \ldots, \lambda_3, n{+}1, \ldots, n{+}\lambda_2, 2n{+}1, \ldots, 2n{+}\lambda_1\}$. Then $\Phi_1(z_\pi)$ equals the Pfaffian of M_π. To check that the Pfaffian of M_π is nonzero, it suffices to produce a specialization of M_π which has full rank. Note that if B_i is the appropriate submatrix from the upper-left corner of A_i, then M_π has the following shape

$$M_\pi = \begin{array}{c} \\ \lambda_3 \\ \lambda_2 \\ \lambda_1 \end{array} \begin{array}{c} \begin{array}{ccc} \lambda_3 & \lambda_2 & \lambda_1 \end{array} \\ \begin{pmatrix} 0 & B_3 & -B_2 \\ -B_3^t & 0 & B_1 \\ B_2^t & -B_1^t & 0 \end{pmatrix} \end{array}.$$

We have $\lambda_1 = \pi_1 + \pi_2$, $\lambda_2 = \pi_1 + \pi_3$, $\lambda_3 = \pi_2 + \pi_3$. If we specialize the A_i such that the B_i are as follows

$$B_1 = \begin{array}{c} \\ \pi_1 \\ \pi_3 \end{array} \begin{array}{c} \begin{array}{cc} \pi_2 & \pi_1 \end{array} \\ \begin{pmatrix} 0 & \mathrm{Id}_{\pi_1} \\ 0 & 0 \end{pmatrix} \end{array}, \quad B_2 = \begin{array}{c} \\ \pi_3 \\ \pi_2 \end{array} \begin{array}{c} \begin{array}{cc} \pi_2 & \pi_1 \end{array} \\ \begin{pmatrix} 0 & 0 \\ -\mathrm{Id}_{\pi_2} & 0 \end{pmatrix} \end{array}, \text{ and } B_3 = \begin{array}{c} \\ \pi_3 \\ \pi_2 \end{array} \begin{array}{c} \begin{array}{cc} \pi_1 & \pi_3 \end{array} \\ \begin{pmatrix} 0 & \mathrm{Id}_{\pi_3} \\ 0 & 0 \end{pmatrix} \end{array},$$

then the specialization of M_π has full rank, since it is the standard block skew-symmetric matrix. □

Example 4.2.9. Consider the case $n = 4$ and $c_1 = 10$. Since $\psi_{1,x}$ is a skew-symmetric 12×12 matrix, we expect the ideal $I_{\kappa_1 \leq 10}$ to be a principal ideal, generated by a polynomial in $S^6(U \otimes S^2(V))$. Applying Proposition 4.2.8, we must have a sum over partitions π of 6 such that $(3)^6 - \pi'$ has at most 4 parts. This forces π' to equal $(3, 3)$, and thus $\pi = (2, 2, 2)$. The generators of $I_{\kappa_1 \leq 10}$ are therefore equal to the 1-dimensional representation $S_{2,2,2}(U) \otimes S_{3,3,3,3}(V)$, corresponding to the Pfaffian of $\psi_{1,x}$. □

Example 4.2.10. Consider the case $n = 4$ and $c = (3, 6, 3)$, which we revisit in Example 4.4.4. Propositions 4.2.5 and 4.2.8 give us the decompositions:

$$(I_{\kappa_0 \leq 3})_4 = S_{2,2}S_{3,3,1,1} \oplus \mathbf{S_{2,1,1}S_{4,2,1,1}} \oplus S_{3,1}S_{3,2,2,1} \oplus S_4S_{2,2,2,2},$$
$$(I_{\kappa_1 \leq 6})_4 = S_{2,2}S_{3,3,1,1} \oplus \mathbf{S_{2,1,1}S_{3,3,2}} \oplus S_{3,1}S_{3,2,2,1} \oplus S_4S_{2,2,2,2}.$$

Both modules are 495-dimensional and consist of quartic polynomials. The ideal $I_{\kappa \leq (3,6,3)}$, which equals $I_{\kappa_6 \leq 6} + I_{\kappa_0 \leq 3}$ by definition, is generated by the 630-dimensional space of quartics obtained by taking the sum of the above decompositions. Notice that, due to the highlighted modules in the above decompositions, neither $I_{\kappa_0 \leq 3}$ nor $I_{\kappa_1 \leq 6}$ belongs to the other. In particular, the 4×4-minor formed by taking columns 1, 2, 5, and 9 from the flattening $\psi_{0,x}$, which is the transpose of (4.1), is not in the ideal of Pfaffians. On the other hand, the Pfaffian formed by taking the rows and columns of (4.2) with indices 1, 2, 5, 6, 7, 9, 10, and 11 is not contained in the ideal of the minors. □

Notice that the formulas in Propositions 4.2.5 and 4.2.8 are multiplicity free, unlike, for example, the ideal generators computed in Example 4.1.15.

4.3 Subspace varieties of partially symmetric tensors

We next give a geometric interpretation for the varieties $\Sigma_{\kappa_0 \leq r}$. These are the partially symmetric analogues of the subspace varieties defined in [34, Defn. 1]; Proposition 4.3.3 forms an analogue to [34, Thm. 3.1].

Definition 4.3.1. *The* subspace variety $\mathrm{Sub}_{m',n'}$ *is the variety of tensors* $x \in (U^* \otimes S^2 V^*)$ *such that there exist vector spaces* $\widetilde{U}^* \subset U^*$ *and* $\widetilde{V}^* \subset V^*$ *of dimensions* m' *and* n' *respectively with* $x \in (\widetilde{U}^* \otimes S^2 \widetilde{V}^*) \subset (U^* \otimes S^2 V^*)$.

Remark 4.3.2. The variety $\mathrm{Sub}_{m',n'}$ has a nice desingularization, analogous to that used to prove the results in [34, §3]. Consider the product of Grassmannians $\mathrm{Gr}(m', U^*) \times \mathrm{Gr}(n', V^*)$, and let E be the total space of the vector bundle $\mathcal{R}_U \otimes S^2 \mathcal{R}_V$, where $\mathcal{R}_U$ and $\mathcal{R}_V$ are the tautological subbundles over $\mathrm{Gr}(m', U^*)$ and $\mathrm{Gr}(n', V^*)$, respectively. Then there is a natural map $\pi \colon E \to \mathrm{Sub}_{m',n'}$, which desingularizes $\mathrm{Sub}_{m',n'}$. Moreover, one can verify that Weyman's geometric technique can be applied in this situation [52, §5]. In fact, a straightforward adaptation of the argument in [34, Thm. 3.1] implies that $\mathrm{Sub}_{m',n'}$ is normal with rational singularities.

We next directly calculate the generators of the ideal of the subspace variety when $m' = m$, which is the case we need.

Proposition 4.3.3. *The defining ideal of* $\mathrm{Sub}_{m,n'}$ *equals* $I_{\kappa_0 \leq n'}$.

The following lemma plays a crucial technical role in the proof of both Proposition 4.3.3 and Theorem 4.4.3, as it provides a criterion for determining the reducedness of some of the ideals that we are studying.

Lemma 4.3.4. *Let* Z *be a* $\mathrm{GL}(U) \times \mathrm{GL}(V)$*-invariant reduced subscheme of the desingularization* E *from Remark 4.3.2. Suppose that* I *is an invariant ideal in* $S^\bullet(U \otimes S^2 V)$*, which contains the ideal of* $\mathrm{Sub}_{m',n'}$*, and whose pullback to* E *defines* Z*. Then* I *is the ideal of* $\pi(Z)$*. In particular,* I *is a radical ideal.*

Proof. Let $J \subseteq S^\bullet(U \otimes S^2 V)$ be the defining ideal of $\pi(Z)$. Recall that $q \colon E \to \mathrm{Gr}(m', U^* \times \mathrm{Gr}(n', V^*)$ is the total space of a vector bundle, as a defined in Remark 4.3.2. Our set-up is the following commutative diagram:

$$\begin{array}{ccccc} Z & \hookrightarrow & E & & \\ \downarrow & & \downarrow{\scriptstyle \pi} & & \\ \pi(Z) & \hookrightarrow & \mathrm{Sub}_{m',n'} & \hookrightarrow & (U^* \otimes S^2 V^*)\,. \end{array}$$

The hypothesis that the pullback of I defines Z in E guarantees $I \subseteq J$. We thus need to show the reverse inclusion.

A point $P \in \mathrm{Gr}(m', U^*) \times \mathrm{Gr}(n', V^*)$ corresponds to vector subspaces $\widetilde{U}^* \subset U^*$ and $\widetilde{V}^* \subset V^*$, and this induces a surjection of rings $\phi_P \colon S^\bullet(U \otimes S^2 V) \to S^\bullet(\widetilde{U} \otimes S^2 \widetilde{V})$. The fiber $q^{-1}(P) \subseteq E$ is isomorphic to the affine space $(\widetilde{U}^* \otimes S^2 \widetilde{V}^*)$, and we define $\widetilde{Z}_P := Z \cap q^{-1}(P)$.

We claim that a polynomial $f \in S^\bullet(U \otimes S^2 V)$ belongs to J if and only if $\phi_P(f)$ vanishes on $\pi(\widetilde{Z}_P)$ for every choice of P. The "only if" direction of the claim is straightforward. For the "if" direction, we first note that, since Z is assumed to be reduced, the condition $f \in J$ is equivalent to the condition that the pullback of f vanishes on every point $y \in \pi(Z)$. This is in turn equivalent to the condition f vanishes on each point of Z, which is implied by the hypothesis that $\phi_P(f)$ vanishes on $\pi(\widetilde{Z}_P)$ for each P.

In fact, since $\mathrm{GL}(U) \times \mathrm{GL}(V)$ acts transitively on $\mathrm{Gr}(m', U^*) \times \mathrm{Gr}(n', V^*)$, we conclude that f vanishes on $\pi(Z)$ if and only if $\phi_P(g \cdot f)$ vanishes on $\pi(\widetilde{Z})$ for *any* fixed choice of P and all $g \in \mathrm{GL}(U) \times \mathrm{GL}(V)$. Therefore, J is the sum of all irreducible Schur submodules M of $S^\bullet(U \otimes S^2 V)$ such that $\phi_P(M)$ vanishes on $\widetilde{Z}_P$. For the rest of the proof we fix $\widetilde{U}^*$ and $\widetilde{V}^*$ and denote the induced map ϕ_P by ϕ and $\widetilde{Z}_P$ by $\widetilde{Z}$.

To show that $J \subseteq I$, let M be an irreducible Schur submodule of the ideal J; we want to show that M is contained in our given ideal I. If M is isomorphic to $S_\mu U \otimes S_\nu V$, then the construction of Schur modules implies that $\phi(M)$ is isomorphic as a $\mathrm{GL}(\widetilde{U}) \times \mathrm{GL}(\widetilde{V})$-representation to $S_\mu \widetilde{U} \otimes S_\nu \widetilde{V}$, which is either trivial or an irreducible representation. We know that $\phi(M)$ vanishes on $\widetilde{Z}$ and thus, since I pulls back to the defining ideal of Z, it follows that $\phi(M)$ is contained in $\phi(I)$. There is thus an irreducible Schur submodule $N \subset I$ such that $\phi(N) = \phi(M)$, and hence N is isomorphic to $S_\mu U \otimes S_\nu V$. If N equals M, then we are done. Otherwise, ϕ sends the submodule $N + M$, spanned by two copies of $S_\mu U \otimes S_\nu V$, to the submodule $\phi(M)$, which is a single copy of $S_\mu \widetilde{U} \otimes S_\nu \widetilde{V}$. Thus, some subrepresentation L of $N + M$ is sent to zero by ϕ. Since L is a representation in the kernel of ϕ, L belongs to the ideal of $\mathrm{Sub}_{m',n'}$, which is contained in I by assumption. It follows that I contains the span of N and L, and hence I contains M. We conclude that $I = J$ as desired. □

Proof of Proposition 4.3.3. First, we prove the claim set-theoretically. The $(n'+1) \times (n'+1)$ minors of $\psi_{0,x}$ vanish if and only if the map has rank at most n'. By linear algebra, this is equivalent to the existence of a change of basis in which $\psi_{0,x}$ uses only the first n' rows, which is the definition of $\mathrm{Sub}_{m,n'}$.

Second, we show that $I_{\kappa_0 \leq n'}$ is radical when $n' = n-1$. Note that $\mathrm{Sub}_{m,n-1}$ has dimension $m\binom{n}{2} + n$, and thus $\mathrm{Sub}_{m,n-1}$ and $\Sigma_{\kappa_0 \leq n-1}$ have codimension $mn - n + 1$. This is the same as the codimension of the maximal minors of a generic $n \times mn$ matrix, so $I_{\kappa_0 \leq n-1}$ is Cohen-Macaulay by [20, Thm. 18.18], and it suffices to show that the affine cone over $\Sigma_{\kappa_0 \leq n-1}$ is reduced at some point. Consider a neighborhood of the point $u_1 \otimes v_1^2 + \cdots + u_1 \otimes v_{n-1}^2$. In

coordinates around this point, $I_{\kappa_0 \leq n-1}$ consists of the maximal minors of the $n \times mn$ matrix:

$$\begin{bmatrix} 1+x_{1,1,1} & \cdots & x_{1,1,n-1} & x_{1,1,n} & x_{2,1,1} & \cdots & x_{2,1,n} & \cdots & x_{m,1,n} \\ \vdots & \ddots & \vdots & \vdots & \vdots & & \vdots & \cdots & \vdots \\ x_{1,1,n-1} & \cdots & 1+x_{1,n-1,n-1} & x_{1,n-1,n} & \vdots & & \vdots & \cdots & \vdots \\ x_{1,1,n} & \cdots & x_{1,n-1,n} & x_{1,n,n} & x_{2,1,n} & \cdots & x_{2,n,n} & \cdots & x_{m,n,n} \end{bmatrix}.$$

The $mn-n+1$ minors which use the first $n-1$ columns form part of a regular sequence, and thus the affine cone over $\Sigma_{\kappa_0 \leq n-1}$ is reduced in a neighborhood of this point. Since $I_{\kappa_0 \leq n-1}$ is a Cohen-Macaulay ideal, it follows that $\Sigma_{\kappa_0 \leq n-1}$ is everywhere reduced.

Third, we show that $I_{\kappa_0 \leq n'}$ defines $\mathrm{Sub}_{m,n'}$ for arbitrary n'. By reverse induction on n', we assume that $I_{\kappa_0 \leq n'}$ equals the ideal of $\mathrm{Sub}_{m,n'}$, and we seek to show equality for $n'-1$. We will apply Lemma 4.3.4, where E is the vector bundle over $\mathrm{Gr}(m, U^*) \times \mathrm{Gr}(n', V^*) = \mathrm{Gr}(n', V^*)$ desingularizing $\mathrm{Sub}_{m,n'}$ as in Remark 4.3.2. Note that, by cofactor expansion, $I_{\kappa_0 \leq n'-1}$ contains $I_{\kappa_0 \leq n'}$, which is the ideal of $\mathrm{Sub}_{m,n'}$ by the inductive hypothesis. We describe Z, which is defined by the pullback of $I_{\kappa_0 \leq n'-1}$, on a local trivialization $(U^* \otimes S^2\widetilde{V}^*) \times Y$ of the vector bundle E, where $\widetilde{V}$ is n'-dimensional and Y is an open subset of $\mathrm{Gr}(n', V^*)$. The pullbacks of the $(n'-1) \times (n'-1)$ minors of $\psi_{0,x}$ do not involve the base Y, and are the $\kappa_0 \leq n'-1$ equations applied to $U^* \otimes S^2\widetilde{V}^*$. These are maximal minors of the matrix $\psi_{0,x}$ for $U^* \otimes S^2\widetilde{V}^*$, and hence they define a reduced subscheme of $U^* \otimes S^2\widetilde{V}^*$ by the previous paragraph. In the local trivialization, their scheme is the product of this reduced scheme with Y, so the preimage of $\Sigma_{\kappa_0 \leq n'}$ in E is reduced. We may thus apply Lemma 4.3.4 and conclude that $I_{\kappa_0 \leq n'-1}$ is reduced. □

Remark 4.3.5. When $m' < m$, the ideal of $\mathrm{Sub}_{m',n'}$ is similarly generated by the sum of $I_{\kappa_0 \leq n'}$ and the irreducible modules in $\bigwedge^{m'+1} U \otimes \bigwedge^{m'+1}(S^2V)$. A decomposition of the latter space, in somewhat different notation, can be found at [41, p. 47].

4.4 Secant varieties of $\mathbb{P}^2 \times \mathbb{P}^{n-1}$ embedded by $\mathcal{O}(1,2)$

In this section, we prove the main result of the chapter, which is to show that the equations given in Definition 4.1.4 generate the defining ideal of the rth secant variety of $\mathbb{P}^2 \times \mathbb{P}^{n-1}$ embedded by $\mathcal{O}(1,2)$ when $r \leq 5$.

We first consider a simpler case: the secant varieties of $\mathbb{P}^1 \times \mathbb{P}^{n-1}$ embedded by $\mathcal{O}(1,2)$. All such secant varieties are defined by κ-equations, which, in this case, are simply the minors of flattenings. The analogous statement for non-symmetric matrices appears as Theorem 1.1 in [34]. However, we know of no proof in the literature for the case of partially symmetric tensors, so we provide one below.

Definition 4.4.1. *For a variety $X \subseteq \mathbb{P}^N$ we denote the affine cone of X in $\mathbb{A}^{N+1}$ by $\widehat{X}$.*

Proposition 4.4.2. *Suppose $m = 2$, and let $Y \subseteq \mathbb{P}(U^* \otimes S^2V^*)$ be the image of $\mathbb{P}(U^*) \times \mathbb{P}(V^*)$ under the embedding by $\mathcal{O}(1,2)$. For any $r > 1$, and any n, the secant variety $\sigma_r(Y)$ is defined ideal-theoretically by the ideal $I_{\kappa_0 \leq r}$.*

Proof. We have $\widehat{\sigma_r}(Y) \subseteq \widehat{\Sigma}_{\kappa_0 \leq r} = \mathrm{Sub}_{2,r}$, where the inclusion follows from Proposition 4.2.3 and the equality follows from Proposition 4.3.3. Since $\mathrm{Sub}_{2,r}$ is integral, it suffices to prove that $\widehat{\sigma_r}(Y)$ and $\mathrm{Sub}_{2,r}$ have the same dimension. By [1, Cor. 1.4(i)], the former has the expected dimension $rn + r$. From the definition of $\mathrm{Sub}_{2,r}$, we can compute its dimension to be $r(n-r) + 2\binom{r+1}{2} = rn + r$. □

For the remainder of this section, we restrict to the case when $\dim U^* = 3$, which is the next partially symmetric case. We let $\dim V^* = n$, and we consider partially symmetric tensors $x \in U^* \otimes S^2V^*$. We fix $\mathbb{P}^N := \mathbb{P}(U^* \otimes S^2(V^*))$ and we let $X \subset \mathbb{P}^N$ denote the embedding of $\mathbb{P}(U^*) \times \mathbb{P}(V^*)$ by $\mathcal{O}(1,2)$. Let $S := S^\bullet(U \otimes S^2(V))$ be the homogeneous coordinate ring of $\mathbb{P}^N$, which contains the ideals $I_{\kappa_j \leq c_j}$ and $I_{\kappa \leq c}$ as in Definition 4.2.2.

Theorem 4.4.3. *For $r \leq 5$, the defining ideal of the variety $\sigma_r(X)$ is $I_{\kappa \leq (r,2r,r)}$.*

Our method of proof is as follows. When n equals r, we relate the ideal $I_{\kappa_1 \leq 2r}$ to the ideal of commuting symmetric matrices. This is a variant of an idea which has appeared in several instances previously [47, 43, 3]. This relation only holds away from a certain closed subvariety of $\mathbb{P}^N$, and in order to extend to all of $\mathbb{P}^N$, we need a bound on the dimension of this variety. Such a bound is given in in [22, §5], and only holds for $r \leq 5$. Finally, we reduce the general case to the case of $n = r$, using Lemma 4.3.4.

Before the proof, we examine the secant varieties of $\mathbb{P}^2 \times \mathbb{P}^3$ in more detail.

Example 4.4.4. Let $X \subseteq \mathbb{P}^{29}$ be the image of $\mathbb{P}^2 \times \mathbb{P}^3$ embedded by $\mathcal{O}(1,2)$. The defining ideal of $\sigma_5(X)$ was previously known. The secant variety $\sigma_5(X)$ is deficient, and is in fact a hypersurface in $\mathbb{P}^{29}$. This hypersurface is defined by the Pfaffian of $\psi_{1,x}$ [43, Thm. 4.1].

For non-symmetric tensors, [34, Thm. 1.1] illustrates that the defining ideal for the second secant variety is generated by the 3×3 minors of the various flattenings. This suggests that a similar result holds in the partially symmetric case, although we know of no explicit reference for such a result. Nevertheless, in the situation of this example, a direct computation with [26] confirms that the defining ideal of $\sigma_2(X)$ is indeed generated by the 3×3 minors of the flattening $\psi_{0,x}$ and by the 3×3 minors of the other flattening of x, i.e. by considering x in $\mathrm{Hom}(U, S^2V^*)$. Theorem 4.4.3 provides an alternate description, illustrating that the 3×3 minors of $\psi_{0,x}$ and the 6×6 principal Pfaffians of $\psi_{1,x}$ also generate the ideal of $\sigma_2(X)$.

As far we are aware, the defining ideals for $\sigma_3(X)$ and $\sigma_4(X)$ were not previously known. The defining ideal of $\sigma_4(X)$ is given by $I_{\kappa \leq (4,8,4)}$. Since the ideals $I_{\kappa_0 \leq 4}$ and $I_{\kappa_2 \leq 4}$ are trivial, this equals the ideal $I_{\kappa_1 \leq 8}$. Thus, $\sigma_4(X)$ is defined by the 10×10 principal Pfaffians of $\psi_{1,x}$.

The case of $\sigma_3(X)$ is perhaps the most interesting, since it requires minors from both $\psi_{0,x}$ and $\psi_{1,x}$ (and, unlike the case of $\sigma_2(X)$, the Pfaffians from $\psi_{1,x}$ do not arise from an alternative flattening). Here $\sigma_3(X)$ is defined by the maximal minors of $\psi_{0,x}$ as well as the

8×8 principal Pfaffians of $\psi_{1,x}$. By Example 4.2.10, we see that neither $I_{\kappa_0 \leq 3}$ nor $I_{\kappa_1 \leq 6}$ is sufficient to generate the ideal of $\sigma_3(X)$.

In fact, neither $I_{\kappa_0 \leq 3}$ nor $I_{\kappa_1 \leq 6}$ is sufficient to define $\sigma_3(X)$ even set-theoretically. For $I_{\kappa_0 \leq 3}$, this follows from the fact that a generic element $y \in \Sigma_{\kappa_0 \leq 3}$ has $\kappa_1(y) = 8$. On the other hand, one may check that if

$$x := \sum_{i=1}^{3} u_i \otimes (v_1 \otimes v_{i+1} + v_{i+1} \otimes v_1) \in U^* \otimes S^2 V^*,$$

then $\kappa(x) = (4, 6, 4)$, and hence $[x]$ belongs to $\Sigma_{\kappa_1 \leq 6}$ but not to $\sigma_3(X)$. □

Remark 4.4.5. Let $x \in U^* \otimes S^2 V^*$ and let $r \leq 5$. Theorem 4.4.3 implies that the border rank of x, considered as an element of $U^* \otimes V^* \otimes V^*$, equals the partially symmetric border rank of x. This is because the ideal $I_{\kappa \leq (r,2r,r)}$ is (up to radical) the restriction to $\mathbb{P}(U^* \otimes S^2 V^*)$ of an ideal on $\mathbb{P}(U^* \otimes V^* \otimes V^*)$ which vanishes on the rth secant variety of $\mathbb{P}(U^*) \times \mathbb{P}(V^*) \times \mathbb{P}(V^*)$ (see Proposition 4.1.5 and Definition 4.2.2 above). This can thus be viewed as evidence for a partially symmetric analogue of Comon's Conjecture [15, §5].

Definition 4.4.6. *If we write $x = e_1 \otimes A_1 + e_2 \otimes A_2 + e_3 \otimes A_3$ for e_1, e_2, e_3 a basis of U and the A_i symmetric matrices, then $\det(t_1 A_1 + t_2 A_2 + t_3 A_3)$ is a polynomial in t_1, t_2, and t_3. We define $P \subset \mathbb{A}^{N+1}$ to be the subset of those x such that this polynomial vanishes identically.*

Remark 4.4.7. Note that $\mathbb{A}^{N+1} - P$ is exactly the $\mathrm{GL}(U^*) \times \mathrm{GL}(V^*)$-orbit of the set $\{e_1 \otimes \mathrm{Id} + e_2 \otimes B + e_3 \otimes C \mid B, C \in S^2 V^*\}$.

Lemma 4.4.8. *Let $n = r$. Then $\widehat{\Sigma}_{\kappa \leq (r,2r,r)} - P$ is an irreducible locus of codimension at least $\binom{r}{2}$ on $\mathbb{A}^{N+1} - P$.*

In fact, the codimension is exactly $\binom{r}{2}$, as will be shown in the proof of Lemma 4.4.10.

Proof. Since $n = r$, and $\kappa_0 = \kappa_2$ are always at most n, we have that $\Sigma_{\kappa_1 \leq r} = \Sigma_{\kappa \leq (r,2r,r)}$. For convenience, we denote this scheme Σ, and we seek to show that $\widehat{\Sigma} - P$ is irreducible and of codimension $\binom{r}{2}$.

We let $\mathcal{W} \subseteq \mathbb{A}^{N+1}$ be the set $\{e_1 \otimes \mathrm{Id} + e_2 \otimes B + e_3 \otimes C \mid B, C \in S^2 V^*\}$ as in Remark 4.4.7, and we identify points in $\mathcal{W}$ with pairs of symmetric matrices (B, C). Let $Z \subseteq \mathcal{W}$ be the subscheme defined by the equations $[B, C] = 0$. By [8, Thm. 3.1], Z, known as the variety of commuting symmetric matrices, is an integral subscheme of codimension $\binom{r}{2}$ in $\mathcal{W}$.

We claim that $\widehat{\Sigma} - P$ is irreducible. To see this, we note the following equivalence of matrices under elementary row and column operations:

$$\begin{bmatrix} 0 & \mathrm{Id} & -B \\ -\mathrm{Id} & 0 & C \\ B & -C & 0 \end{bmatrix} \sim \begin{bmatrix} 0 & \mathrm{Id} & 0 \\ -\mathrm{Id} & 0 & 0 \\ 0 & 0 & BC - CB \end{bmatrix}.$$

Therefore, the scheme-theoretic intersection of $\widehat{\Sigma}$ with $\mathcal{W}$ is exactly Z, the variety of commuting symmetric matrices. By Remark 4.4.7 and the fact that κ_1 is $\mathrm{GL}(U^*)\times\mathrm{GL}(V^*)$-invariant, we see that $\widehat{\Sigma} - P$ is exactly the $\mathrm{GL}(U^*) \times \mathrm{GL}(V^*)$ orbit of the irreducible variety Z, and therefore irreducible.

Finally, since $Z = \mathcal{W} \cap \widehat{\Sigma}$, the codimension of $\widehat{\Sigma} - P$ in $\mathbb{A}^{N+1}$ is at least the codimension of Z in $\mathcal{W}$, which is $\binom{r}{2}$. □

The following result is contained in [22, Proof of Cor. 5.6].

Lemma 4.4.9. *If $n = r \leq 5$, then the codimension of P in $\mathbb{A}^{N+1}$ is strictly greater than $\binom{n}{2}$.*

Lemma 4.4.10. *Let $n = r \leq 5$. Then $\sigma_r(X)$ is defined scheme-theoretically by $I_{\kappa_1\leq 2r} = I_{\kappa\leq(r,2r,r)}$. Moreover, the ring $S/I_{\kappa_1\leq 2r}$ is Gorenstein, i.e. $\sigma_r(X)$ is arithmetically Gorenstein.*

Proof. The ideal of the principal $(2r + 2) \times (2r + 2)$-Pfaffians of a generic skew-symmetric matrix is a Gorenstein ideal of codimension $\binom{r}{2}$ [29, Thm. 17]. Our ideal $I_{\kappa_1\leq 2r}$ is a linear specialization of this ideal, and by Lemmas 4.4.8 and 4.4.9, it must be irreducible and have the same codimension. Therefore, the linear specialization is defined by a regular sequence, so $\Sigma_{\kappa_1\leq 2r}$ is also arithmetically Gorenstein and irreducible.

Hence, $\widehat{\Sigma}_{\kappa_1\leq 2r}$ is either reduced or everywhere non-reduced. As in the proof of Lemma 4.4.8, let $\mathcal{W} \subseteq \mathbb{A}^{N+1}$ be the linear space defined by $A_1 = \mathrm{Id}$, and consider the scheme-theoretic intersection $\widehat{\Sigma}_{\kappa_1\leq 2r} \cap \mathcal{W}$. Again, the codimension of $\widehat{\Sigma}_{(r,2r,r)} \cap Z$ in $\mathcal{W}$ is $\binom{r}{2}$, so the generators of the ideal of Z form a regular sequence on the local ring of any point of $\widehat{\Sigma}_{(r,2r,r)}$ contained in $\mathcal{W}$. The intersection is isomorphic to the variety of commuting symmetric matrices from the proof of Lemma 4.4.8, which is reduced. This implies that $\widehat{\Sigma}_{\kappa\leq(r,2r,r)}$ is reduced as well, and thus that $\Sigma_{\kappa\leq(r,2r,r)}$ is reduced. □

When $r > 5$ we have a partial result. Let $J_P \subseteq S^\bullet(U \otimes S^2(V))$ be the ideal defining P.

Corollary 4.4.11. *For any r, the variety $\sigma_r(X)$ is defined by the prime ideal $(I_{\kappa\leq(r,2r,r)} : J_P^\infty)$.*

Note that computing the saturations as in Corollary 4.4.11 can be non-trivial.

Proof of Theorem 4.4.3. Lemma 4.4.10 proves the theorem when $n = r$ and so we just need to extend this result to the cases when $n \neq r$. We let $N' = 3\binom{r}{2} - 1$, so that $\mathbb{P}^{N'}$ is the projective space of partially symmetric $3 \times r \times r$ tensors. We write $X' \subset \mathbb{P}^{N'}$ for the image of $\mathbb{P}^2 \times \mathbb{P}^{r-1}$ embedded by $\mathcal{O}(1, 2)$.

First, suppose that $n < r$. We pick an inclusion of V^* into $\mathbb{A}^r$, and also a projection from $\mathbb{A}^r$ back to V^*. These define an inclusion $\mathbb{P}^N \to \mathbb{P}^{N'}$ and a rational map $\pi\colon \mathbb{P}^{N'} \to \mathbb{P}^N$ respectively. Because the projection is linear, it commutes with taking secant varieties, so $\sigma_r(X) = \pi(\sigma_r(X'))$. Applying Lemma 4.4.10, we get the first equality of

$$\pi(\sigma_r(X')) = \pi(\Sigma_{\kappa_1\leq 2r}) \supset \pi(\Sigma_{\kappa_1\leq 2r} \cap \mathbb{P}^N) = \Sigma_{\kappa_1\leq 2r} \cap \mathbb{P}^N \supset \sigma_r(X) = \pi(\sigma_r(X')).$$

Note that the middle equality follows from the fact that π is the identity on $\mathbb{P}^N$. We conclude that $\sigma_r(X)$ is defined by $I_{\kappa_1 \leq 2r}$, which is the statement of the theorem, since the conditions on κ_0 and κ_2 are trivial when $n < r$.

Second, we want to prove the theorem when $n > r$, for which we use Lemma 4.3.4. We consider the subspace variety $\mathrm{Sub}_{3,r} \subset \mathbb{A}^{N+1}$ and its desingularization $\pi\colon E \to \mathrm{Sub}_{3,r}$. By Proposition 4.3.3, $\mathrm{Sub}_{3,r}$ is the affine cone over $\Sigma_{\kappa_0 \leq r}$, which contains $\widehat{\Sigma}_{\kappa \leq (r,2r,r)}$. We set $Z := \pi^{-1}(\widehat{\Sigma}_{\kappa \leq (r,2r,r)})$. Note that, along any fiber $U^* \otimes S^2\widetilde{V}^*$ of $q\colon E \to \mathrm{Gr}(r, V^*)$, we have that $Z \cap (U^* \otimes S^2\widetilde{V}^*)$ is defined by the $\kappa_1 \leq 2r$ equations applied to $U^* \otimes S^2\widetilde{V}^*$. It follows that $Z \subseteq E$ is defined by the pullback of $I_{\kappa \leq (r,2r,r)}$. Since $\widetilde{V}^*$ is r-dimensional, Lemma 4.4.10 implies that $Z \cap (U \otimes S^2\widetilde{V}^*)$ is the cone over the rth secant variety of $\mathbb{P}(U^*) \times \mathbb{P}(\widetilde{V}^*)$ in $U^* \otimes S^2\widetilde{V}^*$. In particular, Z is reduced. We thus have the inclusions

$$\pi(Z) \subset \widehat{\sigma}_r(X) \subset \widehat{\Sigma}_{\kappa \leq (r,2r,r)} = \pi(Z).$$

The first inclusion is clear, the second is by Proposition 4.2.3, and the equality follows from Lemma 4.3.4. Therefore, these schemes must be equal, which is the desired statement. □

We conclude by observing that Theorem 4.4.3 is false for $r = 7$. (We do not know whether or not it holds for $r = 6$.)

Example 4.4.12. Set $n = \dim V^* = 6$, in which case $\Sigma_{\kappa \leq (7,14,7)} = \Sigma_{\kappa_1 \leq 14}$. Let X be the Segre-Veronese variety of $\mathbb{P}^2 \times \mathbb{P}^5$ embedded by $\mathcal{O}(1,2)$ in $\mathbb{P}^{62}$. We use a simple dimension count to show that the secant $\sigma_7(X)$ is properly contained in $\Sigma_{\kappa_1 \leq 14}$.

The secant variety $\sigma_7(X)$ is not defective [1, Corollary 1.4(ii)], so it has the expected dimension, namely $\dim \sigma_7(X) = 7 \cdot \dim X + 6 = 55$. On the other hand, since $I_{\kappa_1 \leq 14}$ is a Pfaffian ideal, its codimension is at most $\binom{4}{2}$. We thus have

$$\dim \Sigma_{\kappa_1 \leq 14} \geq \dim \mathbb{P}^{62} - \binom{4}{2} = 62 - 6 = 56.$$

Since $56 > 55$, it follows that $\sigma_7(X) \subsetneq \Sigma_{\kappa_1 \leq 14}$.

Note that $\dim V^* = 6$ is the smallest dimension such that the 7th secant variety is properly contained within $\mathbb{P}^N$.

□

Chapter 5

How are SNPs distributed in genes?

In this chapter, we use the equations from Chapter 4 to study the distribution of single nucleotide polymorphisms (SNPs) within the human genes. The occurrence of SNPs is shaped by the functional constraints imposed by the fact that they code for amino acids within proteins. As we will see, the equations of secant varieties will allow us to detect these constraints from the data.

5.1 Testing mixture models

In this section, we introduce the notion of a mixture model for a $m \times n \times n$ contingency table, and develop a statistical test for membership in such a model. For more about mixture models and secant varieties in algebraic statistics, see Chapter 4 of [19]. First, we define a probabilistic model for producing triples of integers from the product $[m] \times [n] \times [n]$.

Definition 5.1.1. *A* mixture model *with r hidden states is described by a probability distribution p_h on the hidden states $\{1, \ldots, r\}$, together with, for each $h \leq r$, a probability distribution u_{hi} on $[m]$ and a probability distribution v_{hj} on $[n]$. The triples are produced by first choosing an integer h with probability p_h and then choosing i, j, and k independently with probabilities u_{hi}, v_{ij}, and v_{ik} respectively. The three integers i, j, and k are output.*

From this description, it follows immediately that the probability of a given triple is

$$P(i,j,k) = \sum_{h=1}^{r} p_h u_{hi} v_{hj} v_{kj}. \tag{5.1}$$

We can incorporate the quantity p_h into the values of u_{hi}, in which case this equation becomes the expression for the (i,j,k)-entry of a semi-symmetric tensor in the rth secant variety of the Segre-Veronese variety. As in Chapter 4, we will represent a probability distribution on $[3] \times [n] \times [n]$ by a triple of symmetric matrices A, B, and C, which contain

the probability distributions when the first variable is 1, 2, or 3 respectively. With this notation, an immediate consequence of Theorem 4.4.3 is:

Proposition 5.1.2. *If (A, B, C) symmetric matrices representing a probability distribution coming from a mixture of r models, then the matrices*

$$\psi_0 = \begin{pmatrix} A & B & C \end{pmatrix} \qquad \text{and} \qquad \psi_1 = \begin{pmatrix} 0 & A & -B \\ -A & 0 & C \\ B & -C & 0 \end{pmatrix} \tag{5.2}$$

have ranks at most r and $2r$ respectively.

Note that the converse to Proposition 5.1.2 does not hold. The relatively minor reason is the Zariski closure which enters into the definition of secant variety. A more significant reason is that even if the entries of a tensor are real, the decomposition implied by Theorem 4.4.3 may require complex entries, whereas the probabilities in 5.1 are all real and non-negative.

Example 5.1.3. Consider the probability distribution given by:

$$A = \frac{1}{12}\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \qquad B = C = \frac{1}{48}\begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix}.$$

The corresponding tensor $e_1 \otimes A + e_2 \otimes B + e_3 \otimes C$ lies in the 4th secant variety. In fact, it can be checked that this tensor can be written as the sum of 4 decomposable tensors with real entries. However, it is not possible to choose a decomposition such that the entries are all real and positive.

To see this, take a decomposition as in (5.1) with $r = 4$. The off-diagonal zeros in A mean that for each h, either $u_{h1} = 0$ or there is at most one index i such that the v_{hi} is non-zero. However, since A has rank 4, u_{h1} must be non-zero for $h = 1, \ldots, 4$, so for fixed h, there can be only one index i such that v_{hi} is non-zero. This implies that B and C are diagonal matrices, which is a contradiction. □

Despite the limitations exhibited by Example 5.1.3, we will build a statistical test based on Proposition 5.1.2. In order to test membership robustly and efficiently, we don't apply the equations from Theorem 4.4.3 directly, but instead use the singular value decomposition of the matrices from which the equations come. This idea has been used in phylogenetics by Eriksson [21].

Recall that the *singular value decomposition* of a real matrix M is an expression $M = U\Sigma V^T$, where U and V are orthogonal matrices and Σ is a (not necessarily square) matrix whose non-zero entries are all on the diagonal. These entries are known as the *singular values* of the matrix, and the number of non-zero singular values is exactly the rank of the matrix M. More importantly, since U and V are required to be orthogonal, small changes to M will have a relatively small effect on the singular values. More precisely,

Theorem 5.1.4 (Theorem 3.3 from [18]). *Let M be a matrix with singular values $\sigma_1 \geq \cdots \geq \sigma_k$. Then the minimum Frobenius distance between M and a rank r matrix is*

$$\sqrt{\sum_{i=r+1}^{k} \sigma_i^2}$$

Our procedure works as follows. We begin with a series of events, each of which is a triple (i, j, k) with $1 \leq i \leq 3$, $1 \leq j \leq n$, and $1 \leq k \leq n$. We form a $3 \times n \times n$ partially symmetric tensor with slices A, B, and C as follows:

$$A_{jk} = \begin{cases} \text{number of occurrences of } (1, j, j) & \text{if } j = k \\ \frac{1}{2} \text{ number of occurrences of } (1, j, k) \text{ or } (1, k, j) & \text{if } j \neq k \end{cases}$$

$$B_{jk} = \begin{cases} \text{number of occurrences of } (2, j, j) & \text{if } j = k \\ \frac{1}{2} \text{ number of occurrences of } (2, j, k) \text{ or } (2, k, j) & \text{if } j \neq k \end{cases}$$

$$C_{jk} = \begin{cases} \text{number of occurrences of } (3, j, j) & \text{if } j = k \\ \frac{1}{2} \text{ number of occurrences of } (3, j, k) \text{ or } (3, k, j) & \text{if } j \neq k \end{cases}$$

We compute the singular value decompositions of the corresponding matrices $\psi_{0,x}$ and $\psi_{1,x}$ in (5.2), and let $\sigma_{0,1} \geq \cdots \geq \sigma_{0,n} \geq 0$ and $\sigma_{1,1} \geq \cdots \geq \sigma_{1,3n} \geq 0$ be their singular values respectively. For each $r < n$, we define:

$$\gamma_r = \sqrt{\sum_{i=r+1}^{n} \sigma_{0,i}^2}$$

and, similarly, for $r < 3n$,

$$\delta_r = \sqrt{\sum_{i=2r+1}^{n} \sigma_{1,i}^2}.$$

For convenience, we take γ_r and γ_r to be 0 if $r \geq n$ or $r \geq 3n/2$ respectively. By Theorem 5.1.4, γ_r and δ_r are the minimum Frobenius distances between ψ_0 and ψ_1 from Proposition 5.1.2 and matrices with rank r and $2r$ respectively. Therefore, γ_r and δ_r measures the extent to which the matrix conditions in Proposition 5.1.2 fail to hold.

In order to obtain a useful interpretation of particular values of γ_r and δ_r, we need to be able to compute p-values for them, which requires knowing the distribution of these statistics, given that they are derived from a mixture model with r hidden states. This distribution can be estimated computationally by sampling contingency tables according to the mixture model. Beginning with an observed contingency table, we must first estimate the

	phase 1				phase 2				phase 3			
	A	G	C	T	A	G	C	T	A	G	C	T
A	944	512	107	15	891	486	66	10	1345	1042	148	20
G		1868	23	92		1474	18	62		2614	45	144
C			1537	455			1494	389			2008	1172
T				719				790				1398

Table 5.1: Occurrence of bases at HapMap loci.

	Phase 0	Phase 1	Phase 2
Homozygous	5068	4649	8165
Heterozygous	1204	1031	2571
Heterozygosity	19%	18%	24%

Table 5.2: Heterozygosity of HapMap loci by phase

mixture model parameters. This estimation can be done using expectation-maximization, similar to the algorithm in Section 2.1 applied to the equations (5.1). However, in this case the inner iteration (2.4) can be replaced by an exact computation of the maximum likelihood parameters in terms of the estimated distribution on the hidden states. Second, these parameters can be used to simulate contingency tables according to the mixture model and their γ_r and δ_r statistics can be computed. The p-value of either statistic can be computed by comparing the observed value to the distribution simulated in this way. If γ_r is the observed statistic, and $\gamma_r^{(1)}, \ldots, \gamma_r^{(N)}$ are the simulated statistics, then the p-value is:

$$p = \frac{\#\{i \mid \gamma_r^{(i)} \geq \gamma_r\} + 1}{N + 1}$$

5.2 The distribution of SNPs in human genes

A single-nucleotide polymorphism is a variation in a genome consisting of the substitution of a single base for another. In this section, we will refer to SNPs between the two haplotypes of a single individual. The HapMap project has collected occurrences of SNPs in many individual humans [16]. We will focus on the first individual in the HapMap data set. Table 5.1 shows the frequencies of all SNPs within a single individual.

In a gene, every triplet of bases, known as a *codon*, determines a single amino acid. Thus, any base within a gene has a phase, defined as its position modulo 3. Table 5.1 counts the occurrence of bases within the first HapMap individual, tabulated by phase.

There are three characteristics of the data in Table 5.1 that we record here. First, the loci are more likely to be homozygous (have the same bases on both strands of the individual) than heterozygous (have different bases). This is made explicit in Table 5.2, where it shows

second base	third base	first base A	G	C	T
A	A	AAA: Lys	GAA: Glu	CAA: Gln	TAA: stop
	G	AAG: Lys	GAG: Glu	CAG: Gln	TAG: stop
	C	AAC: Asn	GAC: Asp	CAC: His	TAC: Tyr
	T	AAT: Asn	GAT: Asp	CAT: His	TAT: Tyr
G	A	AGA: Arg	GGA: Gly	CGA: Arg	TGA: stop
	G	AGG: Arg	GGG: Gly	CGG: Arg	TGG: Trp
	C	AGC: Ser	GGC: Gly	CGC: Arg	TGC: Cys
	T	AGT: Ser	GGT: Gly	CGT: Arg	TGT: Cys
C	A	ACA: Thr	GCA: Ala	CCA: Pro	TCA: Ser
	G	ACG: Thr	GCG: Ala	CCG: Pro	TCG: Ser
	C	ACC: Thr	GCC: Ala	CCC: Pro	TCC: Ser
	T	ACT: Thr	GCT: Ala	CCT: Pro	TCT: Ser
T	A	ATA: Ile	GTA: Val	CTA: Leu	TTA: Leu
	G	ATG: Met	GTG: Val	CTG: Leu	TTG: Leu
	C	ATC: Ile	GTC: Val	CTC: Leu	TTC: Phe
	T	ATT: Ile	GTT: Val	CTT: Leu	TTT: Phe

Table 5.3: Coding of bases by amino acid triplets.

	A	G	C	T
Phase 1	20.1%	34.8%	29.2%	15.9%
Phase 2	20.6%	30.9%	30.5%	18.0%
Phase 3	18.2%	30.1%	32.5%	19.2%

Table 5.4: Haplotype base frequencies by phase in polymorphic sites from the HapMap data.

that the proportion of heterozygous loci is less than a quarter, with the exact number varying by phase.

Our second observation is that the final phase has significantly more polymorphic loci than the others. As shown in Table 5.3, for many codons, changing the last base does not change the corresponding amino acid. Since most changes to the amino acid sequence will decrease fitness, selective pressure reduces the occurrence of mutations in the first two phases.

Finally, we note that the heterozygous loci are more likely to be A-G or C-T than another pair. The bases A and G, known as purines both consist of two cycles and thus are more likely mutate into each other, whereas the pyrimidines, bases C and T consist of a single chemical cycle [45, Sec. 4.1].

We now describe the first of two models for the occurrence of SNPs in the human genome.

	number of hidden states			
	2	3	4	5
p-value for γ_r	≤ 0.0001	≤ 0.0001		
p-value for δ_r	≤ 0.0001	≤ 0.0001	0.142	0.002

Table 5.5: Probabilities of obtaining a γ statistic at least as large as that of the contingency table in Table 5.1.

At each position, there is a particular base which occurred at that location in the most recent common ancestor of the two haplotypes, and we suppose that the distribution of the ancestral base depends on the phase. In our first model, we assume that the observed bases in the present-day haplotypes have evolved from the ancestral base according to a general Markov model, but one which doesn't depend on the phase. Thus, while the occurrence of polymorphic loci depends on the phase, their subsequent evolution does not. This model is a mixture model as in Definition 5.1.1 with four hidden states, which are the 4 ancestral bases. We will call this model the "uniform selection" model, because having identical transition processes from the ancestral base to the present day is consistent with having identical selection pressures on each phase.

Our second model is a generalization which incorporates the possibility of variable selective pressure in different phases. Having a distinct transition matrix for each phase would be too general, and essentially lead to a separate analysis of the data within each phase. Therefore, we introduce only a single parameter for the level of selective pressure at each phase. We let T be the transition matrix under conditions of maximal selective pressure. For each base, we suppose that there is a phase-dependent probability of evolving according to T, and otherwise it follows the opposite extreme, of maximal mutation a fixed distribution, with no dependence on the ancestral base. Although this possibility seems extreme, by allowing it to mix with the transition matrix T, we obtain a range of intermediate levels of selective pressure. We'll call this model "linear selection."

The linear selection model can be re-interpreted as a mixture model with five hidden states. The five hidden states correspond to the four possible ancestral bases plus one for the accelerated mutation rate. Since the accelerated mutation rate doesn't depend on the ancestral base, we can treat it as a fifth hidden state, in addition to the four ancestral bases. Therefore, we can test both this model and the original one using the tests from Section 5.1.

We applied the γ and δ tests from Section 5.1 to the SNP frequencies for the first HapMap individual which were shown in Table 5.1. The p-values for these statistics are shown in Table 5.5, each based on simulations of 10000 contingency tables (so that the minimum possible p-value is $1/10001 \approx 0.0001$). Note that the δ statistic only applies to mixture models with at most 4 hidden states. As expected, the p-values for the models with fewer than 4 submodels are as small as possible, showing the strong effect of the 4 ancestral bases.

Finding 5.2.1. *Although the test does not achieve significance to reject the uniform selection*

2	Asn, Asp, Cys, Glu, Gln, His, Lys, Phe, Tyr
4	Val, Pro, Thr, Ala, Gly
6	Leu, Arg, Ser
others	Ile, Met, Trp, stop

Table 5.6: Categorization of amino acids based on number of codons which code for that amino acid.

	number of hidden states	
amino acid category	4	5
2	0.010	0.328
4	0.087	0.370
6	0.011	0.007

Table 5.7: Resulting p-values for the δ statistic applied to SNP distributions in each of the amino acid categories from Table 5.6.

model, it does for the linear selection model.

What this tells us is that the uniform selection model fits moderately well, but the linear selection model is not a good fit for compensating for the variation in selection levels.

One reason for the poor fit is that the selective pressure depends not just on the phase, but also on the base being coded, as shown in Table 5.3. For example, the triplet GCT codes for alanine, and the same would be true for the final base, T, replaced by any other base. On the other hand, aspartic acid is only coded by the triplets GAT and GAC. Therefore, in the case of alanine, the selective pressure on the final base is less, but in the case of aspartic acid, the lower selective pressure only applies to mutations between T and C.

Therefore, in Table 5.6, we categorize the codons by their coding pattern. The first category consists of the 2-codon amino acids which are each coded by 2 different triplets, always differing by a transition in the last base. The 4-codon amino acids of those which are determined by the only the first two bases of the triplet, independent of the final base. Finally, there are three amino acids which can be coded by 6 possible triplets each. Note that almost all of the codons fit into these three categories, but there are 3 amino acids, together with the stop codons, which do not, and these will be excluded from our subsequent analysis.

We repeated our mixture model analysis of the SNP distribution, but restricted to SNPs in each of the amino acid categories. By definition, a SNP involves different possible haplotypes, so we base these categorizations on the amino acid in the reference sequence. Table 5.7 lists p-values for the δ_r statistics for when r is 4 or 5. For fewer than 4 hidden states, both the γ and δ tests were significant at the limit of what was simulated, so these are not shown. The p-values for the 5 hidden state model are only significant for the 6-codon amino acids. From this, we conclude that the failure of the same model for the SNPs as a whole in Table 5.5 is due to this class of amino acids. In these cases, the amino acid table allows for

extra mutations in the final phase, but there is some selective pressure for transitions over transversions in the second phase. On the other hand, for the 2- and 4-codon amino acids, the linear selection model is sufficient.

We have shown that, as expected, the selective pressures on SNPs depend on the phase within the codon. However, this dependence is not as simple as having a higher level of polymorphism in the final phase. Instead, the existence of 6-codon amino acids means that the selective pressures vary quantitatively as well as quantitatively. On the other hand, for amino acids with simpler coding patterns, the SNPs can be effectively modeled by a simple phase-dependent increase in the mutation rate.

We conclude that the study of tensors can be brought to bear on the analysis of genomic data. We hope that further developments will lead to more applications of tensors and algebraic statistics to computational biology.

Bibliography

[1] Hirotachi Abo and Maria Chiara Brambilla. Secant varieties of Segre-Veronese varieties $\mathbb{P}^m \times \mathbb{P}^n$ embedded by $\mathcal{O}(1,2)$. *Experiment. Math.*, 18(3):369–384, 2009.

[2] Dan Bates and Frank Sottile. Khovanskii-Rolle continuation for real solutions. arXiv:0908.4579, 2009.

[3] Shuchao Bi. Tensor rank, simultaneous diagonalization, and some related matrix varieties. preprint, 2010.

[4] K. Birnbaum, J. W. Jung, J. Y. Wang, G. M. Lambert, J. A. Hirst, D. W. Galbraith, and P. N. Benfey. Cell type-specific expression profiling in plants via cell sorting of protoplasts from fluorescent reporter lines. *Nature Methods*, 2(8):615, 2005.

[5] K. Birnbaum, D. E. Shasha, J. Y. Wang, J. W. Jung, G. M. Lambert, D. W. Galbraith, and P. N. Benfey. A gene expression map of the *Arabidopsis* root. *Science's STKE*, 302(5652):1956, 2003.

[6] M. Bonke, S. Thitamadee, A. P. Mähönen, M. T. Wauser, and Y. Helariutta. APL regulates vascular tissue identity in *Arabidopsis*. *Nature(London)*, 426(6963):181–186, 2003.

[7] Siobhan M. Brady, David A. Orlando, J. Y. Lee, J. Y. Wang, J. Koch, J. R. Dinneny, D. Mace, U. Ohler, and Philip N. Benfey. A high-resolution root spatiotemporal map reveals dominant expression patterns. *Science*, 318(5851):801–806, 2007.

[8] J. P. Brennan, M. V. Pinto, and W. V. Vasconcelos. The Jacobian module of a Lie algebra. *Trans. Amer. Math. Soc.*, 321(1):183–196, 1990.

[9] W. Busch and J. U. Lohmann. Profiling a plant: expression analysis in *Arabidopsis*. *Current Opinion in Plant Biology*, 10(2):136–141, 2007.

[10] Dustin Cartwright. An algorithm for finding postive solutions to polynomial equations. preprint, 2010.

[11] Dustin Cartwright, Daniel Erman, and Luke Oeding. Secant varieties of $\mathbb{P}^2 \times \mathbb{P}^n$ embedded by $\mathcal{O}(1,2)$. preprint, 2010.

[12] Dustin A. Cartwright, Siobhan M. Brady, David A. Orlando, Bernd Sturmfels, and Philip N. Benfey. Reconstruction spatiotemporal gene expression data from partial observations. *Bioinformatics*, 25(19):2581–2587, 2009.

[13] Dustin A. Cartwright, Daniel Erman, Mauricio Velasco, and Bianca Viray. Hilbert schemes of 8 points. *Algebra & Number Theory*, 3(7):763–795, 2009.

[14] B. Chaudhuri, F. Hörmann, S. Lalonde, S. M. Brady, D. A. Orlando, P. N. Benfey, and W. B. Frommer. Protonophore- and pH-insensitive glucose and sucrose accumulation detected by FRET nanosensors in *Arabidopsis* root tips. *Plant J.*, 56(6):948–962, 2008.

[15] Pierre Comon, Gene Golub, Lek-Heng Lim, and Bernard Mourrain. Symmetric tensors and symmetric tensor rank. *SIAM J. Matrix Anal. Appl.*, 30(3):1254–1279, 2008.

[16] The International HapMap Consortium. A second generation human haplotype map of over 3.1 million snps. *Nature*, 449:851–861, 2007.

[17] J. N. Darroch and D. Ratcliff. Generalized iterative scaling for log-linear models. *Annals of Math. Stat.*, 43(5):1470–1480, 1972.

[18] James W. Demmel. *Applied Numerical Linear Algebra*. Society for Industrial and Applied Mathematics (SIAM), 1997.

[19] Mathias Drton, Bernd Sturmfels, and Seth Sullivant. *Lectures on Algebraic Statistics*, volume 39 of *Oberwolfach Seminars*. Birkhäuser, 2009.

[20] David Eisenbud. *Commutative Algebra with a View Toward Algebraic Geometry*, volume 150 of *Graduate Texts in Mathematics*. Springer, 2004.

[21] Nicholas Eriksson. Tree construction using singular value decomposition. In *Algebraic Statistics for Computational Biology*, pages 347–358. Cambridge University Press, 2005.

[22] Daniel Erman and Mauricio Velasco. A syzygetic approach to the smoothability of zero-dimensional schemes. *Adv. in Math.*, 224:1143–1166, 2010.

[23] William Fulton and Joe Harris. *Representation theory*, volume 129 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1991. A first course, Readings in Mathematics.

[24] Luis David Garcia, Michael Stillman, and Bernd Sturmfels. Algebraic geometry of Bayesian networks. *J. Symbolic Comput.*, 39(3-4):331–355, 2005.

[25] Debashis Ghosh. Mixture models for assessing differential expression in complex tissues using microarray data. *Bioinformatics*, 20:1663–1669, 2004.

[26] Daniel R. Grayson and Michael E. Stillman. Macaulay 2, a software system for research in algebraic geometry. Available at http://www.math.uiuc.edu/Macaulay2/.

[27] Anjali S. Iyer-Pascuzzi and Philip N. Benfey. Transcriptional networks in root cell fate specification. *Biochim. Biophys. Acta.*, 1789(4):315–25, April 2009.

[28] Yuling Jiao, S Lori Tausta, Neeru Gandotra, Ning Sun, Tie Liu, Nicole K Clay, Teresa Ceserani, Meiqin Chen, Ligeng Ma, Matthew Holford, Hui yong Zhang, Hongyu Zhao, Xing-Wang Deng, and Timothy Nelson. A transcriptome atlas of rice cell types uncovers cellular, functional, and developmental hierarchies. *Nature Genetics*, 41:258–263, 2009.

[29] H. Kleppe and D. Laksov. The algebraic structure and deformation of Pfaffian schemes. *J. Algebra*, 64(1):167–189, 1980.

[30] J. M. Landsberg and Laurent Manivel. On the ideals of secant varieties of Segre varieties. *Found. Comput. Math.*, 4(4):397–422, 2004.

[31] J. M. Landsberg and Laurent Manivel. Generalizations of Strassen's equations for secant varieties of Segre varieties. *Comm. Algebra*, 36(2):405–422, 2008.

[32] J. M. Landsberg and G. Ottaviani. Equations for secant varieties to Veronese varieties. arXiv:1006.0180, 2010.

[33] J. M. Landsberg and Zach Teitler. On the ranks and border ranks of symmetric tensors. *Found. Comput. Math.*, 10(3):339–366, 2010.

[34] J. M. Landsberg and Jerzy Weyman. On the ideals and singularities of secant varieties of Segre varieties. *Bull. Lond. Math. Soc.*, 39(4):685–697, 2007.

[35] J. M. Landsberg and Jerzy Weyman. On secant varieties of compact Hermitian symmetric spaces. *J. Pure Appl. Algebra*, 213(11):2075–2086, 2009.

[36] Jean B. Lasserre, Monique Laurent, and Philipp Rostalski. Semidefinite characterization and computation of real radical ideals. *Foundations of Computational Mathematics*, 8(5):607–647, 2008.

[37] Jean B. Lasserre, Monique Laurent, and Philipp Rostalski. A prolongation-projection algorithm for computing the finite real variety of an ideal. *Theoretical Computer Science*, 410(27–29):2685–2700, 2009.

[38] Daniel D. Lee and H. Sebastian Seung. Algorithms for non-negative matrix factorization. *Adv. Neural Info. Proc. Syst.*, 13:556–562, 2001.

[39] J.Y. Lee, J. Colinas, J.Y. Wang, D. Mace, U. Ohler, and P.N. Benfey. Transcriptional and posttranscriptional regulation of transcription factor expression in *Arabidopsis* roots. *Proceedings of the National Academy of Sciences*, 103(15):6055–6060, 2006.

[40] M. Levine and E.H. Davidson. Gene regulatory networks for development. *Proc Natl Acad Sci USA*, 102(14):4936–42, 2005.

[41] I. G. Macdonald. *Symmetric functions and Hall polynomials*. Oxford Mathematical Monographs. The Clarendon Press Oxford University Press, New York, second edition, 1995. With contributions by A. Zelevinsky, Oxford Science Publications.

[42] A.P. Mahonen, M. Bonke, L. Kauppinen, M. Riikonen, P.N. Benfey, and Y. Helariutta. A novel two-component hybrid molecule regulates vascular morphogenesis of the *Arabidopsis* root. *Genes & Development*, 14(23):2938–2943, 2000.

[43] Giorgio Ottaviani. Symplectic bundles on the plane, secant varieties and Lüroth quartics revisited. In *Vector bundles and low codimensional subvarieties: state of the art and recent developments*, volume 21 of *Quad. Mat.*, pages 315–352. Dept. Math., Seconda Univ. Napoli, Caserta, 2007.

[44] Giorgio Ottaviani. An invariant regarding Waring's problem for cubic polynomials. *Nagoya Math. J.*, 193:95–110, 2009.

[45] Lior Pachter and Bernd Sturmfels. *Algebraic Statistics for Computational Biology*. Cambridge University Press, 2005.

[46] A. Slapak and A. Yeredor. Near-optimal weighting in characteristic-function based ICA. In *18th European Signal Processing Conference (EUSIPCO-2010)*, pages 890–894, August 23-27 2010.

[47] Volker Strassen. Rank and optimal computation of generic tensors. *Linear Algebra Appl.*, 52/53:645–685, 1983.

[48] Emil Toeplitz. Über ein Flächennetz zweiter Ordnung. *Math. Ann.*, 11:434–463, 1877.

[49] M. A. A. van Leeuwen, A. M. Coehn, and B. Lisser. *LiE, A Package for Lie Group Computations*. Computer Algebra Nederland, 1992.

[50] D. Venet, F. Pecasse, C. Maenhaut, and H. Bersini. Separation of samples into their constituents using gene expression data. *Bioinformatics*, 17:S279–S287, 2001.

[51] Jan Verschelde. PHCPACK: A general-purpose solver for polynomial systems by homotopy continuation.

[52] Jerzy Weyman. *Cohomology of Vector Bundles and Syzygies*, volume 149 of *Cambridge Tracts in Mathematics*. Cambridge University Press, Cambridge, 2003.