
#### Abstract

SCHOLTEN, GEORGY HENDRIK. Combinatorial and Real Algebraic Structures in Statistics and Optimization. (Under the direction of Cynthia Vinzant.)

In recent years, mathematicians working in nonlinear algebra have actively explored the use of algebraic geometry, combinatorics and symbolic computations to answer questions in a diverse field of applications, with an emphasis on statistics and optimization. This dissertation is comprised of three projects, all of which relied on techniques from real algebraic geometry and combinatorics and have connections to applications coming from statistics, optimization and population genetics.

The first project is about image of a linear space under partial coordinate inversion. The structure of an affine variety constructed in that way is governed by an underlying hyperplane arrangement. We show that circuit polynomials form a universal Gröbner basis for the ideal of polynomials vanishing on this variety. To prove this, we rely on a degeneration to the Stanley-Reisner ideal of a simplicial complex solely determined by the matroid of the linear space and the set of inverted coordinates. If the linear space is real, then the semi-inverted linear space is also an example of a hyperbolic variety, meaning that all of its intersection points with a large family of linear spaces are real.

In the second project, we study the problem of maximum likelihood estimation restricted to logconcave probability density functions. We explore in what sense exact solutions to this problem, which take the form of the exponential of a piecewise linear function, are possible. First, we show that the heights given by the maximum likelihood estimate are generically transcendental. For a cell in one dimension, the maximum likelihood estimator is expressed in closed form using the generalized W-Lambert function. Even more, we show that finding the log-concave maximum likelihood estimate is equivalent to solving a collection of polynomial-exponential systems of a special form. Even in the case of two equations, very little is known about solutions to these systems. As an alternative, we use Smale's alpha-theory to refine approximate numerical solutions and to certify solutions to log-concave density estimation.

In the third project, we study the univariate truncated moment problem of piecewise-constant density functions on the interval $[0,1]$ and its consequences for an inference problem in population genetics. We show that, up to closure, any collection of $n$ moments is achieved by a step function with at most $n-1$ breakpoints and that this bound is tight. We use this to show that any point in the $n$-th coalescence manifold in population genetics can be attained by a piecewise constant population history with at most $n-2$ changes. Both the moment cones and the coalescence manifold are projected spectrahedra and we describe the problem of finding a nearest point on them as a semidefinite program.


© Copyright 2021 by Georgy Hendrik Scholten
All Rights Reserved
by
Georgy Hendrik Scholten

A dissertation submitted to the Graduate Faculty of North Carolina State University in partial fulfillment of the requirements for the Degree of Doctor of Philosophy

Mathematics

Raleigh, North Carolina
2021

## APPROVED BY:

## DEDICATION

To my parents, Galina and Hendrik, and my brother, Anton.

## BIOGRAPHY

Georgy Scholten was born in Lille, France. He grew up in the center region of France, in Blois, where he attended the Lycée Notre Dame des Aydes. He moved with his family to Raleigh, North Carolina, in 2013 and started attending NC State University. He graduated in May 2016 with a bachelor in mathematics and started his graduate studies in August of the same year. While a student, he enjoyed competing with the swim club at NC State, playing the drums, and exploring North Carolina and its surroundings. Georgy will continue his mathematical career as a postdoctoral researcher at the Sorbonne Université in Paris, France.

## ACKNOWLEDGEMENTS

I would like to acknowledge the importance of all the connections created over the course of my academic journey. I have had the privilege to get to know many fascinating people, work on deeply engaging problems and create amazing friendships; those all have had a huge influence on who I am today. First and foremost, I would like to thank my advisor, Cynthia Vinzant, who went above and beyond in time and effort to teach me how to conduct mathematical research. Her positive attitude, work ethics and encouragements have tremendously helped me grow my mathematical interests throughout my graduate studies. I would like to thank Hoon Hong, the first professor I took classes with as undergraduate student at NC State. I had the chance to take his computer algebra class early on in my studies, and that made me realize I wanted to go further into mathematics. I would like to thank Kaie Kubjas for putting together a great team of researchers and giving me the opportunity to contribute to a very interesting research project.

I would like to thank my committee members, Hoon Hong, David Papp, Seth Sullivant and Jonathan Williams, for good questions and insightful feedback on my dissertation. There are numerous professors I have taken classes and worked with at NC State, here are a few that I would like to thank in particular: Alina Duca, Molly Fenn, Min Kang, Irina Kogan, Ricky Liu, Nathan Reading, Radmila Sazdanovic and Agnes Szanto.

I would like to thank my co-authors, working with you was a great experience: Yonggu Kim, J. Rafael Sendra, Alexandros Grosdos, Alexander Heaton, Olga Kuznetsova, Miruna-Stefana Sorea and Zvi Rosen. I have met an amazing community of graduate students, postdocs and professors, both at NC State, conferences and summer schools I attended. I would like to thank them for everything they taught me and the welcoming environment they created. I hope I will get the chance to collaborate with many of you in the future.

Many of my fellow graduate students are now also some of my dearest friends. I would like to thank Faye Pasley-Simons and Michael Ruddy for supporting me and helping me over the years, providing feedback and advice on various writings, presentations, statements and job applications. It would be too long to thank each of the following people individually, but I want to mention their names here and would like to thank them in person next time we meet: Alexeï Bondarev, Cashous and Megan Bortner, Taylor Botham, Jane Coons, William Crumpler, Robert DeYeso, Steven Gilmore, Nicholas Henderson, Benjamin Hollering, Bianca and Tyler Knapp, Tyler Maltba, Amanda and Michael Merritt, Harley Brandstadter and Robin Morillo, Katherine Pearce and Danny Patrick, Cody Trivette, Jillian Tucker, David White, and Zev Woodstock.

The work appearing in this thesis was partially supported by the National Science Foundation under DMS grants \#1620014 and \#1943363.

## TABLE OF CONTENTS

LIST OF TABLES ..... vii
LIST OF FIGURES ..... viii
Chapter 1 Introduction ..... 1
1.1 Background from Algebraic Geometry ..... 1
1.1.1 Projective Space ..... 3
1.1.2 Rational Maps ..... 5
1.1.3 Semialgebraic Sets ..... 6
1.1.4 Real Algebraic Geometry and Nonnegative Polynomials ..... 6
1.2 Convexity ..... 7
1.2.1 Convex Sets ..... 7
1.2.2 Convex Cones ..... 9
1.2.3 Polyhedra and Polytopes ..... 10
1.2.4 Spectrahedra ..... 10
1.2.5 Applications to Optimization ..... 11
1.3 Discrete Geometry and Algebraic Combinatorics ..... 12
1.3.1 Simplicial Complexes ..... 12
1.3.2 Initial Ideals ..... 14
1.3.3 Connections with Algebraic Geometry ..... 15
1.4 Contents of the Following Chapters ..... 15
Chapter 2 Semi-Inverted Linear Spaces ..... 16
2.1 Background ..... 16
2.1.1 Reciprocal Linear Spaces ..... 16
2.1.2 Matroids ..... 17
2.1.3 Linear Matroids ..... 19
2.1.4 Stanley Reisner Ideals ..... 19
2.2 Main Results ..... 20
2.3 A Semi-Broken Circuit Complex ..... 21
2.4 Proof of the Main Theorem ..... 25
2.5 Supports ..... 29
2.6 Real Points and Hyperplane Arrangements ..... 31
Chapter 3 Log-Concave Maximum Likelihood Estimation ..... 36
3.1 Motivations and Background ..... 36
3.2 Geometry of log-concave maximum likelihood estimation ..... 38
3.3 Transcendentality and closed-form solutions ..... 43
3.3.1 Score equation matrix invertibility and transcendentality ..... 43
3.3.2 An Alternative Proof of Invertibility for the Simplex Case ..... 47
3.3.3 One cell in one dimension ..... 50
3.3.4 Two cells in one dimension ..... 53
3.3.5 Optimization for the One Dimensional Case ..... 54
3.4 Certifying solutions with Smale's $\alpha$-theory ..... 56
3.4.1 Smale's $\alpha$-theory ..... 57
3.4.2 Polynomial-exponential systems ..... 58
3.4.3 A procedure for $\alpha$-certifying ..... 61
Chapter 4 Conic Hulls of Moment Curves and Connections to Coalescence Manifolds ..... 66
4.1 Background and Motivation ..... 66
4.2 Moments of step functions ..... 67
4.3 Increasing and decreasing step functions ..... 73
4.4 Connection with coalescence manifold ..... 76
4.5 The Coalescence Manifold in 3-Space ..... 80
4.6 Connections with semidefinite programming ..... 85
4.7 Discussion and open questions ..... 88
BIBLIOGRAPHY ..... 91

## LIST OF TABLES

Table 3.1 Number of binary digits needed to certify $m+1$ points with weights coming from an asymmetric binomial distribution.

## LIST OF FIGURES

Figure 1.1 The projection of a cylinder. ..... 8
Figure 1.2 Affine slice of the convex hull of $\gamma:[0,1] \rightarrow \mathbb{R}^{4}$, for $\gamma(t)=\left(1, t, t^{2}, t^{3}\right)$ ..... 12
Figure 1.3 A Tent Function appearing in [79]. ..... 14
Figure 2.1 Intersections of $\mathcal{L}^{\perp}+u$ with $\operatorname{inv}_{I}^{-}(\mathcal{L})$ from Example 2.6.7. ..... 34
Figure 3.1 The optimal tent function for the sample of 14 points in Example 3.1.1 ..... 38
Figure 3.2 The Samworth body for $X=(2,5,7)$. ..... 40
Figure 3.3 The red tent function corresponds to a vector $y$ in the red region of the Samworth body. The solid green tent function corresponds to a vector $y$ on the curve separating red and green regions of the Samworth body. The dotted green function is not convex. Its height vector $y$ belongs to the green region of the Samworth body and both green sets of heights give the same tent function. ..... 41
Figure 3.4 Maximizing $S_{\Delta}$ over $y$ restricted to $Y_{\Delta}$. ..... 43
Figure 3.5 Generalized Lambert function $W\left(\rho+1 ;-\rho^{-1}-1 ;-\rho\right)$. ..... 52
Figure 3.6 The tent function and log-concave MLE on a sample of 100 points from a Beta $(2,2)$ distribution. ..... 55
Figure 3.7 The tent function and log-concave MLE on a sample of 100 points from a Laplace ( 0,4 ) distribution. ..... 55
Figure 3.8 The height functions for (a) Example 3.4.12; (b) Example 3.4.13; (c) Example 3.4.15 ..... 63
Figure 3.9 Subdivisions in Example 3.4.16. (a) Subdivision $\Delta_{1}$ (b) Subdivision $\Delta_{2}$ (c) The subdivision induced by $y_{\Delta_{1}}^{*}$. ..... 65
Figure 4.1 The cone $M(A)$ along with the curve segments $v_{A}([0,1])$ and $v_{B}([0,1])$ for $A=\{1,2\}$ and $B=\{0,1\}$ ..... 69
Figure 4.2 The parameter space of $M_{1}(A)$ and $M_{1}(A), M_{2}(A)$ for $A=\{0,2,5\}$. ..... 73
Figure 4.3 The functions p, $\tilde{\mathrm{p}}^{*}$, and $R_{\mathrm{p}}$ from Example 4.4.8. ..... 79
Figure 4.4 The sets $M_{1}(A), M_{2}(A), M_{3}(A)$ in $\left\{m_{0}=1\right\}$ for $A=\{0,2,5,9\}$. ..... 80
Figure 4.5 The $S$ and $W$ simplex. ..... 81
Figure 4.6 Schlegel diagram for the boundary of $P$, together with the image of each facet under the moment map. ..... 82
Figure 4.7 A disconnected fiber of $\mu_{A}$ for $A=\{0,2,5,9\}$. ..... 85Figure 4.8 The central image depicts $M_{2}(A)$ in yellow. The orange region is $M^{\uparrow}(A)$ andthe green region $M^{\downarrow}(A)$; their union is $M_{1}(A)$. The triangle above each pointdepicts the fiber as a subset of the $\left(s_{1}, s_{2}\right)$-simplex.89

## Chapter 1

## Introduction

The purpose of this chapter is to introduce the necessary background notations and references on algebraic geometry, polyhedral geometry, convexity theory and algebraic combinatorics required for the understanding of further chapters.

### 1.1 Background from Algebraic Geometry

In this section, we define a few fundamental concepts in algebraic geometry required to work with affine and projective varieties, that is, the vanishing sets of systems of polynomial equations. For an in-depth introduction to the concepts introduced in this section, we refer the reader to [24, 40, 46, 47]. Let us work over a graded ring $R$ with a direct sum decomposition

$$
R=\bigoplus_{i=0}^{\infty} R_{i},
$$

endowed with a multiplicative structure $R_{i} R_{j} \subset R_{i+j}$ for all $i, j \in \mathbb{Z}^{+}$. Furthermore, we require $R$ to be Noetherian, that is, any ascending chain of ideals in $R$ eventually stabilizes. In this thesis, we will work over the polynomial ring $\mathbb{K}\left[x_{1}, \ldots, x_{n}\right]$, where $\mathbb{K}$ will be specified to be the field of complex numbers $\mathbb{C}$ or real numbers $\mathbb{R}$. Later in this section, we will introduce notions from projective geometry, in which case we work with projective coordinates and homogeneous polynomials over the ring $\mathbb{K}\left[x_{0}, x_{1}, \ldots, x_{n}\right]$.

Definition 1.1.1. Let $R=\mathbb{K}\left[x_{1}, \ldots, x_{n}\right]$ and $S$ be a subset of $\mathbb{K}^{n}$. We define $\mathcal{I}(S) \subset R$ to be the polynomial ideal vanishing on $S$ :

$$
I(S)=\{f \in R: f(s)=0 \text { for all } s \in S\} .
$$

Definition 1.1.2. The algebraic variety of an ideal $I \subset \mathbb{C}\left[x_{1}, \ldots, x_{n}\right]$ is a collection of points where the ideal vanishes

$$
\mathcal{V}(I)=\left\{x \in \mathbb{C}^{n}: f(x)=0 \text { for all } f \in I\right\} .
$$

The real algebraic variety $\mathcal{V}_{\mathbb{R}}(I)$ is the restriction to the real part of $\mathcal{V}(I)$

$$
\mathcal{V}_{\mathbb{R}}(I)=\left\{x \in \mathbb{R}^{n}: f(x)=0 \text { for all } f \in I\right\}
$$

We say that an ideal $I \subseteq R$ is proper if it does not equal the whole ring $R$. An ideal $I$ of a ring $R$ is called prime if $f g \in I$ implies $f \in I$ or $g \in I$ and is primary if $f g \in I$ implies $f \in I$ or $g^{m} \in I$ for some $m \in \mathbb{N}$.

Definition 1.1.3. The Zariski topology on $\mathbb{K}^{n}$ is a topological space where closed sets are defined to be algebraic varieties in $\mathbb{K}^{n}$ and open sets are their complements. The Zariski-closure of a set $S \subset \mathbb{K}^{n}$, denoted $\operatorname{cl}(S)$, is the smallest algebraic variety that contains $S$.

Definition 1.1.4 ([40]). The Krull dimension $\operatorname{dim}(R)$ of a ring $R$ is the supremum of the lengths of chains of distinct prime ideals in $R$, in other words, it is the largest $d$ such that the following chain of primes ideals exists:

$$
P_{0} \subsetneq P_{1} \subsetneq \ldots \subsetneq P_{d} \subsetneq R .
$$

The dimension of the ideal $I \subset R$ and of the variety $\mathcal{V}(I)$ is equal to the Krull dimension of the quotient ring $R / I$. Equivalently, $\operatorname{dim}(\mathcal{V}(I))$ is the largest $d$ such that a chain of irreducible subvarieties of $\mathcal{V}(I)$ exists:

$$
V_{0} \subset V_{1} \subset \ldots \subset V_{d} \subset \mathcal{V}(I)
$$

Example 1.1.5. The polynomial ring $\mathbb{R}\left[x_{1}, \ldots, x_{n}\right]$ has dimension $n$ :

$$
\langle 0\rangle \subset\left\langle x_{1}\right\rangle \subset \ldots \subset\left\langle x_{1}, \ldots, x_{n-1}\right\rangle \subset\left\langle x_{1}, \ldots, x_{n}\right\rangle
$$

Furthermore, we observe that there is never a prime ideal properly contained between $\left\langle x_{1}, \ldots, x_{i-1}\right\rangle$ and $\left\langle x_{1}, \ldots, x_{i}\right\rangle$. Meaning there is no additional ideal that could be inserted into this sequence, hence the dimension of the ring is equal to $n$.

Definition 1.1.6. The radical of $I$ is denoted $\sqrt{I}$ and defined as

$$
\sqrt{I}=\left\{f \in R: f^{m} \in I \text { for some } m \in \mathbb{N}\right\}
$$

Proposition 1.1.7. Any ideal $I \subset R$ admits an irredundant primary decomposition into a finite intersection of primary ideals $Q_{i}$ minimal with respect to inclusion:

$$
I=Q_{1} \cap Q_{2} \cap \ldots \cap Q_{r} \quad \text { such that } \quad Q_{j} \nsupseteq Q_{1} \cap \ldots \cap Q_{j-1} \cap Q_{j+1} \cap \ldots \cap Q_{r}
$$

The prime ideals $\sqrt{Q_{i}}$ are called the associated minimal primes of $I$. The radical ideal $\sqrt{I}$ is the intersection of the associated minimal primes of I. The ideal I is said to be equidimensional if all its associated minimal primes are of equal dimension.

Proposition 1.1.8. Suppose $I \subseteq J \subset R$ are both equidimensional of dimension $d$ and $I$ is radical. If $Q_{i}$
is a primary ideal appearing in a minimal primary decomposition of $J$ of the form $Q_{1} \cap \ldots \cap Q_{r}$, with $\operatorname{dim}\left(Q_{i}\right)=d$, then $Q_{i}$ is prime.

Definition 1.1.9. An ideal $m \subsetneq R$ is said to be maximal if no ideal $I$ distinct from $m$ can be properly contained in between $m$ and the whole ring $R$

$$
m \subsetneq I \subsetneq R .
$$

Definition 1.1.10 ( $\S 9.3$ of [24]). Let $I$ be an ideal in $R=\mathbb{K}\left[x_{1}, \ldots, x_{n}\right]$. The affine Hilbert function $\mathrm{HF}_{J}: \mathbb{Z} \rightarrow \mathbb{Z}$ is the function on nonnegative integers $s$ defined by:

$$
\mathrm{HF}_{J}(s)=\operatorname{dim} R_{\leq s} / I_{\leq s}=\operatorname{dim} R_{\leq s}-\operatorname{dim} I_{\leq s}
$$

where $R_{\leq s}$ denotes the set of polynomials of total degree $\leq s$ in $R$. Similarly, we let

$$
I_{\leq s}=I \cap R_{\leq s}
$$

denote the set of polynomials in $I$ of total degree $\leq s$.
The affine Hilbert polynomial of $I$ is the polynomial $h_{I}(s)$ that agrees with $\operatorname{HF}_{I}(s)$ for sufficiently large $s \in \mathbb{N}$ and it takes the form

$$
h_{I}(s)=\sum_{i=0}^{d} b_{i}\binom{s}{d-i} \quad \text { where } b_{i} \in \mathbb{Z} \text { and } d \in \mathbb{Z}_{>0} .
$$

Given a nonnegative integer $n$, the binomial notation used above denotes a polynomial in $s$ of the form

$$
\binom{s}{n}=\left\{\begin{array}{ll}
s(s-1) \ldots(s-n+1) / n! & \text { if } n>0 \\
1 & \text { if } n=0
\end{array} .\right.
$$

Theorem 1.1.11 (Hilbert's Nullstellensatz). Let $I \subseteq \mathbb{C}\left[x_{1}, \ldots, x_{n}\right]$ be an ideal, a polynomial $p$ vanishes on $\mathcal{V}(I)$ if and only if there exists an $n \in \mathbb{N}$ such that $p^{n} \in I$. Put more concisely,

$$
\mathcal{I}(\mathcal{V}(I))=\sqrt{I} .
$$

### 1.1. 1 Projective Space

Projective geometry was first thought of in the arts, most notably when painters started to study perspective. Projective space was introduced in more mathematical terms along the 17 -th century, in order to incorporate points at infinity into Euclidean space in a consistent way. It is about 200 years later (1830) that Julius Plücker introduced coordinate systems for projective spaces. Projective space is a fundamental concept in algebraic geometry and is found throughout applications, such as computer vision and polynomial optimization. The complex projective space $\mathbb{P}^{n}$ consists of all lines through the origin in
$\mathbb{C}^{n+1}$ and the real projective space $\mathbb{P R}^{n}$, of all lines through the origin in $\mathbb{R}^{n+1}$. That is,

$$
\begin{aligned}
\mathbb{P}^{n} & =\left(\mathbb{C}^{n+1}-\{0\}\right) / \sim \\
\mathbb{P}^{n} & =\left(\mathbb{R}^{n+1}-\{0\}\right) / \sim
\end{aligned}
$$

where the equivalence relation $\sim$ denotes equality between $\left(x_{0}, \ldots, x_{n}\right)$ and $\lambda\left(x_{0}, \ldots, x_{n}\right)$, for all $\lambda \neq 0$. A set of points $\left\{\left[w_{1}\right], \ldots,\left[w_{m}\right]\right\} \subset \mathbb{P}^{n}$ is linearly dependent if $\left\{w_{1}, \ldots, w_{m}\right\} \subset \mathbb{C}^{n+1}$ is linearly dependent.

Definition 1.1.12. A homogeneous ideal $J \subseteq R=\oplus R_{d}$ is an ideal generated by a set of elements $\left\{f_{1}, \ldots, f_{r}\right\}$, where each $f_{j}$ is contained in a single $R_{d}$. In other words, if $R=\mathbb{K}\left[x_{0}, \ldots, x_{n}\right]$, then the polynomial $f_{j} \in R_{d}$ is homogeneous of total degree $d$ in the variables $x_{0}, x_{1}, \ldots x_{n}$.

Definition 1.1.13. A projective variety $X=\mathcal{V}(J)$ is the vanishing locus of a homogeneous ideal $J$. In a more abstract reformulation, $X$ is a subset of $\mathbb{P}^{n}$ whose intersection with every set $U_{i}=\left\{\left[x_{0}, \ldots, x_{n}\right]\right.$ : $\left.x_{i} \neq 0\right\}$ is an affine variety, for all $i \in[0,1, \ldots, n]$.

Definition 1.1.14. Given a projective variety $X \subset \mathbb{P}^{n}$, let $\mathcal{J}(X)$ denote the ideal of homogeneous polynomials vanishing on $X$. The coordinate ring of a projective variety $X$ is the quotient ring $\mathbb{C}\left[x_{0}, \ldots, x_{n}\right] / \mathcal{J}(X)$.

Definition 1.1.15 (§9.3 of [24]). Given a homogeneous ideal $J \subseteq R=\mathbb{K}\left[x_{0}, \ldots, x_{n}\right]$ and $s$ a nonnegative integer, we let

$$
J_{s}=J \cap R_{s}
$$

denote the set of homogeneous polynomials in $J$ of total degree $s$. The Hilbert function of $J$ is given by

$$
\mathrm{HF}_{R / J}(s)=\operatorname{dim} R_{S} / I_{S}
$$

For $s$ large enough, the Hilbert function $\operatorname{HF}_{R / J}(s)$ agrees with the Hilbert polynomial

$$
h_{J}(s)=\sum_{i=0}^{d} b_{i}\binom{s}{d-i}
$$

where $b_{i} \in \mathbb{Z}$ for all $i \in[d]$ and $b_{0} \geq 0$.
The dimension of a projective variety $X$ is equal to the degree of the Hilbert polynomial of $\mathcal{J}(X)$ and the degree of $X$ is equal to $b_{0}$, the leading term coefficient of Hilbert polynomial of $\mathcal{J}(X)$.

Remark 1.1.16 (Definition 18.1 [46]). Over an algebraically closed field, the degree of a variety $X \subset \mathbb{C}^{n}$ is equal to the number of intersection points of $X$ with a linear space in general position of dimension $\operatorname{codim}(X)=n-\operatorname{dim}(X)$.

Let $J$ be a homogeneous ideal and let $X=\mathcal{V}(J)$ be the associated variety in $\mathbb{P}^{n}$. The projective variety $X$ is said to be non-degenerate if $X$ is not contained in any hyperplane $H \subset \mathbb{P}^{n}$. The degree of
the projective variety $X$ of dimension $m$ is equal to the number of complex intersection points between $X$ and a general subspace of dimension $n-m$.

Proposition 1.1.17 (Proposition 9.16 in [47]). Let $V \subset \mathbb{C}^{n} \simeq U_{0} \subset \mathbb{P}^{n}$ be an affine variety embedded in projective space, with ideal $\mathcal{I}(V) \subset \mathbb{C}\left[y_{1}, \ldots, y_{n}\right]$. Let $J \subset \mathbb{C}\left[x_{0}, \ldots, x_{n}\right]$ denote the homogenization of $\mathcal{I}(V)$. Then $\mathcal{J}(V)=J$ and $X(J)=\operatorname{cl}(V)$.

Theorem 1.1.18 (Bertini). Let $X \subset \mathbb{P}^{n}$ be irreducible and non-degenerate of dimension $m \geq 1$ and $H$ a hyperplane in general position, then the following hold:

1. $\operatorname{dim}(X \cap H)=m-1$
2. $X \cap H$ is non-degenerate in $H$
3. $X \cap H$ is irreducible if $m \geq 2$.

### 1.1.2 Rational Maps

Definition 1.1.19. Given a ring $R$ over a field $\mathbb{K}$, the ring of rational functions over $R$ is defined to be

$$
\mathbb{K}\left(x_{1}, \ldots, x_{n}\right)=\left\{f / g: f, g \in \mathbb{K}\left[x_{1}, \ldots, x_{n}\right] \text { and } g \neq 0\right\}
$$

Definition 1.1.20. A rational map $\rho: \mathbb{K}^{n} \rightarrow \mathbb{K}^{m}$, for $n, m \in \mathbb{N}$ is defined by

$$
\rho(x)=\left(\rho_{1}(x), \ldots, \rho_{m}(x)\right)
$$

where $\rho_{i} \in \mathbb{K}\left(x_{1}, \ldots, x_{n}\right)$ for each $i \in\{1, \ldots, m\}$.
Example 1.1.21. [Page 150 [47]] An important rational map we will use in Chapter 4 is the truncated moment map $\gamma: \mathbb{R} \rightarrow \mathbb{R}^{n+1}$,

$$
\gamma(t)=\left(1, t, t^{2}, \ldots, t^{n}\right) .
$$

The projective multivariate analogue of this map is known as the Veronese map. Given some $d \geq 0$, the $d$-th Veronese morphism $v_{d}: \mathbb{P}^{n} \rightarrow \mathbb{P}^{\binom{n+d}{d}-1}$ is given by

$$
v_{d}\left(\left[x_{0}: \ldots: x_{n}\right]\right)=\left[m_{1}: \ldots: m_{\binom{n+d}{d}}\right]
$$

where $\left\{m_{1}, \ldots, m_{\binom{n+d}{d}}\right\}$ is the collection of all monomials of degree $d$ in $n+1$ variables.
Proposition 1.1.22 (Corollary 11.13 of [46]). Let $X$ be an irreducible projective variety and $\rho: X \rightarrow \mathbb{P}^{n}$ a rational map; let $Y=\rho(X)$ be its image. For any $q \in Y$, let $\lambda(q)$ denote the dimension of the fiber over $q$ in $X$, that is $\lambda(q)=\operatorname{dim}\left(\rho^{-1}(q)\right)$. We denote by $\lambda$ the minimum value attained by $\lambda(q)$ on $Y$, then the following holds:

$$
\operatorname{dim}(X)=\operatorname{dim}(Y)+\lambda
$$

### 1.1.3 Semialgebraic Sets

In this section, we direct our attention to subsets of $\mathbb{R}^{n}$ defined by regions of non-negativity for systems of polynomials in $\mathbb{R}\left[x_{1}, \ldots, x_{n}\right]$, which are called semialgebraic sets. Semialgebraic sets are particularly nice because, unlike real varieties, they are closed under projection and taking complements. For more on semialgebraic sets, we refer the reader to Chapter 2 of [19].

Definition 1.1.23. A basic semialgebraic set $S$ in $\mathbb{R}^{n}$ is a set of the form

$$
S=\left\{x \in \mathbb{R}^{n}: f_{1}(x)>0, \ldots, f_{j}(x)>0 \text { and } g_{1}(x)=0, \ldots, g_{k}(x)=0\right\}
$$

where $f_{1}, \ldots, f_{j}, g_{1}, \ldots, g_{k} \in \mathbb{R}\left[x_{1}, \ldots, x_{n}\right]$.
Definition 1.1.24. The closure of a set $S \subset \mathbb{R}^{n}$, denoted $\operatorname{cl}(S)$ is the intersection of all closed sets in $\mathbb{R}^{n}$ containing $S$. A basic closed semialgebraic set $S$ in $\mathbb{R}^{n}$ is a set of the form

$$
S=\left\{x \in \mathbb{R}^{n}: f_{1}(x) \geq 0, \ldots, f_{j}(x) \geq 0\right\}
$$

where $f_{1}, \ldots, f_{j} \in \mathbb{R}\left[x_{1}, \ldots, x_{n}\right]$.
Definition 1.1.25. A general semialgebraic set $S \subset \mathbb{R}^{n}$ is obtained by taking a finite number of unions, intersections and complements of basic semialgebraic sets.

Definition 1.1.26. To a semialgebraic set $S$ we can associate the ideal of real polynomials vanishing on $S$ :

$$
\mathcal{I}(S)=\left\{f \in \mathbb{R}\left[x_{1}, \ldots, x_{n}\right]: f(s)=0 \text { for all } s \in S\right\}
$$

The dimension of a semialgebraic set $S \subset \mathbb{R}^{n}$ is equal to the Krull dimension of $R / \mathcal{I}(S)$, the ring of real polynomials defined on $S$.

Remark 1.1.27. The dimension of an affine linear space $L$, denoted $\operatorname{dim}(L)$ is equal to the cardinality of the basis of $L$. This property is invariant under translation, hence it agrees with our intuitive understanding of dimension: a 0 -dimensional affine space is a point, a 1-dimensional space is a line, a 2-dimensional space is a plane and a $(n-1)$-dimensional space is a hyperplane.

### 1.1.4 Real Algebraic Geometry and Nonnegative Polynomials

When we choose to work over the reals, we lose the algebraic closed property and with it, a lot of the powerful machinery introduced in algebraic geometry. In this section, we go over a few notions about real varieties and semialgebraic sets.

Proposition 1.1.28 (Descartes' rule of signs). Let $p \in \mathbb{R}[x]$ be a univariate polynomial with real coefficients. The number of positive real roots of $p$, counting multiplicity, is bounded above by the number of consecutive sign changes in the nonzero coefficients of $p$.

Theorem 1.1.29 (Tarski-Seidenberg). Let $S \subset \mathbb{R}^{n}$ be a semialgebraic set and let $\Pi: \mathbb{R}^{n} \rightarrow \mathbb{R}^{n-1}$ denote the projection map which takes $\left(x_{1}, \ldots, x_{n}\right)$ to $\left(x_{1}, \ldots, x_{n-1}\right)$. The projection $\Pi(S)$ is also a semialgebraic set in $\mathbb{R}^{n-1}$.

Definition 1.1.30. The pre-ordering $P O$ generated by a set of polynomials $\left\{f_{1}, \ldots f_{r}\right\} \in \mathbb{R}\left[x_{1}, \ldots, x_{n}\right]$ is

$$
P O\left(f_{1}, \ldots, f_{r}\right)=\left\{\sum_{\alpha \in\{0,1\}^{r}} s_{\alpha} f_{1}^{\alpha_{1}} \ldots f_{r}^{\alpha_{r}}\right\},
$$

where $s_{\alpha} \in \mathbb{R}\left[x_{1}, \ldots, x_{n}\right]$ is a sum of squares.
Theorem 1.1.31 (Positivestellensatz). Let $S \subseteq \mathbb{R}^{n}$ be a closed semialgebraic set defined by the inequalities $f_{1}(s) \geq 0, \ldots, f_{r}(s) \geq 0$ and let $T=P O\left(f_{1}, \ldots, f_{r}\right)$ be the associated pre-ordering. For any $f \in \mathbb{R}\left[x_{1}, \ldots, x_{n}\right]$, the following are equivalent:

1. $f(s)>0$ for all $s \in S$
2. there exists $g, h \in T$ such that $g f=1+h$.

Theorem 1.1.32 (Schmüdgen). Suppose $S$ is a compact semialgebraic set defined by the inequalities $f_{1}(s) \geq 0, \ldots, f_{r}(s) \geq 0$ and that some $f \in \mathbb{R}\left[x_{1}, \ldots, x_{n}\right]$ is strictly positive on $S$, then $f \in P O\left(f_{1}, \ldots, f_{r}\right)$.

### 1.2 Convexity

Convexity is a geometric property that plays a central role in optimization theory and is a reoccurring property throughout the subsequent chapters of this dissertation.

### 1.2.1 Convex Sets

A convex set $C$ is a subset of $\mathbb{R}^{n}$ which, for any pair of points in $C$, contains the line segment connecting them. That is, for all $x, y \in C$ and $\lambda \in[0,1]$ we have

$$
\lambda x+(1-\lambda) y \in C .
$$

Definition 1.2.1. The convex hull of a subset $S$ of $\mathbb{R}^{n}$ is the intersection of all convex sets containing $C$

$$
\operatorname{conv}(S)=\bigcap_{\substack{S \subseteq C \\ C \text { is convex }}} C
$$

We let $H=H(l, z)$ denote the (affine) hyperplane in $\mathbb{R}^{n}$ and $H^{+}$the associated halfspace defined by a linear form $l \in\left(\mathbb{R}^{n}\right)^{*}$ and constant $z \in \mathbb{R}$ :

$$
\begin{aligned}
H & =\left\{x \in \mathbb{R}^{n}: l(x)=z\right\} \\
H^{+} & =\left\{x \in \mathbb{R}^{n}: l(x) \geq z\right\} .
\end{aligned}
$$

Definition 1.2.2. The affine span of $S \subset \mathbb{R}^{n}$ is the set of all possible linear combinations of points in $S$

$$
\operatorname{Aff}(S)=\left\{\sum_{i=1}^{k} \lambda_{i} x_{i}: k>0, x_{i} \in S, \lambda_{i} \in \mathbb{R}, \sum_{i=1}^{k} \lambda_{i}=1\right\} .
$$

Definition 1.2.3. A face $F$ of a convex set $C \subset \mathbb{R}^{n}$ is a convex subset which, for any given pair $x, y \in C$, satisfies

$$
\frac{1}{2}(x+y) \in F \Rightarrow x, y \in F
$$

An exposed face $F$ of $C$ is the intersection of $C$ with a supporting hyperplane $H$, meaning that $S$ is contained in the halfspace $S \subseteq H^{+}$and $F=C \cap H$. The dimension of the face $F$ is equal to the dimension of its affine span $\operatorname{Aff}(F)$.


Figure 1.1 The projection of a cylinder.

Example 1.2.4. The projected cylinder of figure 1.1 is a convex semialgebraic set. The two rounded boundaries, including the red points, are all extreme points. The two flat faces are 1-dimensional exposed faces but the 4 red points are non-exposed 0 -dimensional faces.

Definition 1.2.5. An extreme point of a convex set $C$ is a 0 -dimensional face of $C$. It can not be expressed as the convex combination of any two elements in $C$ distinct from itself:

$$
x=\lambda y_{1}+(1-\lambda) y_{2} \text { for some } y_{1}, y_{2} \in S \text { and } \lambda \in(0,1) \Rightarrow x=y_{1}=y_{2}
$$

Proposition 1.2.6. Let $C \subset \mathbb{R}^{n}$ be a convex set and $\ell \in\left(\mathbb{R}^{n}\right)^{*}$ a linear functional we want to maximize (respectively minimize) over $C$. Either $\ell$ is a constant function or its maximum (respectively minimum) over $C$ is attained on the boundary of $C$.

Theorem 1.2.7 (Minkowski). Let $S \subseteq \mathbb{R}^{n}$ be a compact convex set, it can be expressed as the convex
hull of its extreme points.

$$
S=\left\{\lambda_{1} x_{1}+\ldots+\lambda_{k} x_{k}: \lambda_{i} \geq 0, \sum_{i} \lambda_{i}=1, x_{i} \text { an extreme point of } S\right\}
$$

Theorem 1.2.8 (Carathéodory's Theorem). Given a subset $S \subseteq \mathbb{R}^{n}$ and let $C=\operatorname{conv}(S)$ be the convex hull of $S$, then any point in $C$ can be written as a convex combination of at most $n+1$ points in $S$.

### 1.2.2 Convex Cones

The cone $K$ is a convex cone if for all $x, y \in K$ and $\lambda_{1}, \lambda_{2} \geq 0$ we have $\lambda_{1} x+\lambda_{2} y \in K$. A subset $F$ of $K$ is called a face of the cone if for all $x, y \in F$ and $\lambda \in \mathbb{R}_{\geq 0}$, we have $\lambda x \in F$ and $x+y \in F$. The extremal rays of and $n$-dimensional cone are the faces of dimension 1 . A cone is said to be pointed if $x \in K$ and $-x \in K$ implies $x=0$, in other words, $K$ contains no lines. A closed pointed convex cone $K$ admits a compact base, that is, there exists an affine hyperplane $H$ which cuts out a compact cross section $H \cap K$ of the cone away from the origin. The extremal rays of $K$ are in one-to-one correspondence with the extreme points of the compact base $H \cap K$.

Definition 1.2.9. Given a set $S \subset \mathbb{R}^{n}$, we denote by $K$ the cone over $S$

$$
K=\operatorname{cone}(S)=\{\lambda x: \lambda \geq 0, x \in S\}
$$

Proposition 1.2.10 (Colollary of theorem 1.2.7). Given a convex cone $K \subseteq \mathbb{R}^{n}$ with a compact base, any point in $K$ is a conical combination of at most $n$ extreme rays of $K$.

Definition 1.2.11. We define $K^{*}$ to be the dual cone to $K$ as the set of all non negative linear functionals on $K$ :

$$
K^{*}=\left\{l \in \mathbb{R}^{*}: l(x) \geq 0 \text { for all } x \in K\right\}
$$

Proposition 1.2.12. A closed convex cone is pointed if and only if its dual cone is full dimensional.
Theorem 1.2.13 (Bipolarity Theorem). Let $K$ be a convex cone, the dual of the dual of $K$ is equal to the closure of $K$ :

$$
K^{* *}=\operatorname{cl}(K)
$$

Example 1.2.14. [[16]] Let $\mathcal{P}_{n, 2 d}$ denote the cone of nonnegative polynomials on $\mathbb{R}^{n}$, homogeneous of degree $2 d$,

$$
\mathcal{P}_{n, 2 d}=\left\{p \in \mathbb{R}\left[x_{1}, \ldots, x_{n}\right]: \operatorname{deg}(p)=2 d \text { and } p(x) \geq 0 \forall x \in \mathbb{R}^{n}\right\}
$$

The dual cone to $\mathcal{P}_{n, 2 d}$ is $\mathcal{M}_{n, 2 d}$, the conic hull of the image of the unit sphere $\mathbb{S}^{n-1}$ under the Veronese map 1.1.21 of degree $2 d$.

$$
\mathcal{M}_{n, 2 d}=\text { ConicalHull }\left\{v_{2 d}(s): s \in \mathbb{S}^{n-1}\right\}
$$

This object is also know as the Veronese Orbitope [84]. We refer the reader to [15] for an in depth explanation of why this relation holds true.

### 1.2.3 Polyhedra and Polytopes

Polyhedra are in some sense the most fundamental type of semialgebraic convex sets, cut out in $\mathbb{R}^{n}$ by finitely many linear inequalities. Yet they are by no means simple objects and their combinatorial structures are intricate and fascinating. Polyhedra are central to geometry and optimization over the reals. As we will see in the following sections and chapters, polytopes and special arrangements thereof can record combinatorial information about algebraic objects, geometric objects and optimization problems. For an extensive journey into the world of polytopes, we refer the reader to Ziegler [104].

Definition 1.2.15 (From [104]). A polyhedron $P \subseteq \mathbb{R}^{n}$ is a semialgebraic set given by the intersection of $m$ closed halfspaces, for some $m \in \mathbb{N}$

$$
\begin{aligned}
P & =\bigcap_{i=1}^{m}\left\{x \in \mathbb{R}^{n}: l_{i}(x) \leq z_{i}\right\} & & \text { where } l_{i} \in\left(\mathbb{R}^{n}\right)^{*} \text { and } z_{i} \in \mathbb{R} \\
& =\left\{x \in \mathbb{R}^{n}: A x \leq z\right\} & & \text { where } A \in \mathbb{R}^{m \times n}, z \in \mathbb{R}^{m} .
\end{aligned}
$$

A polytope is a bounded polyhedron. There are two ways equivalent ways to think about polytopes, either as just stated, a bounded set defined by finitely many linear inequalities or the convex hull of a finite set of points $v_{i} \in \mathbb{R}^{n}$.

Theorem 1.2.16 (Main theorem for polytopes [104]). A subset $P \subseteq \mathbb{R}^{n}$ is the convex hull of a finite collection of points $V \in \mathbb{R}^{n}$

$$
P=\operatorname{conv}(V)
$$

if and only if it is a bounded intersection of halfspaces defined by some $A \in \mathbb{R}^{m \times n}$ and $z \in \mathbb{R}^{m}$

$$
P=\left\{x \in \mathbb{R}^{n}: A x \leq z\right\} .
$$

Definition 1.2.17. Given a point configuration $X$ in $\mathbb{R}^{n}$, a subdivision $\Delta$ of $X$ is a finite collection of $n$-dimensional polytopes $\sigma_{i}$, such that the union of polytopes in $\Delta$ equals $\operatorname{conv}(X)$, the vertex set of polytopes in $\Delta$ is contained in $X$ and any two distinct polytopes $\sigma_{1}, \sigma_{2}$ in $\Delta$ are either disjoint or they intersect along the lower dimensional face $\sigma_{1} \cap \sigma_{2}$.

### 1.2.4 Spectrahedra

Spectrahedra are a family of convex bodies defined by linear matrix inequalities. They play a central role in convex optimization. For a more in depth, yet very concise introduction to spectrahedra, we recommend the AMS notices article [99].

Proposition 1.2.18. Given a symmetric real matrix $A \in \mathbb{R}^{d \times d}$, the following are equivalent:

1. $A$ is positive semidefinite: $A \geqslant 0$
2. $x^{T} A x \geq 0$ for all $x \in \mathbb{R}^{d}$
3. All of the eigenvalues $\lambda_{i}(A)$ are nonnegative, for $i=1 \ldots d$.
4. A admits an eigenvalue decomposition $A=U D U^{T}$, where $U$ is an orthogonal matrix, i.e. $U^{T} U=\operatorname{Id}(d)$, and $D$ a diagonal matrix with nonnegative entries on the diagonal.
5. All principal minors of $A$ are nonnegative.

Definition 1.2.19. Let $\mathcal{S}_{+}^{d}$ denote the cone of $d \times d$ positive semidefinite (PSD) matrices.
Proposition 1.2.20. The cone of PSD matrices is closed and convex of dimension $n(n+1) / 2$. The boundary of this cone consists of singular positive semidefinite matrices and the extreme rays of the cone correspond to rank-one PSD matrices.

A spectrahedron $\mathcal{S}$ is a linear slice of the cone of symmetric positive semidefinite matrices $\mathcal{S}_{+}^{d}$.

$$
\mathcal{S}=\left\{x \in \mathbb{R}^{n}: A(x) \geqslant 0\right\}
$$

where $A(x)=A_{0}+\sum_{i=1}^{n} x_{i} A_{i}$, where each $A_{i} \in \mathbb{R}^{d \times d}$.
Example 1.2.21. The moment curve is the image of the map $\gamma: \mathbb{R} \rightarrow \mathbb{R}^{n}$ given by

$$
\gamma(t)=\left(1, t, t^{2}, \ldots, t^{n-1}\right)
$$

This curve segment is the collection of moments of Dirac measures $\delta_{t}$ for $t \in[0,1]$ :

$$
\int f(x) \delta_{t}(x)=f(t)
$$

We refer to those Dirac measures supported on a single point in $[0,1]$ as point masses on $[0,1]$. The convex hull of this curve segment is the spectrahedron depicted in Figure 1.2. The dual to this spectrahedron is the collection of univariate polynomials nonnegative on $[0,1]$ of degree at most three.

### 1.2.5 Applications to Optimization

We briefly introduce two important classes of convex optimization problems, which both consist of maximizing a linear functional over a convex semialgebraic set. The first type being Linear Programming (LP), where $K$ is a polyhedral cone and the second type being Semi-Definite Programming (SDP), where $K$ is the cone of semidefinite matrices. Both LP and SDP problems are solvable in polynomial time via interior point methods. For a thorough foray into convex optimization and semidefinite programming, we recommend the book [17].

Definition 1.2.22. Linear Programming is the process of maximizing a linear functional subject to linear constraints, which can be re-stated as:

$$
\max c^{T} x \quad \text { s.t. } A x \leq b
$$

where $c \in \mathbb{R}^{n}, A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^{m}$.


Figure 1.2 Affine slice of the convex hull of $\gamma:[0,1] \rightarrow \mathbb{R}^{4}$, for $\gamma(t)=\left(1, t, t^{2}, t^{3}\right)$.

If the linear functional $c^{T} x$ is non-constant on the polyhedron defined by $A x \leq b$, then the optimal solution $x^{*}$ is attained on the boundary of the feasible set.

Definition 1.2.23. In semidefinite programming, we optimize over the cone of positive semidefinite real symmetric $d \times d$-matrices. The feasibility sets of semidefinite programs are spectrahedra.

A function $f: S \rightarrow \mathbb{R}$, with a convex domain of definition $S$, is said to be convex if its epigraph is convex. That is, for any two points $x_{1}, x_{2} \in S, f$ has to satisfies

$$
f\left(t x_{1}+(1-t) x_{2}\right) \leq t f\left(x_{1}\right)+(1-t) f\left(x_{2}\right) \text { for all } t \in[0,1]
$$

A function $f: S \rightarrow \mathbb{R}$ is said to be concave if $(-f)$ is convex.

### 1.3 Discrete Geometry and Algebraic Combinatorics

In this section, we introduce a few concepts at the intersection of combinatorics and algebraic geometry. We refer the reader to [68] for an introduction to algebraic combinatorics.

### 1.3.1 Simplicial Complexes

We introduce simplicial complexes, first as purely combinatorial abstractions and then as arrangements of simplices in $\mathbb{R}^{n}$. These complexes will play a particularly useful role in Chapter 2 when generating squarefree monomial ideals, and in Chapter 3 in the form of maximal subdivisions of point configurations. One of the classical motivation for working with simplicial complexes is rooted in the study of topological spaces in algebraic topology [48]. We refer to [53] for a deeper dive into the world of simplicial complexes and their applications.

Definition 1.3.1. An abstract simplex $\sigma$ on the ground set $[m]=\{1, \ldots, m\}$ of dimension $k \in \mathbb{N}$, also referred to as an abstract $k$-simplex, is a subset of $[m]$ of cardinality $k+1$. An abstract simplicial complex $\Delta$ on $[m]$ is a finite collection of abstract simplices $\sigma \subseteq[m]$, called faces of $\Delta$. A simplicial complex has to be closed under containment and intersection:

1. given a face $\sigma_{1} \in \Delta$ and $\sigma_{2} \subsetneq \sigma_{1}$, then $\sigma_{2}$ also has to be an element of $\Delta$.
2. if $\sigma_{1}, \sigma_{2} \in \Delta$, then their intersection $\sigma_{1} \cap \sigma_{2}$ also must belong to $\Delta$.

Definition 1.3.2. The dimension of a face $\sigma \in \Delta$ is equal to $\operatorname{dim}(\sigma)=|\sigma|-1$. The dimension of the simplicial complex $\Delta$ is given by the largest dimension of its faces:

$$
\operatorname{dim}(\Delta)=\max _{\sigma \in \Delta}(\operatorname{dim}(\sigma))
$$

The maximal faces in $\Delta$, those properly contained in no other face, are called facets of $\Delta$.
Definition 1.3.3. Given a simplicial complex $\Delta$ on $[m] \backslash\{i\}$, let the cone of $\Delta$ over $i$ be:

$$
\operatorname{cone}(\Delta, i)=\Delta \cup\{\sigma \cup\{i\}: \sigma \in \Delta\}
$$

which is a simplicial complex on the ground set $[m]$ and whose facets are in bijection with the facets of $\Delta$. The link of a face $S$ in $\Delta$ is defined as

$$
\operatorname{link}_{\Delta}(S)=\{\sigma \in \Delta: \sigma \cap S=\emptyset, \sigma \cup S \in \Delta\}
$$

We define the neighborhood $\mathcal{N}(j)$ of a vertex $j$ in $\Delta$ to be the set of vertices

$$
\mathcal{N}(j)=\{i:\{i, j\} \subseteq \sigma \text { for some } \sigma \in \Delta\}
$$

Definition 1.3.4. A $k$-simplex $\sigma$ in $\mathbb{R}^{n}$ is the convex hull of $k+1$ points in general position in $\mathbb{R}^{n}$. A simplicial complex in $\mathbb{R}^{n}$ is a finite collection of $n$-simplices satisfying analogous properties to the ones of an abstract simplicial complex, that is, two distinct faces $\sigma_{1}, \sigma_{2}$ of $\Delta$ can only be disjoint or intersect along the lower dimensional face $\sigma_{1} \cap \sigma_{2}$.

The $n$-simplex in $\mathbb{R}^{n}$ is the full dimensional polytope in $\mathbb{R}^{n}$ with the fewest vertices and fewest facets. It can be equivalently expressed as the compact intersection of $n+1$ half-spaces.

Definition 1.3.5. Let $P$ be the convex hull of a point configuration $X=\left(x_{1}, x_{2}, \ldots, x_{m}\right) \subset \mathbb{R}^{n}$. For a fixed real vector $y \in \mathbb{R}^{m}$, we define a function $h_{X, y}$ on $\mathbb{R}^{n}$, called the tent function, as the smallest concave function such that

$$
h_{X, y}\left(x_{i}\right) \geq y_{i} \quad \text { for all } i \in[m]
$$

The tent function $h_{X, y}$ is piecewise linear on $P$ with linear pieces equal to upper facets of the convex hull of the points $\left(x_{1}, y_{1}\right),\left(x_{2}, y_{2}\right), \ldots,\left(x_{m}, y_{m}\right)$ in $\mathbb{R}^{n+1}$. We define $h_{X, y}(x)=-\infty$ at all points $x \in \mathbb{R}^{n}$ outside $P$. If $h_{X, y}\left(x_{i}\right)=y_{i}$ for $i=1, \ldots, m$, then $y$ is called relevant.


Figure 1.3 A Tent Function appearing in [79].

Definition 1.3.6 ([29]). Given a point configuration $X=\left(x_{1}, x_{2}, \ldots, x_{m}\right) \subset \mathbb{R}^{n}$, a triangulation of $X$ is a simplicial complex $\Delta=\left\{\sigma_{1}, \ldots, \sigma_{k}\right\}$ with all vertices contained in $X$, satisfying

$$
\operatorname{conv}(X)=\bigcup_{i=1}^{k} \sigma_{i}
$$

A triangulation $\Delta$ of the point configuration $X$ is said to be maximal if every element of $X$ appears in the vertex set of $\Delta$. A subdivision is called regular if its full dimensional cells $\sigma_{i}$ are combinatorially equivalent to the regions of linearity of a tent function on $X$ for some height vector $y \in \mathbb{R}^{m}$.

### 1.3.2 Initial Ideals

For $w \in\left(\mathbb{R}_{\geq 0}\right)^{n}$ and $f=\sum_{\alpha} c_{\alpha} \mathbf{x}^{\alpha} \in \mathbb{C}\left[x_{1}, \ldots, x_{n}\right]$, where $\mathbf{x}^{\alpha}$ denotes $x_{1}^{\alpha_{1}} \ldots x_{n}^{\alpha_{n}}$, define the degree and initial form of $f$ with respect to $w$ to be

$$
\operatorname{deg}_{w}(f)=\max \left\{w^{T} \alpha: c_{\alpha} \neq 0\right\} \quad \text { and } \quad \operatorname{in}_{w}(f)=\sum_{\alpha: w^{T}} c_{\alpha=\operatorname{deg}_{w}(f)} c_{\alpha} \mathbf{x}^{\alpha}
$$

In that sense, every choice of weight vector $w$ defines a monomial order on $\left[x_{1}, \ldots, x_{n}\right]$.
Definition 1.3.7. $\mathrm{in}_{w}(I)$ is the ideal generated by initial forms of polynomials in $I$, i.e.

$$
\operatorname{in}_{w}(I)=\left\langle\operatorname{in}_{w}(f): f \in I\right\rangle
$$

A nice property that we will make use of in next chapter is the following.
Proposition 1.3.8 (Proposition 4, §9.3 of [24]). An ideal $I \subset \mathbb{C}\left[x_{1}, \ldots, x_{n}\right]$ and its initial ideal $\mathrm{in}_{w}(I)$, for all $w \in \mathbb{R}_{\geq 0}^{n+1}$, share the same affine Hilbert function

$$
\operatorname{HF}_{I}(s)=\operatorname{HF}_{\mathrm{in}_{w}(I)}(s)
$$

It follows that $\operatorname{dim}(R / I)=\operatorname{dim}\left(R / \mathrm{in}_{w}(I)\right)$, thus the Krull dimension of the rings $R / I$ and $R / \mathrm{in}_{w}(I)$
coincide.
As we will see in the next section, initial ideals are a key concept in defining Gröbner bases.

### 1.3.3 Connections with Algebraic Geometry

A set of polynomials $G=\left\{g_{1}, \ldots, g_{k}\right\} \subset I$ satisfying $\operatorname{in}_{w}(I)=\left\langle\operatorname{in}_{w}\left(g_{1}\right), \ldots \operatorname{in}_{w}\left(g_{k}\right)\right\rangle$ defines a Gröbner basis of $I$ with respect to $w$. A finite subset $F$ of an ideal $I \subset \mathbb{C}\left[x_{1}, \ldots, x_{n}\right]$ is a universal Gröbner basis for $I$ if it is a Gröbner basis with respect to every monomial order on $\mathbb{C}\left[x_{1}, \ldots, x_{n}\right]$. An equivalent definition using weight vectors is given as follows: A finite collection of polynomials $F \subset I$ is a universal Gröbner basis for $I$ if and only if for every $w \in\left(\mathbb{R}_{\geq 0}\right)^{n}$, the polynomials $\operatorname{in}_{w}(F)$ generate $\operatorname{in}_{w}(I)$. See [95, Chapter 1].

For homogeneous $J, \operatorname{in}_{w}(J)$ is a flat degeneration of $J$. For $f \in \mathbb{C}\left[x_{1}, \ldots, x_{n}\right], w \in \mathbb{Z}^{n}$, define $t^{w} \cdot f=t^{\operatorname{deg}_{w}(f)} f\left(t^{-w_{1}} x_{1}, \ldots, t^{-w_{n}} x_{n}\right) \in \mathbb{C}\left[t, x_{1}, \ldots, x_{n}\right]$ and $\bar{J}^{w}=\left\langle t^{w} \cdot f: f \in J\right\rangle \subset \mathbb{C}\left[t^{ \pm}, x_{1}, \ldots, x_{n}\right]$. The ideal $\bar{J}^{w}$ defines an ideal in $\mathbb{A}^{1}(\mathbb{C}) \times \mathbb{P}^{n-1}(\mathbb{C})$, namely the Zariski-closure

$$
\mathcal{V}\left(\bar{J}^{w}\right)=\operatorname{cl}\left(\left\{\left(t,\left[t^{w_{1}} x_{1}: \ldots: t^{w_{n}} x_{n}\right]\right) \text { such that } t \in \mathbb{C}^{*}, x \in \mathcal{V}(J)\right\}\right)
$$

Letting $t$ vary from 1 to 0 gives a flat deformation from $\mathcal{V}(J)$ to the variety $\mathcal{V}\left(\operatorname{in}_{w}(J)\right)$. Formally, for any $\gamma \in \mathbb{C}$, let $\bar{J}^{w}(\gamma)$ denote the ideal in $\mathbb{C}\left[x_{1}, \ldots, x_{n}\right]$ obtained by substituting $t=\gamma$. Then $\bar{J}^{w}(1)$ equals $J, \bar{J}^{w}(0)$ equals $\operatorname{in}_{w}(J)$, and for $\gamma \in \mathbb{C}^{*}$, the variety of $\bar{J}^{w}(\gamma)$ consists of the points $\left\{\left[\gamma^{w_{1}} x_{1}: \ldots: \gamma^{w_{n}} x_{n}\right]: x \in \mathcal{V}(J)\right\}$. All the ideals $\bar{J}^{w}(\gamma)$ have the same Hilbert series. In particular, taking $\gamma=0,1$ shows that $J$ and $\operatorname{in}_{w}(J)$ have the same Hilbert series.

### 1.4 Contents of the Following Chapters

The underlying common theme to the 3 following chapters is the combinatorial approach to studying geometric objects arising from algebraic geometry and optimization. In chapter 2, we study the image of linear spaces under partial coordinate inversion. Through a combinatorial construction, we are able to compute universal Gröbner bases for the vanishing ideals of these semialgebraic sets.

In chapter 3, we study the exact solutions to log-concave maximum likelihood estimation problems. We pay particular attention to how the accuracy on the output of this optimization problem is affected by the regular subdivision it induces on the ground set.

In chapter 4 , we study a variation of the classical truncated moment problem on a compact subset of $\mathbb{R}$, a classical problem in functional analysis and real algebraic geometry [86]. We expand this question to moments of step function and a finite subset $A \subset \mathbb{N}$. Our main objects of interest are the full moment cone $M(A)$ and the moment cone of step functions with at most $k$ steps $M_{k}(A)$. The motivation to study the moment cone $M(A)$ is rooted in population genetics, through its connection to the coalescence manifold [80]. The moment cone $M(A)$ is a convex semi algebraic set [93], we provide a semidefinite program to compute the minimal distance to the boundary of the moment cone.

## Chapter 2

## Semi-Inverted Linear Spaces

The content of this chapter is based on joint work with C. Vinzant appearing in Algebraic Combinatorics [88]. The variety of interest in this chapter is the image of a linear space $\mathcal{L} \subseteq \mathbb{C}^{n}$ under partial coordinate inversion, that is, we apply the map $x_{i} \rightarrow 1 / x_{i}$ to the subset of coordinates indexed by the set $I \subseteq\{1, \ldots, n\}$. We denote the variety obtained in this manner $\operatorname{inv}_{I}(\mathcal{L})$. We draw connections to two well studied objects in algebraic combinatorics: the hyperplane arrangement $[2,32]$ defining $\mathcal{L}$ and the associated matroid $M(\mathcal{L})[2,32,34,69]$. We construct a universal Gröbner basis of the vanishing ideal of $\operatorname{inv}_{I}(\mathcal{L})$ in theorem 2.2.1,

### 2.1 Background

In 2006, Proudfoot and Speyer showed that the coordinate ring of a reciprocal linear space (i.e. the closure of the image of a linear space under coordinate-wise inversion) has a flat degeneration into the Stanley-Reisner ring of the broken circuit complex of a matroid [73]. This completely characterizes the combinatorial data of these important varieties, which appear across many areas of mathematics, including in the study of matroids and hyperplane arrangements [96], interior point methods for linear programming [30], and entropy maximization for log-linear models in statistics [67].

### 2.1.1 Reciprocal Linear Spaces

In the special case of $I=[n]$, the variety $\operatorname{inv}_{I}(\mathcal{L})$ is well-studied in the literature. Proudfoot and Speyer study the coordinate ring of the variety $\operatorname{inv}_{[n]}(\mathcal{L})$ and relate it to the broken circuit complex of a matroid [73]. One of their motivations is connections with the cohomology of the complement of a hyperplane arrangement [96] and the close relationship with the Orlik-Terao algebra [74], [96]. These varieties also appear in the algebraic study of interior point methods for linear programming [30] and entropy maximization for log-linear models in statistics [67].

If the linearspace $\mathcal{L}$ is invariant under complex conjugation, the variety $\operatorname{inv}_{[n]}(\mathcal{L})$ also has a special real-rootedness property. Specifically, if $\mathcal{L}^{\perp}$ denotes the orthogonal complement of $\mathcal{L}$, then for any $u \in \mathbb{R}^{n}$, all the intersection points of $\operatorname{inv}_{[n]}(\mathcal{L})$ and the affine space $\mathcal{L}^{\perp}+u$ are real. This was first
shown in different language by Varchenko [98] and used extensively in [30]. One implication of this real-rootedness is that the discriminant of the projection away from $\mathcal{L}^{\perp}$ is a nonnegative polynomial [85]. Another is that $\operatorname{inv}_{[n]}(\mathcal{L})$ is a hyperbolic variety, in the sense of [90]. In fact, the Chow form of the variety $\operatorname{inv}_{[n]}(\mathcal{L})$ has a definite determinantal representation, certifying its hyperbolicity [56]. We generalize some of this results to $\operatorname{inv}_{I}(\mathcal{L})$.

In closely related work [3], Ardila and Boocher study the closure of a linear space $\mathcal{L}$ inside of $\left(\mathbb{P}^{1}\right)^{n}$. For any $I \subseteq[n], \operatorname{inv}_{I}(\mathcal{L})$ can be considered an affine chart of this projective closure. Specifically, let $X \subseteq\left(\mathbb{P}^{1}\right)^{n}$ denote the closure of the image of $\mathcal{L}$ under the map

$$
\left(x_{1}, \ldots, x_{n}\right) \mapsto\left(\left[x_{1}: y_{1}\right],\left[x_{2}: y_{2}\right], \ldots,\left[x_{n}: y_{n}\right]\right),
$$

with $y_{1}=\ldots=y_{n}=1$. The restriction of $X$ to the affine chart $x_{i}=1$ for $i \in I$ and $y_{j}=1$ for $j \in[n] \backslash I$ is isomorphic to $\operatorname{inv}_{I}(\mathcal{L})$.

### 2.1.2 Matroids

Matroids are a combinatorial generalization of the notion of independence relations we know from linear algebra. The motivation for using matroids in chapter is to build an analogue of the brokencircuit complex, a special simplicial complex which maximal non-faces are in correspondence with the circuits of a given matroid. Matroids are a classical tool in combinatorial geometry with a multitude of applications and variations, such as oriented matroids [13] and valuated matroids [34]. For a more in depth source of information on matroids and their applications to algebra and geometry, we refer to [2, 69].

We introduce four cryptomorphic definitions of the concept of matroids. Traditionally, matroids are defined over some abstract ground set $E$, but we do not require such level of generality, hence we fix $E$ to be equal to $[n]=[1,2, \ldots, n]$.

Definition 2.1.1. A matroid $M=([n], \mathcal{I})$ is a collection $\mathcal{I}$ of independent subsets of $[n]$ satisfying:

1. $\emptyset \in I$
2. if $A \in I$ and $B \subset A$ then $B \in I$
3. if $A, B \in I$ and $|A|<|B|$ then there exists an element $b \in B-A$ such that $A \cup b \in I$.

Maximal independent sets of a matroid are called bases and minimal dependent sets are called circuits.

Definition 2.1.2. A matroid $M=([n], \mathcal{B})$ is a collection $\mathcal{B}$ of bases on $[n]$ satisfying the exchange of basis axioms:

1. $\mathcal{B} \neq \emptyset$
2. For any $A, B \in \mathcal{B}$ if $a \in A-B$ then there exists some $b \in B-A$ such that $(A \cup b-a) \in \mathcal{B}$.

Definition 2.1.3. A matroid $M=([n], C)$ is a collection $C$ of circuits on $[n]$ satisfying the exchange of circuits axioms:

1. $\emptyset \notin C$.
2. If $C \in C$ and $C^{*}$ is a proper subset of $C$, then $C^{*} \notin C$.
3. If $C_{1}, C_{2} \in C$ are two distinct circuits with $c \in C_{1} \cap C_{2}$, then there exist a circuit $C_{3} \in C$ contained in $C_{1} \cup C_{2}-c$.

Notation 2.1.4. Given $X \subset[n]$, let us denote by $X^{c}$ the set $[n]-X$.
Definition 2.1.5. [69] A matroid $M=([n], \mathcal{F})$ is a collection $\mathcal{F}$ of flats on $[n]$ satisfying:

1. $[n] \in \mathcal{F}$.
2. If $F_{1}, F_{2} \in \mathcal{F}$, then $F_{1} \cap F_{2} \in \mathcal{F}$.
3. Let $F \in \mathcal{F}$ and $\left\{F_{1}, F_{2}, \ldots, F_{k}\right\}$ be the set of minimal members of $\mathcal{F}$ which properly contain $F$. The set $\left\{F_{1}-F, \ldots, F_{k}-F\right\}$ forms a partition of $F^{c}$.

We use $\mathcal{B}(M), \mathcal{C}(M)$ and $\mathcal{F}(M)$ to denote the set of bases, circuits, and flats of a matroid $M$. For ease of notation, we drop the $(M)$ when it is clear what matroid is referred to. To summarize our notations: Let $M$ be a matroid on the ground set [ $n$ ]

1. $I$ is the collection of independent subsets of $M$
2. $C$ is the collection of circuits of $M$
3. $\mathcal{B}$ is the collection of bases of $M$
4. $\mathcal{F}$ is the collection of flats of $M$

Lemma 2.1.6. Given the collection of circuits $C$ of a matroid $M$ and a set $C \subset[n]$, we have that $C \in C$ if and only if $C \notin I$ and $(C-x) \in \mathcal{I}$, for all $x \in C$.

Corollary 2.1.7 (1.2.6 [69]). Let $B$ be a basis of $M$ and if e belongs to $B^{c}$, then $(B \cup e)$ contains a unique circuit $C$, and that circuit must contain $e$.

An element $i \in[n]$ is called a loop if $\{i\}$ is a circuit, and a co-loop if $i$ is contained in every basis of $M$. The rank of a subset $S \subseteq[n]$ is the largest size of an independent set in $S$. A flat is a set $F \subseteq[n]$ that is maximal for its rank, meaning that $\operatorname{rank}(F)<\operatorname{rank}(F \cup\{i\})$ for any element $i \in[n]$ that is not already contained in $F$.

Given a matroid $M$ on a ground set [ $n$ ] and some element $i \in[n]$, the deletion of $M$ by $i$, denoted $M \backslash i$, is the matroid on the ground set $[n] \backslash i$ whose independent sets are subsets $I \subset[n] \backslash i$ that are independent in $M$. If $i$ is not a co-loop of $M$, then

$$
\mathcal{B}(M \backslash i)=\{B \in \mathcal{B}(M): i \notin B\} \text { and } C(M \backslash i)=\{C \in \mathcal{C}(M): i \notin C\} .
$$

More generally, the deletion of $M$ by a subset $S \subset[n]$, denoted $M \backslash S$, is the matroid obtained from $M$ by successive deletion of the elements of $S$. The restriction of $M$ to a subset $S$, denoted $\left.M\right|_{S}$, is the deletion of $M$ by $[n] \backslash S$.

If $i$ is not a loop of $M$, then the contraction of $M$ by $i$, denoted $M / i$, is the matroid on the ground set $[n] \backslash\{i\}$ whose independent sets are subsets $I \subset[n] \backslash i$ for which $I \cup\{i\}$ is independent in $M$. Then

$$
\mathcal{B}(M / i)=\{B \backslash i: B \in \mathcal{B}(M), i \in B\} \quad \text { and } \quad C(M / i)=\text { min. elts. of }\{C \backslash i: C \in C(M)\}
$$

If $i$ is a loop of $M$, then we define the contraction of $M / i$ to be the deletion $M \backslash i$. The contraction of $M$ by a subset $S \subset[n]$, denoted $M / S$, is obtained from $M$ by successive contractions by the elements of $S$.

### 2.1.3 Linear Matroids

Given a $d$-dimensional linearspace $\mathcal{L} \subset \mathbb{C}^{n}$ we can associate to it a matroid $M(\mathcal{L})$ as follows: Write the linearspace $\mathcal{L} \subset \mathbb{C}^{n}$ as the rowspan of a $d \times n$ matrix $A=\left(a_{1}, \ldots, a_{n}\right)$. A set $I \subseteq[n]$ is independent in $M(\mathcal{L})$ if the vectors $\left\{a_{i}: i \in I\right\}$ are linearly independent in $\mathbb{C}^{d}$. For any invertible matrix $U \in \mathbb{C}^{d \times d}$, the vectors $\left\{a_{i}: i \in I\right\}$ are linearly independent if and only if the vectors $\left\{U a_{i}: i \in I\right\}$ are also independent, meaning that this condition is independent of the choice of basis for $\mathcal{L}$. Indeed, $I \subseteq[n]$ is independent in $M(\mathcal{L})$ if and only if the coordinate linear forms $\left\{x_{i}: i \in I\right\}$ are linearly independent when restricted to $\mathcal{L}$.

For linear matroids, deletion corresponds to projection and contraction corresponds to intersection in the following sense. For $S \subset[n]$, let $\mathcal{L} \backslash S$ denote the linear subspace of $\mathbb{C}^{[n] \backslash S}$ obtained by projecting $\mathcal{L}$ away from the coordinate space $\mathbb{C}^{S}=\operatorname{span}\left\{e_{i}: i \in S\right\}$. Let $\mathcal{L} / S$ denote the intersection of $\mathcal{L}$ with $\mathbb{C}^{[n] \backslash S}$. Then

$$
M(\mathcal{L}) \backslash S=M(\mathcal{L} \backslash S) \text { and } \quad M(\mathcal{L}) / S=M(\mathcal{L} / S)
$$

The following definition will play a central role in the next Chapter. We will extend this definition by introducing $\mathcal{I}$-broken circuit complexes.

Definition 2.1.8. Given a matroid $M=M([n], C)$ with the standard ordering $1<2<\ldots<n$, a brokencircuit of $M$ is a subset of the form $C \backslash \min (C)$ where $C \in C(M)$.

### 2.1.4 Stanley Reisner Ideals

A Stanley-Reisner ideal is a square-free monomial ideal that we define from a simplicial complex $\Delta$.
Definition 2.1.9 ([68, Def. 1.6]). Let $\Delta$ be a simplicial complex on vertices $\{1, \ldots, n\}$. The StanleyReisner ideal of $\Delta$ is the square-free monomial ideal

$$
I_{\Delta}=\left\langle\mathbf{x}^{S}: S \subseteq[n], S \notin \Delta\right\rangle
$$

generated by monomial corresponding to the non-faces of $\Delta$. The Stanley-Reisner ring of $\Delta$ is the quotient ring $\mathbb{C}[\mathbf{x}] / \mathcal{I}_{\Delta}$.

An interesting observation is that combinatorial properties of a matroid can be encoded in a simplicial complex, called the broken circuit complex. The broken-circuit complex of a matroid $M([n], C)$, with the usual ordering $1<2<\ldots<n$ on [ $n$ ], is the simplicial complex with vertices 1 through $n$, whose faces are the subsets of $[n]$ not containing any broken circuit.

The ideal $I_{\Delta}$ is radical and it equals the intersection of prime ideals

$$
\mathcal{I}_{\Delta}=\bigcap_{F \text { a facet of } \Delta}\left\langle x_{i}: i \notin F\right\rangle
$$

This writes the variety $\mathcal{V}\left(\mathcal{I}_{\Delta}\right)$ as the union of coordinate subspaces span $\left\{e_{i}: i \in F\right\}$ where $F$ is a facet of $\Delta$. In particular, if $\Delta$ has $k$ facets of dimension $d-1$, then $\mathcal{V}\left(I_{\Delta}\right)$ is a variety of dimension $d$ and degree $k$. See [68, Chapter 1] for more.

### 2.2 Main Results

We extend the results of Proudfoot and Speyer to the image of a linear space $\mathcal{L} \subset \mathbb{C}^{n}$ under inversion of some subset of coordinates. For $I \subseteq\{1, \ldots, n\}$, consider the rational map $\operatorname{inv}_{I}: \mathbb{C}^{n} \rightarrow \mathbb{C}^{n}$ defined by

$$
\left(\operatorname{inv}_{I}(x)\right)_{i}=\left\{\begin{array}{cc}
1 / x_{i} & \text { if } i \in I \\
x_{i} & \text { if } i \notin I
\end{array}\right.
$$

Let $\operatorname{inv}_{I}(\mathcal{L})$ denote the Zariski-closure of the image of $\mathcal{L}$ under this map, which is an affine variety in $\mathbb{C}^{n}$. One can interpret $\operatorname{inv}_{I}(\mathcal{L})$ as an affine chart of the closure of $\mathcal{L}$ in the product of projective spaces $\left(\mathbb{P}^{1}\right)^{n}$, as studied in [3], or as the projection of the graph of $\mathcal{L}$ under the map $x \rightarrow \operatorname{inv}_{[n]}(x)$, studied in [41], onto complementary subsets of the $2 n$ coordinates. We give a degeneration of the coordinate ring of $\operatorname{inv}_{I}(\mathcal{L})$ to the Stanley-Reisner ring of a simplicial complex generalizing the broken circuit complex of a matroid. This involves constructing a universal Gröbner basis for the ideal of polynomials vanishing on $\operatorname{inv}_{I}(\mathcal{L})$.

Let $\mathbb{C}[\mathbf{x}]$ denote the polynomial ring $\mathbb{C}\left[x_{1}, \ldots, x_{n}\right]$ and for any $\alpha \in\left(\mathbb{Z}_{\geq 0}\right)^{n}$, let $\mathbf{x}^{\alpha}$ denote $\prod_{i=1}^{n} x_{i}^{\alpha_{i}}$. For a subset $S \subseteq[n]$, we will also use $\mathbf{x}^{S}$ to denote $\prod_{i \in S} x_{i}$. As in [73], the circuits of the matroid $M(\mathcal{L})$ corresponding to $\mathcal{L}$ give rise to a universal Gröbner basis for the ideal of polynomials vanishing on $\operatorname{inv}_{I}(\mathcal{L})$. We say that a linear form $\ell(x)=\sum_{i \in[n]} a_{i} x_{i}$ vanishes on $\mathcal{L}$ if $\ell(x)=0$ for all $x \in \mathcal{L}$. The support of $\ell, \operatorname{supp}(\ell)$, is $\left\{i \in[n]: a_{i} \neq 0\right\}$. The minimal supports of linear forms vanishing on $\mathcal{L}$ are the circuits of the matroid $M(\mathcal{L})$. For every circuit $C \subset[n]$ of $M(\mathcal{L})$, there is a unique (up to scaling) linear form $\ell_{C}=\sum_{i \in C} a_{i} x_{i}$ vanishing on $\mathcal{L}$ with support $C$. To each circuit, we associate the polynomial

$$
\begin{equation*}
f_{C}(\mathbf{x})=\mathbf{x}^{C \cap I} \cdot \ell_{C}\left(\operatorname{inv}_{I}(\mathbf{x})\right)=\sum_{i \in C \cap I} a_{i} \mathbf{x}^{C \cap I \backslash\{i\}}+\sum_{i \in C \backslash I} a_{i} \mathbf{x}^{C \cap I \cup\{i\}} \tag{2.1}
\end{equation*}
$$

Theorem 2.2.1. Let $\mathcal{L} \subseteq \mathbb{C}^{n}$ be a d-dimensional linearspace and let $\mathcal{I} \subseteq \mathbb{C}[\mathbf{x}]$ be the ideal of polynomials vanishing on $\operatorname{inv}_{I}(\mathcal{L})$. Then $\left\{f_{C}: C\right.$ is a circuit of $\left.M(\mathcal{L})\right\}$ is a universal Gröbner basis for
I. In particular, for $w \in\left(\mathbb{R}_{+}\right)^{n}$ with distinct coordinates, the initial ideal $\mathrm{in}_{w}(\mathcal{I})$ is the Stanley-Reisner ideal of the semi-broken circuit complex $\Delta_{w}(M(\mathcal{L}), I)$.

The simplicial complex $\Delta_{w}(M(\mathcal{L}), I)$ will be defined in Section 2.3 . For real linear spaces $\mathcal{L}$, the variety $\operatorname{inv}_{I}(\mathcal{L})$ relates to the regions of a hyperplane arrangement as follows.

Theorem 2.2.2. Let $\mathcal{L} \subset \mathbb{C}^{n}$ be invariant under complex conjugation. The following are equal:

1. the degree of the affine variety $\operatorname{inv}_{I}(\mathcal{L})$,
2. the number of facets of the semi-broken circuit complex $\Delta_{w}(M(\mathcal{L}), I)$, and
3. for generic $u \in \mathbb{R}^{n}$, the number of regions in $\left(\mathcal{L}^{\perp}+u\right) \backslash\left\{x_{i}=0\right\}_{i \in I}$ whose recession cones have trivial intersection with $\mathbb{R}^{I}=\left\{x \in \mathbb{R}^{n}: x_{j}=0\right.$ for $\left.j \notin I\right\}$.

In Section 2.3 we define the simplicial complex $\Delta_{w}(M(\mathcal{L}), I)$ and show that it satisfies a deletioncontraction relation analogous to that of the broken circuit complex of a matroid. Section 2.4 contains the proof of Theorem 2.2.1. We characterize the strata of $\operatorname{inv}_{I}(\mathcal{L})$ given by its intersection with coordinate subspaces in Section 2.5. Finally, in Section 2.6, we show that for a real linear space $\mathcal{L}, \operatorname{inv}_{I}(\mathcal{L})$ is a hyperbolic variety, in the sense of [56, 90], and prove Theorem 2.2.2.

### 2.3 A Semi-Broken Circuit Complex

Let $M$ be a matroid on elements $[n]$ and suppose $I \subseteq[n]$. A vector $w \in \mathbb{R}^{n}$ with distinct coordinates gives an ordering on $[n]$, where $i<j$ whenever $w_{i}<w_{j}$. Without loss of generality, we can assume $w_{1}<\ldots<w_{n}$, which induces the usual order $1<\ldots<n$. Given a circuit $C$ of $M$ we define an I-broken circuit $M$ to be

$$
b_{I}(C)= \begin{cases}C \backslash \min (C) & \text { if } C \subseteq I \\ (C \cap I) \cup \max (C \backslash I) & \text { if } C \nsubseteq I\end{cases}
$$

Now we define the I-broken circuit complex of $M$ to be

$$
\begin{equation*}
\Delta_{w}(M, I)=\{\tau \subseteq[n]: \tau \text { does not contain an } I \text {-broken circuit of } M\} \tag{2.2}
\end{equation*}
$$

Note that an [ $n$ ]-broken circuit is a broken circuit in the usual sense and $\Delta_{w}(M,[n])$ is the well-studied broken circuit complex of $M$. The $I$-broken circuit complex shares many properties with the classical one.

Theorem 2.3.1. Let $\Delta_{w}(M, I)$ be the I-broken circuit complex defined in (2.2).
(a) If $i \in I$ is a loop of $M$, then $\Delta_{w}(M, I)=\emptyset$.
(b) If $i \in I$ is a coloop of $M$, then $\Delta_{w}(M, I)$ is isomorphic to the $\operatorname{cone}\left(\Delta_{w}(M / i, I \backslash i), i\right)$.
(c) If $i=\max (I)$ is neither a loop nor a coloop of $M$, then

$$
\Delta_{w}(M, I)=\Delta_{w}(M \backslash i, I \backslash i) \cup \operatorname{cone}\left(\Delta_{w}(M / i, I \backslash i), i\right) .
$$

Proof. (a) If $i \in I$ is a loop, then $C=\{i\}$ is a circuit of $M$ with $b_{I}(C)=\emptyset$.
(b) If $i \in I$ is a coloop, then no circuit of $M$, and hence no $I$-broken circuit, contains $i$. The circuits of $M$ are exactly the circuits of the contraction $M / i$ and the $I$-broken circuits of $M$ are the $(I \backslash i)$-broken circuits of $M / i$. Therefore $\tau$ is a face of $\Delta_{w}(M, I)$ if any only if $\tau \backslash i$ is a face of $\Delta_{w}(M / i, I \backslash i)$.
(c) (؟) Let $\tau$ be a face of $\Delta_{w}(M, I)$. We will show that if $i \notin \tau$, then $\tau$ is a face of $\Delta_{w}(M \backslash i, I \backslash i)$ and if $i \in \tau$, then $\tau \backslash i$ is a face of $\Delta_{w}(M / i, I \backslash i)$.

If $i \notin \tau$ and $C$ is a circuit of the deletion $M \backslash i$, then $C$ is a circuit of $M$, and $b_{I}(C)=b_{I \backslash i}(C)$ is an $I$-broken circuit of $M$ and therefore is not contained in $\tau$. If $i \in \tau$ and $C$ is a circuit of the contraction $M / i$, then either $C$ or $C \cup\{i\}$ is a circuit of $M$. In the first case, we again have that $b_{I}(C)=b_{I \backslash i}(C)$ is not contained in $\tau$ and thus not contained in $\tau \backslash i$. Secondly, suppose that $C \cup\{i\}$ is a circuit of $M$. If $C \subseteq I$, then $b_{I}(C \cup\{i\})$ is equal to $C \cup\{i\} \backslash \min (C \cup\{i\})$. Since $i$ is the maximum element of $I$, this equals $C \backslash \min (C) \cup\{i\}$. This set is not contained in $\tau$. Therefore $b_{I}(C)=C \backslash \min (C)$ is not contained in $\tau \backslash i$. If $C \nsubseteq I$, then the $I$-broken circuit of $C \cup\{i\}$ is $(C \cap I) \cup\{i\} \cup \max (C \backslash I)$, which equals $b_{I}(C) \cup\{i\}$. Since $\tau$ cannot contain an $I$-broken circuit of $M, \tau \backslash i$ does not contain $b_{I}(C)$.
(Э) Let $\tau$ be a face of $\Delta_{w}(M \backslash i, I \backslash i)$ and suppose $C$ is a circuit of $M$. If $i \notin C$, then $C$ is also a circuit of $M \backslash i$, meaning that $b_{I}(C)$ is not contained in $\tau$. If $i \in C$ and $C \subseteq I$, then $i=\max (C)$. Since $i$ is not a loop, this implies that $i \in b_{I}(C)$, which cannot be contained in $\tau$. Similarly, if $i \in C$ and $C \not \subset I$, then $i \in b_{I}(C)$ and $b_{I}(C) \not \subset \tau$.

Finally, let $\tau$ be a face of $\Delta_{w}(M / i, I \backslash i)$ and let $C$ be a circuit of $M$. If $i \in C$, then $C \backslash i$ is a circuit of $M / i$. Then $b_{I}(C)$ equals $b_{I \backslash i}(C \backslash i) \cup\{i\}$. Since $\tau$ cannot contain $b_{I}(C \backslash i), \tau \cup\{i\}$ does not contain $b_{I}(C)$. If $i \notin C$, then $C$ is a union of circuits of $M / i$, see [69, §3.1, Exercise 2]. If $C \subseteq I$, then there is a circuit $C^{\prime} \subseteq C$ of $M / i$ containing $\min (C)$. Then $C^{\prime} \subseteq I \backslash i$ and $b_{I \backslash i}\left(C^{\prime}\right)$ is a subset of $b_{I}(C)$. Similarly, if $C \nsubseteq I$, then there is a circuit $C^{\prime} \subseteq C$ of $M / i$ containing $\max (C \backslash I)$, giving $b_{I \backslash i}\left(C^{\prime}\right) \subseteq b_{I}(C)$. In either case, $\tau$ is a face of $\Delta_{w}(M / i, I \backslash i)$ and cannot contain the broken circuit $b_{I \backslash i}\left(C^{\prime}\right)$ and therefore $\tau \cup\{i\}$ cannot contain $b_{I}(C)$.

Corollary 2.3.2. If $M$ is a matroid of rank $d$ with no loops in $I$, then $\Delta_{w}(M, I)$ is a pure simplicial complex of dimension $d-1$.

Proof. We induct on the size of $I$. If $I=\emptyset$, then for every circuit $C$, the broken circuit $b_{I}(C)$ is the maximum element $\max (C)$. In this case, the simplicial complex $\Delta_{w}(M, I)$ consists of one maximal face $B$, where $B$ is the lexicographically smallest basis of $M(\mathcal{L})$. Here $B$ consists of the elements $i \in[n]$ for which the rank of $[i]$ in $M(\mathcal{L})$ is strictly larger than the rank of [i-1]. Every other element is the maximal element of some circuit of $M$.

Now suppose $|I|>0$. If $i \in I$ is a coloop of $M$, then the contraction $M / i$ is a matroid of rank $d-1$ with no loops in $I \backslash i$. Then by induction and Theorem 2.3.1(b), $\Delta_{w}(M, I)=\operatorname{cone}\left(\Delta_{w}(M / i, I \backslash i), i\right)$ is a pure simplicial complex of dimension $d-1$. Finally, suppose $i \in I$ is neither a loop nor a coloop of
$M$. Then the deletion $M \backslash i$ is a matroid of rank $d$ and no element of $I \backslash i$ is a loop of $M \backslash i$. It follows that $\Delta_{w}(M \backslash i, I \backslash i)$ is a pure simplicial complex of dimension $d-1$. The contraction $M / i$ is a matroid of rank $d-1$, meaning that $\Delta_{w}(M / i, I \backslash i)$ is either empty (if $I \backslash i$ contains a loop of $\left.M / i\right)$, or a pure simplicial complex of dimension $d-2$. In either case the decomposition in Theorem 2.3.1 finishes the proof.

We can also see this via connections with the external activity complex defined by Ardila and Boocher [3]. Following their convention, for subsets $S, T \subseteq[n]$, we use $x_{S} y_{T}$ to denote the set $\left\{x_{i}: i \in\right.$ $S\} \cup\left\{y_{j}: j \in T\right\}$.

Definition 2.3.3. [3, Theorem 1.9] Let $M$ be a matroid and suppose $u \in \mathbb{R}^{n}$ has distinct coordinates. Then $u$ induces an order on [ $n$ ] where $i<j$ if and only if $u_{i}<u_{j}$. The external activity complex $B_{u}(M)$ is the simplicial complex on the ground set $\left\{x_{i}, y_{i}: i \in[n]\right\}$ whose minimal non-faces are $\left\{x_{\min _{\wedge_{u}}(C)} y_{C \backslash \min _{<_{u}}(C)}: C \in C\right\}$.

Given a weight vector $w \in\left(\mathbb{R}_{\geq 0}\right)^{n}$ with distinct coordinates, define $u \in \mathbb{R}^{n}$

$$
u_{i}=w_{i} \text { for } i \in I \quad \text { and } \quad u_{j}=-w_{j} \text { for } j \notin I .
$$

With this translation of weights, we can realize the semi-broken circuit complex $\Delta_{w}(M, I)$ as the link of a face in the external activity complex $B_{u}(M)$. Formally, the link of a face $\sigma$ in the simplicial complex $\Delta$ is the simplicial complex

$$
\operatorname{link}_{\Delta}(\sigma)=\{\tau \in \Delta: \tau \cup \sigma \in \Delta \text { and } \tau \cap \sigma=\emptyset\} .
$$

It is the set of faces that are disjoint from $\sigma$ but whose unions with $\sigma$ lie in $\Delta$.
Proposition 2.3.4. Define weight vectors $u, w \in \mathbb{R}^{n}$ as above. If the matroid $M$ has no loops in $I$, then the semi-broken circuit complex $\Delta_{w}(M, I)$ is isomorphic to the link of the face $x_{I} y_{[n] \backslash I}$ in the external activity complex $B_{u}(M)$.

Proof. First we show that $\sigma=x_{I} y_{[n] \backslash I}$ is actually a face of $B_{u}(M)$ by arguing that $\sigma$ does not contain the minimal non-face $x_{\min _{<u}(C)} y_{C \backslash \min _{<u}(C)}$ for any circuit $C$ of $M$. If $C$ is contained in $I$, then so is $C \backslash \min _{<_{u}}(C)$. Since $M$ has no loops in $I$, this is nonempty and we can take $i \in C \backslash \min _{<_{u}}(C)$. Then $y_{i}$ belongs to the non-face $x_{\min _{<u}(C)} y_{C \backslash \min _{<u}(C)}$, but not $\sigma$. On the other hand, if $C$ is not contained in $I$, we know $\min _{<_{u}}(C)$ is contained in the complement of $I$, since the weight vector entries satisfy $u_{j}<u_{i}$ for all $i \in I$ and $j \notin I$. Hence $x_{\min _{<u}(C)}$, an element of the minimal non-face associated to $C$, does not belong to $\sigma$.

Now we argue that $\Delta_{w}(M, I)$ and the link of $\sigma$ in $B_{u}(M)$ are isomorphic by identifying their nonfaces. Note that the link of $\sigma$ is supported on the vertex set $x_{[n] / I} y_{I}$. The bijection of vertices is then just $j \leftrightarrow x_{j}$ for $j \notin I$ and $i \leftrightarrow y_{i}$ for $i \in I$. Note that $\tau \subseteq x_{[n] \backslash I} y_{I}$ is a face of the link of $\sigma$ in $B_{u}(M)$ if and only if for every circuit $C, \tau$ does not contain the intersection of the non-face $x_{\min _{n_{u}}(C)} y_{C \backslash \min _{n_{u}}(C)}$ with $x_{[n] \backslash I} y_{I}$. It suffices to check that these intersections are exactly the $I$-broken circuits of $M$.

If $C$ is contained in $I$, then $w$ and $u$ give the same order on elements of $C$ and $b_{I}(C)$ equals $C \backslash \min _{<_{u}}(C)=C \backslash \min _{<_{w}}(C)$. Since $x_{\min _{<_{u}}(C)} \in \sigma$, we find that

$$
b_{I}(C)=C \backslash \min _{<_{u}}(C) \leftrightarrow y_{C \backslash \min _{<u}(C)}=x_{\min _{<u}(C)} y_{C \backslash \min _{<u}(C)} \backslash \sigma
$$

If $C$ is not contained in $I$, then $b_{I}(C)$ equals $(C \cap I) \cup \max _{<_{w}}(C \backslash I)$. Since $u$ reverses the order on $[n] \backslash I$, this equals $(C \cap I) \cup \min _{<_{u}}(C \backslash I)$. Then

$$
b_{I}(C)=(C \cap I) \cup \min _{<u}(C \backslash I) \leftrightarrow x_{\min _{<u}(C \backslash I)} y_{C \cap I}=x_{\min _{<u}(C)} y_{C \backslash \min _{<u}(C)} \backslash \sigma,
$$

where the equality $x_{\min _{<u}(C \backslash I)}=x_{\min _{<_{u}}(C)}$ holds because $u_{j}<u_{i}$ for all $i \in I$ and $j \notin I$. This shows that under this bijection of the vertices, the semi-broken circuit complex equals the link of $\sigma$ in the external activity complex.

Corollary 2.3.5. The semi-broken circuit complex is shellable.
Proof. In [4], Ardila, Castillo, and Sampler show that the external activity complex, $B_{u}(M)$, is shellable. Then by [14, Prop. 10.14], the link of any face in $B_{u}(M)$ is also shellable.

Example 2.3.6. Let $M$ be the matroid from Example 2.4.3, $I=\{1,2,3\}$, and $u$ be the weight vector associated to $w$ as described above. It induces the linear order $5<4<1<2<3$ on the ground set of the matroid $M(\mathcal{L})$.

We outline the connection between the external activity complex $B_{u}(M)$ and the semi-broken circuit complex by tracking two bases $B_{1}=\{1,3,4\}, B_{2}=\{2,3,5\}$ of the matroid $M(\mathcal{L})$ in the construction of the two simplicial complexes. For each basis, we split the complement [5] $\backslash B_{i}$ into externally active and externally passive elements. (See [3, §2.5] for the definitions of externally active and passive.) For $B_{1},\{2\}$ is externally passive and $\{5\}$ is externally active. Then by [3, Theorem 5.1], the associated facet of $B_{u}(M)$ is $F_{1}=x_{1} x_{2} x_{3} x_{4} y_{1} y_{3} y_{4} y_{5}$. By deleting $\sigma=x_{1} x_{2} x_{3} y_{4} y_{5}$ from $F_{1}$, we obtain the facet $x_{4} y_{1} y_{3}$ of link ${ }_{\Delta}(\sigma)$, corresponding to the facet $\{1,3,4\}$ of $\Delta_{w}(M, I)$. For $B_{2}=\{2,3,5\}$, the externally passive elements are the entire complement $\{1,4\}$, hence the associated facet of $B_{u}(M)$ is $F_{2}=x_{1} x_{2} x_{3} x_{4} x_{5} y_{2} y_{3} y_{5}$. Since $F_{2}$ does not contain $\sigma$, it does not contribute a facet to the link of $\sigma$ in $B_{u}(M)$.

The connection between this simplicial complex and the semi-inverted linear space $\operatorname{inv}_{I}(\mathcal{L})$ is that when $w \in\left(\mathbb{R}_{+}\right)^{n}$ has distinct coordinates, the ideal generated by the initial forms

$$
\left\{\operatorname{in}_{w}\left(f_{C}\right): C \text { is a circuit of } M(\mathcal{L})\right\}
$$

is the Stanley-Reisner ideal of $\Delta_{w}(M, I)$. In fact, the initial form of $f_{C}$ is $\operatorname{in}_{w}\left(f_{C}\right)=\mathbf{x}^{b_{I}(C)}$. The ideal generated by these initial forms is then the Stanley-Reisner ideal $I_{\Delta_{w}(M, I)}=\left\langle\mathrm{in}_{w}\left(f_{C}\right): C \in C(M)\right\rangle$.

### 2.4 Proof of the Main Theorem

In this section, we prove Theorem 2.2.1. To do this, we first use a flat degeneration of $\operatorname{inv}_{I}(\mathcal{L})$ to establish a recursion for its degree.

Proposition 2.4.1. Suppose $\mathcal{L}$ is a linear subspace of $\mathbb{C}^{n}$ and $I \subseteq[n]$. Let $D(\mathcal{L}, I)$ denote the degree of the affine variety $\operatorname{inv}_{I}(\mathcal{L})$. Then $D(\mathcal{L}, I)$ satisfies the following recurrence:
(a) If $i \in I$ is a loop of $M(\mathcal{L})$, then $\operatorname{inv}_{I}(\mathcal{L})$ is empty and $D(\mathcal{L}, I)=0$.
(b) If $i \in I$ is a co-loop of $M(\mathcal{L})$, then $D(\mathcal{L}, I)=D(\mathcal{L} / i, I \backslash i)$.
(c) If $i \in I$ is neither a loop nor a coloop of $M(\mathcal{L})$ then

$$
D(\mathcal{L} \backslash i, I \backslash i)+D(\mathcal{L} / i, I \backslash i) \leq D(\mathcal{L}, I) .
$$

The proof of Theorem 2.2.1 will show that there is actually equality in part (c).
Proof. Without loss of generality, take $i=1$.
(a) If $1 \in I$ is a loop of $M(\mathcal{L})$ then $\mathcal{L}$ is contained in the hyperplane $\left\{x_{1}=0\right\}$. Therefore the map $\operatorname{inv}_{I}$ is undefined at every point of $\mathcal{L}$ and the $\operatorname{image}_{\operatorname{inv}}^{I}(\mathcal{L})$ is empty. By convention, we take the degree of the empty variety to be zero.
(b) If 1 is a co-loop of $M(\mathcal{L})$, then $\mathcal{L}$ is a direct sum of $\operatorname{span}\left(e_{1}\right)$ and $\mathcal{L} / 1$, meaning that any element in $\mathcal{L}$ can be written as $a e_{1}+b$ where $a \in \mathbb{C}$ and $b \in \mathcal{L} / 1$. For points at which the map $\operatorname{inv}_{I}$ is defined, $\operatorname{inv}_{I}\left(a e_{1}+b\right)=a^{-1} e_{1}+\operatorname{inv}_{I \backslash 1}(b)$. From this, we see that $\operatorname{inv}_{I}(\mathcal{L})$ is the direct sum of $\operatorname{span}\left(e_{1}\right)$ and $\operatorname{inv}_{I \backslash 1}(\mathcal{L} / 1)$, meaning that $\operatorname{inv}_{I}(\mathcal{L})$ and $\operatorname{inv}_{I \backslash 1}(\mathcal{L} / 1)$ have the same degree.
(c) Let $\mathcal{I}$ denote the ideal of polynomials vanishing on $\operatorname{inv}_{I}(\mathcal{L})$ and $\mathcal{J}=\overline{\mathcal{I}}$ denote its homogenization in $\mathbb{C}\left[x_{0}, x_{1}, \ldots, x_{n}\right]$. Take $w=e_{1} \in \mathbb{R}^{n+1}$ and consider $\operatorname{in}_{w}(\mathcal{J})$, as defined in Section 1.3.3, We will show that the variety of in ${ }_{w}(\mathcal{J})$ contains the image in $\mathbb{P}^{n}$ of both $\{0\} \times \operatorname{inv}_{I \backslash 1}(\mathcal{L} \backslash 1)$ and $\mathbb{A}^{1}(\mathbb{C}) \times \operatorname{inv}_{I \backslash 1}(\mathcal{L} / 1)$. Since both these varieties have dimension equal to $\operatorname{dim}(\mathcal{L})$, the degree of the variety $\operatorname{of~}^{\operatorname{in}_{w}(\mathcal{J}) \text { is at }}$ least the sum of their degrees. The claim then follows by the equality of the Hilbert series of $\mathcal{J}$ and $\mathrm{in}_{w}(\mathcal{J})$.

If $j \in I \backslash 1$ is a loop of $M(\mathcal{L})$, then $j$ is a loop of $M(\mathcal{L} \backslash 1)$ and $D(\mathcal{L} \backslash 1, I \backslash 1)=0$. Otherwise the set $U_{I}$ is Zariski-dense in $\mathcal{L}$, where $U_{I}$ denotes the intersection of $\mathcal{L}$ with $\left(\mathbb{C}^{*}\right)^{I} \times \mathbb{C}^{[n] \backslash I}$.

Let $\pi_{I}$ denote the coordinate projection $\mathbb{C}^{n} \rightarrow \mathbb{C}^{I}$. On $U_{I}$, the maps $\pi_{I \backslash 1} \circ \operatorname{inv}_{I}$ and $\operatorname{inv}_{I \backslash 1} \circ \pi_{I \backslash 1}$ are equal:

$$
\pi_{I \backslash 1}\left(\operatorname{inv}_{I}(x)\right)=\operatorname{inv}_{I \backslash 1}\left(\pi_{I \backslash 1}(x)\right)=\sum_{j \in I \backslash 1} x_{j}^{-1} e_{j}+\sum_{j \neq I} x_{j} e_{j} .
$$

In particular, the points $\operatorname{inv}_{I \backslash 1}\left(\pi_{I \backslash 1}\left(U_{I}\right)\right)$ are Zariski dense in $\operatorname{inv}_{I \backslash 1}(\mathcal{L} \backslash 1)$. Now let $x$ be a point of $\operatorname{inv}_{I}\left(U_{I}\right)$. Then $[1: x]$ belongs to the variety of $\mathcal{J}$ and, for every $t \in \mathbb{C}$, the point $\left(t, t^{e_{1}} \cdot[1: x]\right)$ belongs to the variety of $\overline{\mathcal{J}}^{w}$, as defined in Section 1.3.3. Taking $t \rightarrow 0$, we see that $\left[1: 0: \pi_{I \backslash 1}(x)\right]$ belongs to the variety of $\mathrm{in}_{e_{1}}(\mathcal{J})$.

If $j \in I \backslash 1$ is a loop of $M(\mathcal{L} / 1)$, then $\operatorname{inv}_{I \backslash 1}(\mathcal{L} / 1)$ is empty and the claim follows. Otherwise the intersection $U_{I \backslash 1}$ of $\mathcal{L} / 1$ with $\{0\} \times\left(\mathbb{C}^{*}\right)^{I \backslash 1} \times \mathbb{C}^{[n] \backslash I}$ is nonempty and Zariski-dense in $\mathcal{L} \cap\left\{x \in \mathbb{C}^{n}\right.$ : $\left.x_{1}=0\right\} \cong \mathcal{L} / 1$. Let $x \in U_{I \backslash 1}$. Since 1 is not a loop of $M(\mathcal{L})$, there is a point $v \in \mathcal{L}$ with $v_{1}=1$. Then for any $\lambda, t \in \mathbb{C}^{*}, x+(t / \lambda) v$ belongs to $\mathcal{L}$ and for all but finitely many values of $t, y(t)=\operatorname{inv}_{I}(x+(t / \lambda) v)$ is defined and has first coordinate $y_{1}(t)=\lambda / t$. Then $[1: y(t)] \in \mathcal{V}(\mathcal{J})$ and $\left(t, t^{e_{1}} \cdot[1: y(t)]\right)$ belongs to $\mathcal{V}\left(\overline{\mathcal{J}}^{w}\right)$. Note that the limit of $t^{e_{1}} \cdot[1: y(t)]=\left[1: \lambda: y_{2}(t): \ldots: y_{n}(t)\right]$ as $t \rightarrow 0$ equals $\left[1: \lambda: \operatorname{inv}_{I \backslash 1}(x)\right]$. Therefore for every point $(\lambda, u) \in \mathbb{A}^{1}(\mathbb{C}) \times \operatorname{inv}_{I \backslash 1}(\mathcal{L} / 1)$, the point $[1: \lambda: u]$ belongs to $\mathcal{V}\left(\operatorname{in}_{e_{1}}(\mathcal{J})\right)$.

Lemma 2.4.2. Let $I \subseteq J \subseteq \mathbb{C}[\mathbf{x}]$ be equidimensional homogeneous ideals of dimension $d$. If $I$ is radical and $\operatorname{deg}(I) \leq \operatorname{deg}(J)$, then $I$ and $J$ are equal.

Proof. Let $I=P_{1} \cap \ldots \cap P_{r}$ and $J=Q_{1} \cap \ldots \cap Q_{s}$ be irredundant primary decompositions of $I$ and $J$. Without loss of generality, we can assume that $\operatorname{dim}\left(Q_{i}\right)=d$ for $1 \leq i \leq u$, and since $\mathcal{V}(J) \subseteq V(I)$, the prime ideals $P_{i}$ can be reindexed such that $P_{i}=\sqrt{Q_{i}}$, meaning $Q_{i} \subseteq P_{i}$. For all $1 \leq i \leq u$, there exists an element $a \in\left(\cap_{j \neq i} P_{j}\right) \cap\left(\cap_{j \neq i} \sqrt{Q_{j}}\right)$ with $a \notin P_{i}$. Then the saturation $I:\langle a\rangle^{\infty}=P_{i}$ is contained in $J:\langle a\rangle^{\infty}=Q_{i}$, implying $P_{i}=Q_{i}$. This writes the ideal $J$ as $J=P_{1} \cap \ldots \cap P_{u} \cap Q_{u+1} \cap \ldots \cap Q_{s}$. The degree of an ideal is equal to the sum of the degrees of the top dimensional ideals in its primary decomposition, hence

$$
\operatorname{deg}(I)=\sum_{i=1}^{r} \operatorname{deg}\left(P_{i}\right) \quad \text { and } \quad \operatorname{deg}(J)=\sum_{i=1}^{u} \operatorname{deg}\left(Q_{i}\right)=\sum_{i=1}^{u} \operatorname{deg}\left(P_{i}\right)
$$

The assumption that $\operatorname{deg}(I) \leq \operatorname{deg}(J)$ implies that $r=u$, which gives the reverse containment $I=P_{1} \cap \ldots \cap P_{u} \supseteq J$.

Proof of Theorem 2.2.1. We proceed by induction on $|I|$. If $|I|=0$, then $\operatorname{inv}_{I}(\mathcal{L})$ is just the linearspace $\mathcal{L}$. Then Theorem 2.2 .1 reduces to the statement that the linear forms supported on circuits form a universal Gröbner basis for $\mathcal{I}(\mathcal{L})$. See e.g. [95, Prop. 1.6].

Now take $|I| \geq 1, w \in\left(\mathbb{R}_{+}\right)^{n}$ with distinct coordinates, and let $M$ denote the matroid $M(\mathcal{L})$. If $M$ has a loop $i$ in $I$, then for the circuit $C=\{i\}$, the circuit polynomial $f_{C}$ equals 1 , which is a Gröbner basis for the ideal of polynomials vanishing on the empty set $\operatorname{inv}_{I}(\mathcal{L})$. Therefore we may suppose that $M$ has no loops in $I$, in which case $\operatorname{inv}_{I}(\mathcal{L})$ is a $d$-dimensional affine variety of degree $D(\mathcal{L}, I)$.

Let $\Delta$ denote the $I$-broken circuit complex $\Delta_{w}(M, I)$ defined in Section 2.3 and let $\Delta_{0}$ denote the simplicial complex on elements $\{0, \ldots, n\}$ obtained from $\Delta$ by coning over the vertex 0 . Let $I_{\Delta_{0}}$ denote the Stanley-Reisner ideal of $\Delta_{0}$, as in Section 2.1.4.

Let $\mathcal{I} \subset \mathbb{C}[\mathbf{x}]$ be the ideal of polynomials vanishing on $\operatorname{inv}_{I}(\mathcal{L})$ and let $\mathcal{J} \subset \mathbb{C}\left[x_{0}, x_{1}, \ldots, x_{n}\right]$ be its homogenization with respect to $x_{0}$. For a circuit polynomial $f_{C}$, its homogenization $\overline{f_{C}}$ belongs to $\mathcal{J}$ and since $w \in\left(\mathbb{R}_{+}\right)^{n}$,

$$
\operatorname{in}_{(0, w)}\left(\overline{f_{C}}\right)=\operatorname{in}_{w}\left(f_{C}\right)= \begin{cases}a_{k} \mathbf{x}^{C \backslash k} & \text { if } C \subseteq I \text { and } k=\operatorname{argmin}\left\{w_{j}: j \in C\right\} \\ a_{k} \mathbf{x}^{C \cap I \cup k} & \text { if } C \nsubseteq I \text { and } k=\operatorname{argmax}\left\{w_{j}: j \in C \backslash I\right\}\end{cases}
$$

Up to a scalar multiple, $\operatorname{in}_{w}\left(f_{C}\right)$ equals the square-free monomial corresponding to the $I$-broken circuit of $C$, namely $\mathbf{x}^{b_{I}(C)}$. It follows that

$$
\left\langle\operatorname{in}_{w}\left(f_{C}\right): C \in \mathcal{C}(M)\right\rangle=I_{\Delta} \text { and }\left\langle\operatorname{in}_{(0, w)}\left(\overline{f_{C}}\right): C \in \mathcal{C}(M)\right\rangle=I_{\Delta_{0}}
$$

From this we see that $I_{\Delta_{0}} \subseteq \operatorname{in}_{(0, w)}(\mathcal{J})$.
Let $i=\operatorname{argmax}\left\{w_{j}: j \in I\right\}$. By the inductive hypothesis, $D(\mathcal{L} \backslash i, I \backslash i)$ and $D(\mathcal{L} / i, I \backslash i)$ are the number of facets of $\Delta_{w}(M \backslash i, I \backslash i)$ and $\Delta_{w}(M / i, I \backslash i)$, respectively. Therefore by Theorem 2.3.1, $\Delta$ and thus $\Delta_{0}$ each have $D(\mathcal{L} / i, I \backslash i)$ facets if $i$ is a coloop of $M$ and $D(\mathcal{L} \backslash i, I \backslash i)+D(\mathcal{L} / i, I \backslash i)$ facets otherwise. By Proposition 2.4.1, it follows that $\Delta_{0}$ has at most $D(\mathcal{L}, I)$ facets and that the Stanley-Reisner ideal $I_{\Delta_{0}}$ has degree $\leq D(\mathcal{L}, I)$.

Since $\Delta_{0}$ is a pure simplicial complex of dimension $d, I_{\Delta_{0}}$ is an equidimensional ideal of dimension $d$. The ideal $\mathcal{J}$ is a prime ideal of dimension $d$, meaning that its initial ideal $\operatorname{in}_{(0, w)}(\mathcal{J})$ is equidimensional of the same dimension, see [60, Prop. 2.4.2].

The ideals $\mathcal{I}_{\Delta_{0}}$ and $\operatorname{in}_{(0, w)}(\mathcal{J})$ then satisfy the hypotheses of Lemma 2.4.2, and we conclude that they are equal. By [60, Prop. 2.6.1], restricting to $x_{0}=1$ gives that

$$
I_{\Delta}=\left\langle\operatorname{in}_{w}\left(f_{C}\right): C \in C(M)\right\rangle=\operatorname{in}_{w}(\mathcal{I})
$$

As this holds for every $w \in\left(\mathbb{R}_{+}\right)^{n}$ with distinct coordinates, it will also hold for arbitrary $w \in\left(\mathbb{R}_{\geq 0}\right)^{n}$ (see [95, Prop. 1.13]). It follows from [95, Cor. 1.9, 1.10] that the circuit polynomials $\left\{f_{C}: C \in C(M)\right\}$ form a universal Gröbner basis for the ideal $\mathcal{I}$.

Example 2.4.3. Consider the 3-dimensional linear space in $\mathbb{C}^{5}$ :

$$
\mathcal{L}=\operatorname{rowspan}\left(\begin{array}{lllll}
1 & 0 & 0 & 1 & 1 \\
0 & 1 & 0 & 1 & 0 \\
0 & 0 & 1 & 0 & 1
\end{array}\right)
$$

The circuits of the matroid $M(\mathcal{L})$ are $C=\{124,135,2345\}$. Take $I=\{1,2,3\}$. Then

$$
f_{124}=x_{1}+x_{2}-x_{1} x_{2} x_{4}, \quad f_{135}=x_{1}+x_{3}-x_{1} x_{3} x_{5}, \quad \text { and } f_{2345}=x_{2}-x_{3}+x_{2} x_{3} x_{4}-x_{2} x_{3} x_{5}
$$

If $w \in\left(\mathbb{R}_{+}\right)^{5}$ with $w_{1}<\ldots<w_{5}$, then the ideal $\left\langle\operatorname{in}_{w}\left(f_{C}\right): C \in C\right\rangle$ is $\left\langle x_{1} x_{2} x_{4}, x_{1} x_{3} x_{5}, x_{2} x_{3} x_{5}\right\rangle$. The corresponding simplicial complex $\Delta_{w}(M, I)$ is 2-dimensional and has seven facets:

$$
\operatorname{facets}\left(\Delta_{w}(M, I)\right)=\{123,125,134,145,234,245,345\}
$$

Indeed, the variety of $\left\langle x_{1} x_{2} x_{4}, x_{1} x_{3} x_{5}, x_{2} x_{3} x_{5}\right\rangle$ is the union the seven coordinate linear spaces span $\left\{e_{i}, e_{j}, e_{k}\right\}$ where $\{i, j, k\}$ is a facet of $\Delta_{w}(M, I)$.

Interestingly, it is not true that the homogenizations $\overline{f_{C}}$ form a universal Gröbner basis for the homogenization $\overline{\mathcal{I}}$. Indeed, consider the weight vector $(2,0,0,1,1,1)$. The ideal generated by the
initial forms of circuit polynomials $\left\langle\operatorname{in}_{w}\left(\overline{f_{C}}\right): C \in C\right\rangle$ is $\left\langle x_{0}^{2} x_{1}+x_{0}^{2} x_{2}, x_{0}^{2} x_{3}\right\rangle$, whereas $\mathrm{in}_{w}(\overline{\mathcal{I}})=$ $\left\langle x_{2} x_{3} x_{4}-x_{1} x_{3} x_{5}-x_{2} x_{3} x_{5}, x_{0}^{2} x_{1}+x_{0}^{2} x_{2}, x_{0}^{2} x_{3}\right\rangle$. Nevertheless, upon restriction to $x_{0}=1$, the two ideals become equal.

Corollary 2.4.4. If $\operatorname{dim}(\mathcal{L})=d$, then the affine Hilbert series of the ideal $\mathcal{I} \subseteq \mathbb{C}[\mathbf{x}]$ of polynomials vanishing on $\operatorname{inv}_{I}(\mathcal{L})$ is

$$
\sum_{m=0}^{\infty} \operatorname{dim}_{\mathbb{C}}\left(\mathbb{C}[\mathbf{x}]_{\leq m} / I_{\leq m}\right) t^{m}=\frac{1}{(1-t)^{d+1}} \sum_{i=0}^{d} f_{i-1} t^{i}(1-t)^{d-i}=\frac{h_{0}+h_{1} t+\ldots+h_{d} t^{d}}{(1-t)^{d+1}}
$$

where $\left(f_{-1}, \ldots, f_{d-1}\right)$ and $\left(h_{0}, \ldots, h_{d}\right)$ are the $f$ and $h$ vectors of $\Delta_{w}(M, I)$. In particular, its degree is the number of facets $f_{d-1}=h_{0}+h_{1}+\ldots+h_{d}$.

Proof. The affine Hilbert series of $\mathcal{I}$ equals the classical Hilbert series of its homogenization $\overline{\mathcal{I}}$, which equals the Hilbert series of $\operatorname{in}_{(0, w)}(\overline{\mathcal{I}})$ for any $w \in \mathbb{R}^{n}$. When the coordinates of $w$ are distinct and positive, $\operatorname{in}_{(0, w)}(\overline{\mathcal{I}})$ is the Stanley-Reisner ideal of $\Delta_{0}=\operatorname{cone}\left(\Delta_{w}(M, I), 0\right)$. Since the Stanley-Reisner ideals of $\Delta=\Delta_{w}(M, I)$ and $\Delta_{0}$ are generated by the same square-free monomials, their Hilbert series differ by a factor of $1 /(1-t)$. The result then follows from well known formulas for the Hilbert series of $I_{\Delta}$, [68, Ch. 1].

Theorem 2.2.1 shows that there is equality in Proposition 2.4.1(c), namely that if $i \in I$ is neither a loop nor a coloop of $M(\mathcal{L})$, then $D(\mathcal{L}, I)=D(\mathcal{L} \backslash i, I \backslash i)+D(\mathcal{L} / i, I \backslash i)$. From this we can derive an explicit formula for the degree of $\operatorname{inv}_{I}(\mathcal{L})$ in the uniform matroid case.

Corollary 2.4.5. For a generic d-dimensional linear space $\mathcal{L} \subseteq \mathbb{C}^{n}$ and $I \subseteq[n]$ of size $|I|=k$, the degree of $\operatorname{inv}_{I}(\mathcal{L})$ equals

$$
D(\mathcal{L}, I)=\sum_{j=k+d-n}^{d}\binom{k}{j}-\binom{k-1}{d}
$$

where we take $\binom{a}{b}=0$ whenever $a<0$ or $b<0$. In particular, for $n \geq k+d$, the degree only depends on $d$ and $k$.

Proof. By assumption $k, d, n$ satisfy the inequalities $0 \leq k \leq n$ and $0 \leq d \leq n$. We proceed by induction on $k$. In the extremal cases, $D(\mathcal{L}, I)$ satisfies

$$
D(\mathcal{L}, I)= \begin{cases}1 & \text { if } k=0 \\ 1 & \text { if } d=n \\ 0 & \text { if } d=0 \text { and } k \geq 1\end{cases}
$$

Indeed, if $I=\emptyset$, then $\operatorname{inv}_{I}(\mathcal{L})=\mathcal{L}$ and $D(\mathcal{L}, I)=1$. If $d=n$, then $\operatorname{inv}_{I}(\mathcal{L})$ is all of $\mathbb{C}^{n}$ and $D(\mathcal{L}, I)=1$. Finally, if $d=0$ and $|I| \geq 1$, then $n \geq|I| \geq 1$, and $\mathcal{L}=\{(0, \ldots, 0)\}$ in $\mathbb{C}^{n}$. The map $\operatorname{inv}_{I}$ is not defined at this point so $\operatorname{inv}_{I}(\mathcal{L})$ is empty and thus has degree 0 .

Suppose $k \geq 1$ and $0<d<n$. Any $i \in I$ is neither a loop nor a coloop, so Proposition 2.4.1(c) and the proof of Theorem 2.2 .1 imply that $D(\mathcal{L}, I)=D(\mathcal{L} \backslash i, I \backslash i)+D(\mathcal{L} / i, I \backslash i)$. Recall that $\mathcal{L} \backslash i$ and $\mathcal{L} / i$
are subspaces in $\mathbb{C}^{n-1}$ of dimensions $d$ and $d-1$, respectively. Since $|I \backslash i|=k-1$, by induction we get that

$$
D(\mathcal{L} \backslash i, I \backslash i)=\sum_{j=k+d-n}^{d}\binom{k-1}{j}-\binom{k-2}{d} \text { and } D(\mathcal{L} / i, I \backslash i)=\sum_{j=k+d-n-1}^{d-1}\binom{k-1}{j}-\binom{k-2}{d-1} .
$$

Since $\binom{k-1}{j}+\binom{k-1}{j-1}=\binom{k}{j}$ and $\binom{k-2}{d}+\binom{k-2}{d-1}=\binom{k-1}{d}$, the sum $D(\mathcal{L} \backslash i, I \backslash i)+D(\mathcal{L} / i, I \backslash i)$ is the desired formula for $D(\mathcal{L}, I)$.

Example 2.4.6. The number of facets of the complex $\Delta_{w}(M, I)$ gives the degree $D(\mathcal{L}, I)$ and if $M$ is the uniform matroid of rank $d$ on $[n]$, we can write out these facets explicitly. Let $w=(1,2 \ldots, n)$ and consider $I=\{1,2, \ldots, k\}$. If $k \leq d$, no circuit is contained in the inverted set $I$, meaning that every broken circuit has the form $(C \cap I) \cup \max \{C \backslash I\}$. Then every facet of $\Delta_{w}(M, I)$ has the form $S \cup\{k+1, \ldots, k+d-j\}$ where $S \subseteq I$ and $|S|=j \leq d$. For fixed $j$, the number of possibilities are $\binom{k}{j}$, and the constraints on $j$ are $k+d-j \leq n$ and $0 \leq j \leq k \leq d$. If $k>d$, then every subset of $\{2, \ldots, k\}$ of size $d$ is a semi-broken circuit. From the list of facets $S \cup\{k+1, \ldots, k+d-j\}$, we need to remove those for which $S \subset\{2, \ldots, k\}$ and $|S|=d$, of which there are $\binom{k-1}{d}$.

### 2.5 Supports

In this section, we characterize the intersection of the variety $\operatorname{inv}_{I}(\mathcal{L})$ with the coordinate hyperplanes. These are exactly the points in the closure of, but not the actual image of, the map inv ${ }_{I}$. Given a point $\mathbf{p} \in \mathbb{C}^{n}$, its support is the set of indices of its nonzero coordinates: $\operatorname{supp}(\mathbf{p})=\left\{i: p_{i} \neq 0\right\}$. For a subset $S \subseteq[n]$, we will use $\mathbb{C}^{S}$ to denote the set of points $\mathbf{p}$ with $\operatorname{supp}(\mathbf{p}) \subseteq S$ and $\bar{S}$ to denote the complement $[n] \backslash S$.

Theorem 2.5.1. Suppose that the matroid $M=M(\mathcal{L})$ has no loops in I. For $S \subseteq[n]$, let $T=S \cup \bar{I}$. If $T$ is a flat of $M$, then the restriction of $\operatorname{inv}_{I}(\mathcal{L})$ to $\mathbb{C}^{S}$ is given by

$$
\operatorname{inv}_{I}(\mathcal{L}) \cap \mathbb{C}^{S}=\operatorname{inv}_{S \cap I}\left(\pi_{T}(\mathcal{L}) \cap \mathbb{C}^{S}\right)
$$

where $\pi_{T}$ denotes the coordinate projection $\mathbb{C}^{n} \rightarrow \mathbb{C}^{T}$. Moreover, there exists $\mathbf{p} \in \operatorname{inv}_{I}(\mathcal{L})$ with $\operatorname{supp}(\mathbf{p})=S$ if and only if $T$ is a flat of $M$ and $T \backslash S$ is a flat of $\left.M\right|_{T}$.

We build up to the proof of Theorem 2.5.1 by considering the cases $\bar{I} \subseteq S$ and $I \subseteq S$.
Lemma 2.5.2. If $S \subseteq[n]$ is a flat of $M$ with $\bar{I} \subseteq S$, then

$$
\operatorname{inv}_{I}(\mathcal{L}) \cap \mathbb{C}^{S}=\operatorname{inv}_{S \cap I}\left(\pi_{S}(\mathcal{L})\right)
$$

where $\pi_{S}$ denotes the coordinate projection $\mathbb{C}^{n} \rightarrow \mathbb{C}^{S}$.

Proof. Recall that $F \subset[n]$ is a flat of $M$ if and only if $|\bar{F} \cap C| \neq 1$ for all circuits $C$ of $M$. Suppose that $S$ is a flat of $M$ and consider the restriction of the circuit polynomials $f_{C}$ to $\mathbb{C}^{S}$. Note that $\bar{S} \subseteq I$, so that for any circuit $C$ with $|C \cap \bar{S}| \geq 2,|C \cap I| \geq 2$ and the circuit polynomial $f_{C}$ is zero at every point of $\mathbb{C}^{S}$.

The circuits for which $|C \cap \bar{S}|=0$ are exactly the circuits contained in $S$, which are the circuits of the matroid restriction $\left.M\right|_{S}$. Moreover the projection $\pi_{S}(\mathcal{L})$ is cut out by the vanishing of the linear forms $\left\{\ell_{C}: C \in \mathcal{C}(M), C \subseteq S\right\}$, which are exactly the linear forms $\left\{\ell_{C^{\prime}}: C^{\prime} \in \mathcal{C}(M \mid s)\right\}$. It follows that the circuit polynomials $\left\{f_{C^{\prime}}: C^{\prime} \in \mathcal{C}\left(\left.M\right|_{S}\right)\right\}$ are a subset of a circuit polynomials of $\mathcal{L}$, namely $\left\{f_{C}: C \in \mathcal{C}(M), C \subseteq S\right\}$. By Theorem 2.2.1, the variety of circuit polynomials is the variety of the semi-inverted linear space, giving that

$$
\operatorname{inv}_{I}(\mathcal{L}) \cap \mathbb{C}^{S}=\mathcal{V}\left(\left\{f_{C}: C \in \mathcal{C}(M), C \subseteq S\right\}\right) \cap \mathbb{C}^{S}=\mathcal{V}\left(\left\{f_{C^{\prime}}: C^{\prime} \in \mathcal{C}\left(\left.M\right|_{S}\right)\right\}\right)=\operatorname{inv}_{S \cap I}\left(\pi_{S}(\mathcal{L})\right)
$$

Lemma 2.5.3. If $S \subseteq[n]$ with $I \subseteq S$, then $\operatorname{inv}_{I}(\mathcal{L}) \cap \mathbb{C}^{S}=\operatorname{inv}_{I}\left(\mathcal{L} \cap \mathbb{C}^{S}\right)$.
Proof. (〇) The affine variety $\operatorname{inv}_{I}\left(\mathcal{L} \cap \mathbb{C}^{S}\right)$ is the Zariski-closure of $\mathcal{L} \cap \mathbb{C}^{S}$ under the map inv ${ }_{I}$. Since
 $\mathbf{p} \in \mathbb{C}^{S}$. The inclusion follows.
$(\subseteq)$ For this we show the reverse inclusion of the ideals of polynomials vanishing on these varieties. Now let $C^{\prime}$ be a circuit $\mathcal{L} \cap \mathbb{C}^{S}$ and $\ell_{C^{\prime}}=\sum_{i \in C^{\prime}} a_{i} x_{i}$ its corresponding linear form. Then for some circuit $C$ of $M, C^{\prime}=C \cap S$ and $\ell_{C^{\prime}}$ equals the restriction $\ell_{C}\left(\pi_{S}(\mathbf{x})\right)$. Applying inv ${ }_{I}$ and clearing denominators then gives

$$
f_{C^{\prime}}(\mathbf{x})=\mathbf{x}^{C^{\prime} \cap I} \ell_{C^{\prime}}\left(\operatorname{inv}_{I}(\mathbf{x})\right)=\mathbf{x}^{C \cap I} \ell_{C}\left(\operatorname{inv}_{I}\left(\pi_{S}(\mathbf{x})\right)\right)=f_{C}\left(\pi_{S}(\mathbf{x})\right) .
$$

The middle equation holds because $I \subseteq S$, meaning that $C \backslash C^{\prime} \subseteq \bar{S} \subseteq \bar{I}$.
Proof of Theorem 2.5.1. Suppose that $T$ is a flat of $M$. Since $\bar{I} \subseteq T$, Lemma 2.5.2 says that the restriction $\left.\operatorname{inv}_{I}(\mathcal{L})\right|_{\mathbb{C}^{T}}$ equals $\operatorname{inv}_{T \cap I}\left(\pi_{T}(\mathcal{L})\right)$. Furthermore since $T \cap I=S \cap I \subseteq S$, we can apply Lemma 2.5.3 to find the intersection of $\operatorname{inv}_{T \cap I}\left(\pi_{T}(\mathcal{L})\right)$ with $\mathbb{C}^{S}$. All together this gives

$$
\begin{equation*}
\operatorname{inv}_{I}(\mathcal{L}) \cap \mathbb{C}^{S}=\left(\operatorname{inv}_{I}(\mathcal{L}) \cap \mathbb{C}^{T}\right) \cap \mathbb{C}^{S}=\operatorname{inv}_{S \cap I}\left(\pi_{T}(\mathcal{L})\right) \cap \mathbb{C}^{S}=\operatorname{inv}_{S \cap I}\left(\pi_{T}(\mathcal{L}) \cap \mathbb{C}^{S}\right) \tag{2.3}
\end{equation*}
$$

Suppose further that $T \backslash S$ is a flat of the matroid $\left.M\right|_{T}$. This implies that the contraction of $\left.M\right|_{T}$ by $T \backslash S$ has no loops. This is the matroid of the linearspace $\pi_{T}(\mathcal{L}) \cap \mathbb{C}^{S}$, which is therefore not contained in any coordinate subspace $\left\{x_{i}=0\right\}$ for $i \in S$. It follows that there is a point $\mathbf{p} \in \pi_{T}(\mathcal{L}) \cap \mathbb{C}^{S}$ of full $\operatorname{support} \operatorname{supp}(\mathbf{p})=S$. Equation (2.3) then shows that $\operatorname{inv}_{S_{S_{I}}(\mathbf{p})}$ is a point of $\operatorname{support} \operatorname{in}^{\operatorname{inv}}{ }_{I}(\mathcal{L})$.

Conversely, suppose that $S=\operatorname{supp}(\mathbf{p})$ for some point $\mathbf{p} \in \operatorname{inv}_{I}(\mathcal{L})$. Then $T$ is a flat of $M$. To see this, suppose for the sake of contradiction that for some circuit $C$ of $M, C \cap \bar{T}=\{j\}$. Then $j$ is the unique
element of $C \cap I$ for which $p_{j}=0$, and evaluating the circuit polynomial $f_{C}$ at the point $\mathbf{p}$ gives

$$
f_{C}(\mathbf{p})=\sum_{i \in C \cap I} a_{i} \mathbf{p}^{C \cap I \backslash\{i\}}+\sum_{i \in C \backslash I} a_{i} \mathbf{p}^{C \cap I \cup\{i\}}=a_{j} \mathbf{p}^{C \cap I \backslash\{j\}} \neq 0
$$

contradicting $\mathbf{p} \in \operatorname{inv}_{I}(\mathcal{L})$. Therefore $T$ is a flat of $M$ and (2.3) holds. It follows that $\mathbf{p}$, or more precisely $\pi_{T}(\mathbf{p})$, is a point of support $S$ in $\pi_{T}(\mathcal{L})$. Therefore $\pi_{T}(\mathcal{L}) \cap \mathbb{C}^{S}$ contains a point of full support, the contraction of the matroid $\left.M\right|_{T}$ by $T \backslash S$ has no loops, and $T \backslash S$ is a flat of the matroid $\left.M\right|_{T}$.

Example 2.5.4. Suppose $\mathcal{L}$ is a generic $d$-dimensional subspace of $\mathbb{C}^{n}$, meaning that $M=M(\mathcal{L})$ is the uniform matroid of rank $d$ on $[n]$. Its flats are the subsets $F \subseteq[n]$ of size $|F|<d$, along with the full set $[n]$. Consider $S \subseteq[n]$ and $T=S \cup \bar{I}$. If $T$ is a flat of $M$, then either $|T|<d$, implying $|\bar{I}|<d$, or $T=[n]$, in which case $I \subseteq S$. If $|T|<d$, then $\left.M\right|_{T}$ is the uniform matroid of rank $|T|$ on the elements $T$. Then every subset of $T$ is a flat of $\left.M\right|_{T}$, meaning that $S$ is the support of a point in $\operatorname{inv}_{I}(\mathcal{L})$. If $T=[n]$, then $T \backslash S=\bar{S}$ is a flat of $\left.M\right|_{T}=M$ if and only if $|\bar{S}|<d$ or $|\bar{S}|=n$. Since $S$ contains $I,|\bar{S}|=n$ only when $I=S=\emptyset$. Therefore if $I \neq \emptyset$, we have $|S|>n-d$. Putting these together gives

$$
S \in \operatorname{supp}\left(\operatorname{inv}_{I}(\mathcal{L})\right) \Longleftrightarrow \begin{cases}S=\emptyset \text { or }|S|>n-d & \text { if } I=\emptyset \\ I \subseteq S \text { and }|S|>n-d & \text { if } 0<|I| \leq n-d \\ |S \cup \bar{I}|<d \text { or } I \subseteq S & \text { if } n-d<|I|\end{cases}
$$

### 2.6 Real Points and Hyperplane Arrangements

In this section, we explore a slight variation of $\operatorname{inv}_{I}$ that preserves a real-rootedness property of certain intersections. For $I \subseteq[n]$, define the rational map inv ${ }_{I}^{-}: \mathbb{C}^{n} \rightarrow \mathbb{C}^{n}$ by

$$
\left.\operatorname{inv}_{I}^{-}(x)\right)_{i}= \begin{cases}-1 / x_{i} & \text { if } i \in I \\ x_{i} & \text { if } i \notin I\end{cases}
$$

Equivalently this is the composition of $\operatorname{inv}_{I}$ with the map that scales coordinates $x_{i}$ for $i \in I$ by -1 . Note that the varieties $\operatorname{inv}_{I}(\mathcal{L})$ and $\operatorname{inv}_{I}^{-}(\mathcal{L})$ are isomorphic, and in particular they have the same degree. For any linearspace $\mathcal{L} \subset \mathbb{C}^{n}$, let $\mathcal{L}^{\perp}$ denote the subspace of vectors $v$ for which $\sum_{i=1}^{n} v_{i} x_{i}=0$ for all $x \in \mathcal{L}$.

Proposition 2.6.1. If $\mathcal{L} \subset \mathbb{C}^{n}$ is invariant under complex conjugation, then for any $u \in \mathbb{R}^{n}$, all of the intersection points of $\operatorname{inv}_{I}^{-}(\mathcal{L})$ with $\mathcal{L}^{\perp}+u$ are real.

Proof. If $\mathcal{L}$ is contained in a coordinate hyperplane $\left\{x_{i}=0\right\}$ where $i \in I$, then $\operatorname{inv}_{I}^{-}(\mathcal{L})$ is empty and the claim trivially follows. Otherwise, the points $x \in \operatorname{inv}_{I}^{-}(\mathcal{L})$ with $x_{i} \neq 0$ for $i \in I$ are necessarily Zariski-dense, and for a generic point $u \in \mathbb{R}^{n}$, the intersection points of $\operatorname{inv}_{I}^{-}(\mathcal{L})$ with $\mathcal{L}^{\perp}+u$ belongs to $\left(\mathbb{C}^{*}\right)^{I} \times \mathbb{C}^{[n] \backslash I}$. Showing that these intersection points are real for generic $u$ implies it for all.

Suppose that a point $a+\mathrm{i} b$ belongs to the intersection of $\operatorname{inv}_{I}^{-}(\mathcal{L})$ with $\mathcal{L}^{\perp}+u$ where $a, b \in \mathbb{R}^{n}$ and $a_{j}+\mathrm{i} b_{j} \neq 0$ for every $j \in I$. Then $(a-u)+\mathrm{i} b$ belongs to $\mathcal{L}^{\perp}$. Since $\mathcal{L}^{\perp}$ is conjugation invariant, it
follows that $b \in \mathcal{L}^{\perp}$. In particular, $b^{T} x=0$ for all $x \in \mathcal{L}$. Since $a+\mathrm{i} b$ belongs to $\operatorname{inv}_{I}^{-}(\mathcal{L}), \operatorname{inv}_{I}^{-}(a+\mathrm{i} b)$ belongs to $\mathcal{L}$, meaning that $b^{T} \operatorname{inv}_{I}^{-}(a+\mathrm{i} b)=0$. Taking imaginary parts gives

$$
0=\operatorname{Im}\left(\sum_{j \in I} \frac{-b_{j}}{a_{j}+\mathrm{i} b_{j}}+\sum_{j \notin I} b_{j}\left(a_{j}+\mathrm{i} b_{j}\right)\right)=\sum_{j \in I} \frac{b_{j}^{2}}{a_{j}^{2}+b_{j}^{2}}+\sum_{j \notin I} b_{j}^{2}
$$

Since every term is nonnegative and their sum is zero, each term must be zero, meaning that $b_{j}=0$ for all $j$ and the point $a+\mathrm{i} b$ is real.

Remark 2.6.2. Propostion 2.6 .1 shows that $\operatorname{inv}_{I}^{-}(\mathcal{L})$ is hyperbolic with respect to $\mathcal{L}^{\perp}$, in the sense of [90]. In fact, one can replace $\mathcal{L}^{\perp}$ in this statement by any linear space of the same dimension whose non-zero Plücker coordinates agree in sign with those of $\mathcal{L}^{\perp}$. This shows that $\operatorname{inv}_{I}^{-}(\mathcal{L})$ is a stable variety. See [56, Section 2] for more.

Proposition 2.6.3. For generic $u \in \mathbb{R}^{n}$, the intersection points of $\operatorname{inv}_{I}^{-}(\mathcal{L})$ with $\mathcal{L}^{\perp}+u$ are the minima of the function

$$
\begin{equation*}
f(x)=\frac{1}{2} \sum_{j \notin I} x_{j}^{2}-\sum_{j \in I} \log \left|x_{j}\right| \tag{2.4}
\end{equation*}
$$

over the regions in the complement of the hyperplane arrangement $\left\{x_{i}=0\right\}_{i \in I}$ in $\mathcal{L}^{\perp}+u$.
Proof. On $\left(\mathbb{R}^{*}\right)^{I} \times \mathbb{R}^{[n] \backslash I}, f$ is infinitely differentiable and we examine its behavior on each orthant. For a sign pattern $\sigma: I \rightarrow\{ \pm 1\}$, let $\mathbb{R}_{\sigma}^{I}$ denote the orthant of points in $\left(\mathbb{R}^{*}\right)^{I}$ with $\sigma(i) x_{i}>0$ for all $i \in I$. Inspecting the Hessian of $f$ shows that it is also strictly convex on $\mathbb{R}_{\sigma}^{I} \times \mathbb{R}^{[n] \backslash I}$. Indeed, the Hessian of $f$ is a diagonal matrix whose $(j, j)$ th entry is equal to $1 / x_{j}^{2}$ for $j \in I$ and 1 for $j \notin I$ and is therefore positive definite on $\left(\mathbb{R}^{*}\right)^{I} \times \mathbb{R}^{[n] \backslash I}$.

Define the (open) polyhedron $\mathcal{P}_{\sigma}$ to be the intersection of $\mathbb{R}_{\sigma}^{I} \times \mathbb{R}^{[n] \backslash I}$ with the affine space $\mathcal{L}^{\perp}+u$. The function $f$ is strictly convex on $\mathcal{P}_{\sigma}$, meaning that any critical point of $f$ over $\mathcal{P}_{\sigma}$ is a global minimum. The affine span of $\mathcal{P}_{\sigma}$ is $\mathcal{L}^{\perp}+u$, so $p \in \mathcal{P}_{\sigma}$ is a critical point of $f$ when $\nabla f(p)$ belongs to $\left(\mathcal{L}^{\perp}\right)^{\perp}=\mathcal{L}$. Since $\nabla f(p)=\operatorname{inv}_{I}^{-}(p)$ and $\operatorname{inv}_{I}^{-}$is an inversion, this implies that $p$ belongs to $\operatorname{inv}_{I}^{-}(\mathcal{L})$. Putting this all together, we find that for a point $p \in \mathcal{P}_{\sigma}$,

$$
p \text { attains the minimum of } f \text { over } \mathcal{P}_{\sigma} \Leftrightarrow \nabla f(p) \in \mathcal{L} \Leftrightarrow p \in \operatorname{inv}_{I}^{-}(\mathcal{L})
$$

We can characterize which connected components of $\left(\mathcal{L}^{\perp}+u\right) \backslash\left\{x_{i}=0\right\}_{i \in I}$ contains a point in $\operatorname{inv}_{I}^{-}(\mathcal{L})$ in terms of the recession cone $\operatorname{rec}\left(\mathcal{P}_{\sigma}\right)=\left(\mathbb{R}_{\sigma}^{I} \times \mathbb{R}^{[n] \backslash I}\right) \cap \mathcal{L}^{\perp}$.

Lemma 2.6.4. The infimum of $f$ over $\mathcal{P}_{\sigma}$ is attained if and only if the intersection of $\mathbb{R}^{I}$ with the recession cone of $\mathcal{P}_{\sigma}$ is trivial, i.e. $\operatorname{rec}\left(\mathcal{P}_{\sigma}\right) \cap \mathbb{R}^{I}=\{0\}$.

Proof. $(\Rightarrow)$ Suppose $\operatorname{rec}\left(\mathcal{P}_{\sigma}\right) \cap \mathbb{R}^{I}$ contains $v \neq 0$. Then for any $p \in \mathcal{P}_{\sigma}$, the univariate function $f(p+t v)=\frac{1}{2} \sum_{j \notin I} p_{j}^{2}-\sum_{i \in I} \log \left|p_{i}+t v_{i}\right|$ is strictly decreasing as $t \rightarrow \infty$ and the infimum of $f$ is not attained on $\mathcal{P}_{\sigma}$.
$(\Leftarrow)$ Suppose that $\operatorname{rec}\left(\mathcal{P}_{\sigma}\right) \cap \mathbb{R}^{I}=\{0\}$. Then the quadratic form $\sum_{j \notin I} x_{j}^{2}$ is positive definite on the recession cone $\operatorname{rec}\left(\mathcal{P}_{\sigma}\right)$. We can write $\mathcal{P}_{\sigma}$ as $Q+\operatorname{rec}\left(\mathcal{P}_{\sigma}\right)$, where $Q$ is a compact polytope. Let $S$ denote the section of the recession cone, $S=\left\{v \in \operatorname{rec}\left(\mathcal{P}_{\sigma}\right):\|v\|_{1}=1\right\}$. For any point $p \in Q$ and $v \in S$, consider the univariate function $t \mapsto f(p+t v)$, which is strictly convex and continuous on $\left\{t: p+t v \in \mathcal{P}_{\sigma}\right\}$. Its derivative

$$
\frac{d}{d t} f(p+t v)=\sum_{j \notin I} v_{j} p_{j}+t \sum_{j \notin I} v_{j}^{2}-\sum_{i \in I} \frac{v_{i}}{p_{i}+t v_{i}}
$$

has a unique root $t \in \mathbb{R}$ for $p+t v \in \mathcal{P}_{\sigma}$. Indeed, by assumption $\sum_{j \notin I} v_{j}^{2}>0$. Then, since $\frac{d^{2}}{d t^{2}} f(p+t v)>0$ where defined, $\frac{d}{d t} f(p+t v)$ is strictly increasing on $\left\{t: p+t v \in \mathcal{P}_{\sigma}\right\}$. If $v \in \mathbb{R}^{[n] \backslash I}$, then this set is all of $\mathbb{R}$ and $\frac{d}{d t} f(p+t v)$ is linear. Otherwise, there is a minimum $t$ for which $p+t v \in \mathcal{P}_{\sigma}$ and $\frac{d}{d t} f(p+t v) \rightarrow-\infty$ as $t$ approaches this minimum, whereas $\frac{d}{d t} f(p+t v)>0$ for sufficiently large $t$. Let $t^{*}(p, v)$ denote this unique root of $\frac{d}{d t} f(p+t v)$. This is a continuous function in $p$ and $v$. Let $T$ denote the maximum of $t^{*}(p, v)$ over $(p, v) \in Q \times S$.

Now we claim that when minimizing $f$ over $\mathcal{P}_{\sigma}$, it suffices to minimize over the compact set $Q+[0, T] S$. Indeed, if $y \in \mathcal{P}_{\sigma}$, then $y=p+t v$ for some $p \in Q, v \in S$ and $t \in \mathbb{R}_{>0}$. If $t>T$, then the point $x=p+T v \in Q+[0, T] S$ satisfies $f(x)<f(y)$. In particular, the minimum of $f$ is bounded from below and is therefore attained on the compact set $Q+[0, T] S$.

Proposition 2.6.5. For generic $u \in \mathbb{R}^{n}$, there is exactly one point of $\operatorname{inv}_{I}^{-}(\mathcal{L})$ in each region of $\left(\mathcal{L}^{\perp}+u\right) \backslash\left\{x_{i}=0\right\}_{i \in I}$ whose recession cone has trivial intersection with $\mathbb{R}^{I}$. The degree of $\operatorname{inv}_{I}^{-}(\mathcal{L})$ equals the number of these regions.

Proof. First we show that for generic $u \in \mathbb{R}^{n}$, the number of intersection points of inv ${ }_{I}^{-}(\mathcal{L})$ with $\mathcal{L}^{\perp}+u$ equals the degree of $\operatorname{inv}_{I}^{-}(\mathcal{L})$. To do this, we show that the closure $\overline{\operatorname{inv}_{I}^{-}(\mathcal{L})}$ in $\mathbb{P}^{n}(\mathbb{C})$ has no points in common with $\mathcal{L}^{\perp}+x_{0} u$ with $x_{0}=0$. For the sake of contradiction suppose that for some $a \in \mathcal{L}^{\perp}$, the point $[0: a]$ belongs to $\overline{\operatorname{inv}_{I}^{-}(\mathcal{L})}$ and let $S=\operatorname{supp}(a)$.

It follows that $a^{T} \mathbf{x}=\sum_{i \in S} a_{i} x_{i}$ vanishes on $\mathcal{L}, g=\mathbf{x}^{S \cap I} \cdot a^{T} \operatorname{inv}_{I}^{-}(\mathbf{x})$ vanishes on $\operatorname{inv}_{I}^{-}(\mathcal{L})$, and the homogenezation $g^{\text {hom }}$ with respect to $x_{0}$ vanishes on the closure $\overline{\operatorname{inv}_{I}^{-}(\mathcal{L})} \subseteq \mathbb{P}^{n}(\mathbb{C})$. In particular, $g^{\text {hom }}(0, a)=0$. If $S \subseteq I$, this contradicts the evaluation

$$
g^{\mathrm{hom}}=g=\sum_{j \in S} a_{j} \mathbf{x}^{S \backslash j} \quad \text { given by } g^{\text {hom }}(0, a)=\sum_{j \in S} a^{S}=a^{S} \cdot|S| \neq 0
$$

Similarly, since $\overline{\operatorname{inv}_{I}^{-}(\mathcal{L})}$ is invariant under complex conjugation, we also have $g^{\text {hom }}(0, \bar{a})=0$, where $\bar{a}$ is the complex conjugate of $a$. If $S \nsubseteq I$, this contradicts the evaluation of

$$
g^{\mathrm{hom}}=-x_{0}^{2} \sum_{j \in S \cap I} a_{j} \mathbf{x}^{S \cap I \backslash j}+\mathbf{x}^{S \cap I} \sum_{j \in S \backslash I} a_{j} x_{j} \text { given by } g^{\text {hom }}(0, \bar{a})=\bar{a}^{S \cap I} \sum_{j \in S \backslash I} a_{j} \overline{a_{j}} \neq 0 .
$$

Therefore all the intersection points of $\overline{\operatorname{inv}_{I}^{-}(\mathcal{L})}$ with $\mathcal{L}^{\perp}+x_{0} u$ have $x_{0} \neq 0$. Then for generic $u$, the number of intersection points of $\operatorname{inv}_{I}^{-}(\mathcal{L})$ and $\mathcal{L}^{\perp}+u$ equals the degree of $\operatorname{inv}_{I}^{-}(\mathcal{L})$.


Figure 2.1 Intersections of $\mathcal{L}^{\perp}+u$ with $\operatorname{inv}_{I}^{-}(\mathcal{L})$ from Example 2.6.7.

By Propositions 2.6.1 and 2.6.3, each of these intersection points is real and thus is a minimizer of the function $f(x)$ of (2.4) over some connected component $\mathcal{P}_{\sigma}$ of $\left(\mathcal{L}^{\perp}+u\right) \backslash\left\{x_{i}=0\right\}_{i \in I}$. By Lemma 2.6.4, the components $\mathcal{P}_{\sigma}$ contains a minimizer if and only if $\operatorname{rec}\left(\mathcal{P}_{\sigma}\right) \cap \mathbb{R}^{I}=\{0\}$.

This together with Corollary 2.4.4 constitutes the proof of Theorem 2.2.2. For special cases of $I$, we find a simpler characterization of the regions counted by $\operatorname{deg}\left(\operatorname{inv}_{I}(\mathcal{L})\right)$.

Corollary 2.6.6. Let $u \in \mathbb{R}^{n}$ be generic. If I is independent in the matroid $M(\mathcal{L})$, then the degree of $\operatorname{inv}_{I}(\mathcal{L})$ equals the total number of regions in $\left(\mathcal{L}^{\perp}+u\right) \backslash\left\{x_{i}=0\right\}_{i \in I}$. If $I=[n]$, then the degree of $\operatorname{inv}_{I}(\mathcal{L})$ equals the number of bounded regions in $\left(\mathcal{L}^{\perp}+u\right) \backslash\left\{x_{i}=0\right\}_{i \in I}$.

Proof. If $I$ is independent in $M(\mathcal{L})$, then $I$ is contained in a basis $B$ of $M(\mathcal{L})$, and $[n] \backslash B$ is a basis of $M\left(\mathcal{L}^{\perp}\right)$ contained in $[n] \backslash I$. In particular, if $x \in \mathcal{L}^{\perp}$ has $x_{j}=0$ for all $j \in[n] \backslash I$, then $x=0$. So $\mathbb{R}^{I} \cap \mathcal{L}^{\perp}=\{0\}$. The recession cone of any region of $\left(\mathcal{L}^{\perp}+u\right) \backslash\left\{x_{i}=0\right\}_{i \in I}$ is contained in $\mathcal{L}^{\perp}$, so its intersection with $\mathbb{R}^{I}$ is trivial.

If $I=[n]$, then $\mathbb{R}^{I}=\mathbb{R}^{n}$. The recession cone of a region in $\left(\mathcal{L}^{\perp}+u\right) \backslash\left\{x_{i}=0\right\}_{i \in I}$ contains a non-zero vector if and only if it is unbounded. Therefore the regions whose recession cones have trivial intersection with $\mathbb{R}^{I}$ are those which are bounded.

Example 2.6.7. Consider the 3 -dimensional linearspace $\mathcal{L}$ from Example 2.4.3 and take the vector $u=(0,0,1,2,2)$. The two-dimensional affine space $\mathcal{L}^{\perp}+u$ consists of points of the form $\left(x_{1}, x_{2}, x_{1}-\right.$ $\left.x_{2}+1,-x_{2}+2,-x_{1}+x_{2}+2\right)$. Since $I=\{1,2,3\}$ is independent in $M(\mathcal{L})$, each of the seven regions in the complement of the hyperplane arrangement $\left\{x_{i}=0\right\}_{i \in I}$ contains a point of $\operatorname{inv}_{I}^{-}(\mathcal{L})$. For $I=\{1,2,3,4\}$, there are four regions whose recession cones intersect $\left\{x_{5}=0\right\}$ nontrivially. The remaining six regions each contain a unique point in $\operatorname{inv}_{I}^{-}(\mathcal{L})$. Finally, for $I=\{1,2,3,4,5\}, \mathbb{R}^{I}$ is all of $\mathbb{R}^{5}$ so the recession cone of $\mathcal{P}_{\sigma}$ intersects $\mathbb{R}^{I}$ nontrivially if and only if $\mathcal{P}_{\sigma}$ is unbounded, meaning that the four bounded regions of the hyperplane arrangement $\left\{x_{i}=0\right\}_{i \in I}$ in $\mathcal{L}^{\perp}+u$ are precisely those that contain points in $\operatorname{inv}_{I}^{-}(\mathcal{L})$. These hyperplane arrangements and intersection points are shown in Figure 2.1.

Rational mapping, combinatorial structure, matroid controlled.

## Chapter 3

## Log-Concave Maximum Likelihood Estimation

The work presented in this chapter is based on a joint project with A. Grosdos, A. Heaton, K. Kubjas, O. Kuznetsova and M-S. Sorea and has been submitted for peer review. The preprint is accessible at: https://arxiv.org/abs/2003.04840. The code for computations appearing in this chapter can be found at: https://github.com/agrosdos/Computing-the-Exact-LogConcave-MLE. We study the process of inferring a log-concave probability density function from a weighted point configuration $(X, w)$ by maximizing the likelihood of observing this data as a sample from said probability density function. We consider a non-parametric approach [102] to Maximum Likelihood Estimation (MLE); instead of specifying a statistical model, we impose a log-concave shape constraint on the density function. Log-concavity is a geometric property appearing frequently in statistics $[7,35]$ and in algebraic combinatorics [52]. We study probability density functions that are log-concave. Despite the space of all such densities being infinite-dimensional, the maximum likelihood estimate is the exponential of a piecewise linear function determined by finitely many quantities, namely the function values, or heights, at the data points. We explore in what sense exact solutions to this problem are possible. First, we show that the heights given by the maximum likelihood estimate are generically transcendental. For a cell in one dimension, the maximum likelihood estimator is expressed in closed form using the generalized $W$-Lambert function. We show that finding the log-concave maximum likelihood estimate is equivalent to solving a collection of polynomial-exponential systems of a special form. We use Smale's $\alpha$-theory to refine approximate numerical solutions and to certify solutions to log-concave density estimation.

### 3.1 Motivations and Background

Nonparametric methods in statistics emerged in the 1950-1960s [42, 81, 71, 6] and fall into two main streams: smoothing methods and shape constraints. Examples of smoothing methods include delta sequence methods such as kernel, histogram and orthogonal series estimators [100], and penalized maximum likelihood estimators, e.g., spline methods [39]. Their defining feature is the need to choose
the smoothing or tuning parameters. It is a delicate process because smoothing parameters depend on the unknown probability density function. In contrast to smoothing methods, shape constrained nonparametric density estimation is fully automatic and does not depend on the underlying probability distribution, though this comes at the expense of worse $L^{1}$ convergence rates for smooth densities [38]. Some previously studied classes of functions include non-increasing [43], convex [45], $k$-monotone [ 9 , 10 ] and $s$-concave [33]. For general references on nonparametric statistics we refer the reader to [92, 89 , 97, 44].

We focus on the class of log-concave densities, which is an important special case of $s$-concave densities. The choice of log-concavity is attractive for several reasons. First of all, most common univariate parametric families are log-concave, including the normal, Gamma with shape parameter greater than one, Beta densities with parameters greater than 1, Weibull with parameter greater than 1 and others. Furthermore, log-concavity is used in reliability theory, economics and political science [7]. In addition to this, log-concave densities have several desirable statistical properties. For example, log-concavity implies unimodality but log-concave density estimation avoids the spiking phenomenon common in general unimodal estimation [35]. Moreover, this class is closed under convolutions and taking pointwise limits [26]. We refer the reader to [83] for an overview of the recent progress in the field.

Let $X=\left(x_{1}, x_{2}, \ldots, x_{m}\right)$ be a point configuration in $\mathbb{R}^{n}$ with weights $w=\left(w_{1}, w_{2}, \ldots, w_{m}\right)$ such that $w_{i} \geq 0$ and $w_{1}+w_{2}+\cdots+w_{m}=1$. The log-concave maximum likelihood estimation (MLE) problem aims to find a Lebesgue density that solves

$$
\begin{equation*}
\max \sum_{i=1}^{m} w_{i} \log \left(f\left(x_{i}\right)\right) \text { s.t. } \log (f) \text { is concave and } \int_{\mathbb{R}^{n}} f(x) d x=1 \tag{3.1}
\end{equation*}
$$

It has been shown that the solution exists with probability 1 and is unique, and its logarithm is a tent function, i.e., a piecewise linear function with regions of linearity inducing a subdivision of the convex hull of $X[101,70,27,79]$, see Figure 3.1 for an example. While MLE is the most widely studied estimator in this setting, it is not the only one, for examples see [31, 28].

The maximum likelihood estimator is attractive because of its consistency under general assumptions [70, 35, 26, 37] and superior performance compared to kernel-based methods with respect to mean integrated squared error, as observed in simulations [27]. At the same time, the convergence rate is still an open question and only lower [55,57] and upper [55, 21] bounds are known. Further theoretical properties have been studied for some special cases of log-concave densities, e.g., $k$-affine densities [54] and totally positive densities [78].

Example 3.1.1. Consider the sample of 14 points in $\mathbb{R}^{2}$ with uniform weights:

$$
X=((0,1),(0,9),(1,4),(2,4),(2,6),(3,3),(5,5),(6,3),(6,9),(7,6),(7,8),(8,9),(9,5),(9,9))
$$

How many cells does the subdivision induced by the logarithm of the optimal log-concave density have?
Using the R package LogConcDEAD with default parameters, one obtains that the logarithm of the
maximum likelihood estimate is a piecewise linear function with seven unique linear pieces. However, when one investigates the optimal density more closely, it appears that several linear pieces are similar. For example, a visual inspection of the optimal density depicted in Figure 3.1 suggests that there are only four unique linear pieces. The true number of unique linear pieces of the optimal density is hard to find.

Theoretically, the algorithm used in LogConcDEAD finds the true optimal density, however, in practice, the answer is a numerical approximation. By changing the parameter sigmatol from default value $10^{-8}$ to $10^{-10}$, LogConcDEAD outputs four unique linear pieces, exactly as we observed in Figure 3.1. Although it might seem obvious that four is the correct number of linear pieces, in reality the situation is more complicated, see Example 3.4.16.


Figure 3.1 The optimal tent function for the sample of 14 points in Example 3.1.1.

### 3.2 Geometry of log-concave maximum likelihood estimation

We start by reviewing the geometry of log-concave maximum likelihood estimation mostly following [79].
Recall from definition 1.3.5 that given a point configuration $X=\left(x_{1}, x_{2}, \ldots, x_{m}\right) \subset \mathbb{R}^{n}$ and a fixed real vector $y \in \mathbb{R}^{m}$, we define the tent function $h_{X, y}: \mathbb{R}^{n} \rightarrow \mathbb{R}$ to be the smallest concave function such that $h_{X, y}\left(x_{i}\right) \geq y_{i}$ for $i=1, \ldots, m$. The tent function $h_{X, y}$ is piecewise linear on $P=\operatorname{conv}(X)$ with linear pieces equal to upper facets of the convex hull of the points $\left(x_{1}, y_{1}\right),\left(x_{2}, y_{2}\right), \ldots,\left(x_{m}, y_{m}\right)$ in $\mathbb{R}^{m+1}$. We define $h_{X, y}(x)=-\infty$ at all points $x \in \mathbb{R}^{n}$ outside $P$. If $h_{X, y}\left(x_{i}\right)=y_{i}$ for $i=1, \ldots, n$, then $y$ is called relevant.

It was shown by Cule, Samworth and Stewart for uniform weights [27] and by Robeva, Sturmfels and Uhler in general [79] that the constrained optimization problem (3.1) of finding the log-concave
maximum likelihood estimate is equivalent to the unconstrained optimization problem

$$
\begin{equation*}
\max _{y \in \mathbb{R}^{m}} w \cdot y-\int_{P} e^{h_{X, y}(t)} d t . \tag{3.2}
\end{equation*}
$$

In [79], Robeva, Sturmfels and Uhler applied the Legendre-Fenchel transformation to the convex function $\int_{P} e^{h_{X, y}(t)} d t$ in order to recast the optimization problem (3.2) as maximizing a linear functional over a convex set called the Samworth body of $X$. Moreover, the log-concave maximum likelihood estimate is a tent function with tent poles at some of the $x_{i}$. Therefore finding the log-concave density which maximizes the likelihood of $(X, w)$ is equivalent to finding a unique optimal height vector $y^{*}$.

Corollary 3.2.1. [79, Corollary 2.6] To find the optimal height vector $y^{*}$ in (3.2) is to maximize the following rational-exponential objective function over $y \in \mathbb{R}^{m}$ :

$$
\begin{equation*}
S\left(y_{1}, \ldots, y_{m}\right)=w \cdot y-\sum_{\sigma \in \Delta} \sum_{i \in \sigma} \frac{\operatorname{vol}(\sigma) \cdot e^{y_{i}}}{\prod_{\alpha \in \sigma \backslash i}\left(y_{i}-y_{\alpha}\right)} \tag{3.3}
\end{equation*}
$$

where $\Delta$ is any regular triangulation that refines the regular subdivision induced by the tent function $h_{X, y}$.

If $y$ induces a regular subdivision $\Delta$ that is not a maximal regular triangulation, then we can consider any maximal regular triangulation that refines $\Delta$. Thus if there are $M$ maximal regular triangulations of $X$, then to find the optimal $y^{*}$ we must compare the optimal values $y_{\Delta_{1}}^{*}, y_{\Delta_{2}}^{*}, \ldots, y_{\Delta_{M}}^{*}$ which are obtained by solving the optimization problem (3.3) $M$ times, once for each maximal regular triangulation $\Delta_{1}, \Delta_{2}, \ldots, \Delta_{M}$.

Notation 3.2.2. We will denote by $S_{\Delta}$ the function given by the right hand side of (3.3) for a fixed triangulation $\Delta$.

Example 3.2.3. Fix $d=1, m=3$ and $X=(2,5,7)$. The configuration $X$ has two triangulations $\Delta_{1}=\{13\}$ and $\Delta_{2}=\{12,23\}$, which are both regular triangulations. Only $\Delta_{2}$ is a maximal triangulation. Hence solving the optimization problem (3.2) is equivalent to maximizing the objective function

$$
\begin{equation*}
S_{\Delta_{2}}=w \cdot y-3 \frac{e^{y_{1}}-e^{y_{2}}}{y_{1}-y_{2}}-2 \frac{e^{y_{2}}-e^{y_{3}}}{y_{2}-y_{3}} . \tag{3.4}
\end{equation*}
$$

If $y_{1}=y_{2}$ or $y_{2}=y_{3}$, then a denominator on the right hand side of (3.4) becomes zero. However, the objective function in the formulation (3.2) can be still simplified to

$$
w \cdot y-3 e^{y_{2}}-2 \frac{e^{y_{2}}-e^{y_{3}}}{y_{2}-y_{3}} \quad \text { or } \quad w \cdot y-3 \frac{\left(e^{y_{1}}-e^{y_{2}}\right)}{y_{1}-y_{2}}-2 e^{y_{2}} .
$$

To visualize the situation, we consider the Samworth body

$$
\mathcal{S}(X)=\left\{y \in \mathbb{R}^{3}: \int_{P} e^{h_{X, y}(t)} d t \leq 1\right\}
$$



Figure 3.2 The Samworth body for $X=(2,5,7)$.
which was introduced in [79]. The unconstrained optimization problem (3.2) is equivalent to the constrained optimization problem of maximizing the linear function $w \cdot y$ over the Samworth body. For different choices of weight vector $w=\left(w_{1}, w_{2}, w_{3}\right)$, we obtain different optimal height vectors $y=\left(y_{1}, y_{2}, y_{3}\right)$ on the surface of the Samworth body, and the height vector determines the triangulation. The Samworth body consists of two regions that can be seen in Figure 3.2. The green region comes from the one-simplex triangulation $\Delta_{1}=\{13\}$, while the red region comes from the two-simplex triangulation $\Delta_{2}=\{12,23\}$. Moreover, one can see lines separating the green region into two pieces and the red region into three pieces (ignore the curve separating the green and the red regions for now). These lines correspond to the degenerate cases where $y_{1}=y_{3}, y_{1}=y_{2}$ or $y_{2}=y_{3}$, and hence the right hand side of (3.3) is not defined. Therefore those lines are simply artifacts of the reformulation (3.3) since in the original unconstrained setting (3.2) these points present no difficulty. The intersection of the three lines is the point $(-\log 5,-\log 5,-\log 5)$.

Consider the curve separating the green and red regions of the Samworth body. This curve is made of all the points $y$ that form a relevant tent function, inducing the subdivision $\Delta_{1}$. To understand the green region, see the piecewise linear functions drawn in Figure 3.3. Since the lowest (dotted) function is not concave, it is invalid as a tent function. Therefore, if the height $y_{2}$ is too low, the optimal tent function will be the (solid-line) linear function. In effect, the optimal tent-function ignores heights $y_{i}$ if they are too low. This basic phenomenon is responsible for the green part of the Samworth body being flat in the $y_{2}$ direction, meaning that it is a pencil of half-lines parallel to the $y_{2}$-axis.

The transition from the red region to the green region is not smooth. For every $y$ on the curve between the green and red regions, there is a two-dimensional cone of weight vectors that give $y$ as an optimal


Figure 3.3 The red tent function corresponds to a vector $y$ in the red region of the Samworth body. The solid green tent function corresponds to a vector $y$ on the curve separating red and green regions of the Samworth body. The dotted green function is not convex. Its height vector $y$ belongs to the green region of the Samworth body and both green sets of heights give the same tent function.
solution. The generators of this cone are described in [79, Theorem 3.7]. The optimal height vector $y^{*}$ for $w=\left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)$ lies on the curve between the red and green regions. It is not a critical point of the function (3.4), because $w$ is not a normal vector to the red region at the point $y^{*}$.

We now return to the general situation and consider the specific approach of critical equations for solving the optimization problem (3.3). Let $X=\left(x_{1}, \ldots, x_{m}\right)$ be a configuration of $n$ points $x_{i} \in \mathbb{R}^{n}$. Fixing a maximal regular triangulation $\Delta$ of our point configuration $X$, we can find the optimal $y_{\Delta}^{*}$ for $S_{\Delta}$ in (3.3) over $y \in \mathbb{R}^{m}$ by solving the system of critical equations $\partial S_{\Delta} / \partial y_{j}=0$. These partial derivatives take the form (see [79, Proof of Lemma 3.4]):

$$
\begin{align*}
\frac{\partial S_{\Delta}}{\partial y_{j}}=w_{j} & -\sum_{\substack{\sigma \in \Delta, j \in \sigma}} \operatorname{vol}(\sigma) e^{y_{j}} \frac{1}{\prod_{\alpha \in \sigma \backslash j}\left(y_{j}-y_{\alpha}\right)}\left(1-\sum_{\alpha \in \sigma \backslash j} \frac{1}{\left(y_{j}-y_{\alpha}\right)}\right) \\
& -\sum_{\substack{\sigma \in \Delta \\
j \in \sigma}} \operatorname{vol}(\sigma) \sum_{i \in \sigma \backslash j} e^{y_{j}} \frac{1}{\prod_{\alpha \in \sigma \backslash j}\left(y_{j}-y_{\alpha}\right)} \frac{1}{\left(y_{j}-y_{k}\right)} . \tag{3.5}
\end{align*}
$$

Definition 3.2.4. For a fixed maximal regular triangulation $\Delta$ of $X$, let $A$ be the matrix such that the system of $n$ critical equations (3.5) can be written in the form

$$
\begin{equation*}
A e^{y}=w, \tag{3.6}
\end{equation*}
$$

where $e^{y}$ is a column vector of exponentials $\left(e^{y_{1}}, e^{y_{2}}, \ldots, e^{y_{m}}\right)^{T}$, and $w$ is a column vector of weights $\left(w_{1}, \ldots, w_{m}\right)^{T}$. The matrix $A$ is called the score equation matrix.

The entries of $A$ are in the field of rational functions in the variables $y_{1}, \ldots, y_{m}$. Diagonal entries of $A$ are

$$
A_{j, j}=\sum_{\substack{\sigma \in \Delta, j \in \sigma}} \operatorname{vol}(\sigma) \frac{1}{\prod_{\alpha \in \sigma \backslash j}\left(y_{j}-y_{\alpha}\right)}\left(1-\sum_{\alpha \in \sigma \backslash j} \frac{1}{\left(y_{j}-y_{\alpha}\right)}\right)
$$

and off-diagonal entries of $A$ are

$$
A_{i, j}=\sum_{\substack{\sigma \in \Delta, i, j \in \sigma}} \operatorname{vol}(\sigma) \frac{1}{\prod_{\alpha \in \sigma \backslash j}\left(y_{j}-y_{\alpha}\right)} \frac{1}{\left(y_{j}-y_{i}\right)} .
$$

The matrix $A$ can be written as a sum of matrices over maximal simplices $\sigma \in \Delta$. This will be described explicitly in the proof of Theorem 3.3.1.

There are two caveats when solving the optimization problem (3.3) using the method of critical equations. First, it is not enough to consider the system of critical equations $\partial S_{\Delta} / \partial y_{j}=0$ only for each of the maximal regular triangulations $\Delta$, since the optimization problem (3.3) is not smooth. We will demonstrate this phenomenon on the point configuration from Example 3.2.3.

Example 3.2.5. Recall that $d=1, m=3$ and $X=(2,5,7)$. The configuration $X$ has two triangulations $\Delta_{1}=\{13\}$ and $\Delta_{2}=\{12,23\}$. Let $w=\left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)$. The output from LogConcDEAD suggests that the optimal tent function is supported on one cell, with heights given by $y_{1}^{*}=-1.816665, y_{2}^{*}=-1.576024$ and $y_{3}^{*}=-1.415597$. However, the vector $y^{*}$ is neither a critical point of $S_{\Delta_{2}}$ nor of the function

$$
S_{\Delta_{1}}=w \cdot y-5 \frac{e^{y_{1}}-e^{y_{3}}}{y_{1}-y_{3}} .
$$

This can be seen by taking partial derivatives of these functions with respect to $y_{1}, y_{2}, y_{3}$ and substituting $y_{1}^{*}, y_{2}^{*}, y_{3}^{*}$. In the case of $\partial S_{\Delta_{1}} / \partial y_{j}=0$, it is particularly easy to see that there are no solutions, since $\partial S_{\Delta_{1}} / \partial y_{2}=w_{2} \neq 0$. In the case of $\partial S_{\Delta_{2}} / \partial y_{j}=0$, the system of critical equations fails to certify in the sense of Section 3.4.

The points $\left(x_{1}, y_{1}^{*}\right),\left(x_{2}, y_{2}^{*}\right),\left(x_{3}, y_{3}^{*}\right)$ being collinear is equivalent to $\left(x_{2}, y_{2}^{*}\right)=\lambda_{1}\left(x_{1}, y_{1}^{*}\right)+\lambda_{3}\left(x_{3}, y_{3}^{*}\right)$ where $\lambda_{1}, \lambda_{3} \geq 0, \lambda_{1}+\lambda_{3}=1$. Since $x_{1}=2, x_{2}=5, x_{3}=7$, we have $\lambda_{1}=\frac{2}{5}, \lambda_{3}=\frac{3}{5}$. Hence $y_{2}=\frac{2}{5} y_{1}+\frac{3}{5} y_{3}$. Substituting this expression into the objective function (3.4) we get

$$
\widetilde{S}_{\Delta_{2}}=\left(w_{1}+\frac{2}{5} w_{2}\right) y_{1}+\left(w_{3}+\frac{3}{5} w_{2}\right) y_{3}-5 \frac{e^{y_{1}}-e^{y_{3}}}{y_{1}-y_{3}}
$$

which for uniform weights $w=\left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)$ becomes

$$
\begin{equation*}
\widetilde{S}_{\Delta_{2}}=\frac{7}{15} y_{1}+\frac{8}{15} y_{3}-5 \frac{e^{y_{1}}-e^{y_{3}}}{y_{1}-y_{3}} . \tag{3.7}
\end{equation*}
$$

We will verify in Example 3.4.13 that $y^{*}$ is a critical point of the function $\widetilde{S}_{\Delta_{2}}$.
The second caveat is that to find the optimal tent function, it is not enough to merely compare the optimal critical points $y_{\Delta}^{*}$ of $\partial S_{\Delta} / \partial y_{j}=0$ for each subdivision $\Delta$. Denote by $Y_{\Delta}$ the set of $y$ that induce a subdivision that is equal to or coarser than $\Delta$. For each $\Delta$, it also has to be checked that $y_{\Delta}^{*}$ is in $Y_{\Delta}$. If the maximum of $S_{\Delta}$ over $Y_{\Delta}$ is not a critical point of $S_{\Delta}$, then the maximum must be on the boundary of $Y_{\Delta}$, see Figure 3.4 for an illustration. The boundary of $Y_{\Delta}$ is stratified into regions $Y_{\widetilde{\Delta}}$ corresponding to the various subdivisions $\widetilde{\Delta}$ which are refined by $\Delta$. Hence one can consider critical points for strictly
coarser subdivisions $\widetilde{\Delta}$. Thus if $y_{\Delta}^{*}$ is not in $Y_{\Delta}$, then $y_{\Delta}^{*}$ should be discarded.


Figure 3.4 Maximizing $S_{\Delta}$ over $y$ restricted to $Y_{\Delta}$.

### 3.3 Transcendentality and closed-form solutions

In this section we use notions from geometric combinatorics to study the structure of (3.2.4). In particular, we will prove that the matrix $A$ is invertible. This will be our main tool in proving the transcendentality of log-concave MLE and deriving closed form solutions in the one-dimensional one cell case using Lambert functions.

### 3.3.1 Score equation matrix invertibility and transcendentality

Towards proving transcendentality, we first investigate the invertibility of the matrix $A$.
Theorem 3.3.1. Consider a point configuration $X=\left(x_{1}, \ldots, x_{m}\right)$ in $\mathbb{R}^{d}$, let $\Delta=\left\{\sigma_{1}, \ldots, \sigma_{m}\right\}$ be a maximal regular triangulation of $X$. The score equation matrix A from (3.6) is invertible.

Recall definition 1.3.3, given a triangulation $\Delta$, we define the neighborhood $\mathcal{N}(j)$ of a vertex $j$ in $\Delta$ to be the set of vertices

$$
\mathcal{N}(j)=\{i:\{i, j\} \subseteq \sigma \text { for some } \sigma \in \Delta\}
$$

Before giving the proof of Theorem 3.3.1, we illustrate the construction in the proof with a small example.

Example 3.3.2. Let $X=\left(x_{1}, x_{2}, x_{3}, x_{4}\right)$ be a four point configuration in $\mathbb{R}^{2}$ with $\Delta=\left\{\sigma_{1}, \sigma_{2}\right\}$, where $\sigma_{1}=\{1,2,3\}$ and $\sigma_{2}=\{2,3,4\}$. Let $A$ be the score equation matrix for the entire regular triangulation
$\Delta$. Let us denote the difference $y_{i}-y_{j}$ by $y_{i j}$. Then $A=A\left(\sigma_{1}\right)+A\left(\sigma_{2}\right)$, where

$$
\begin{aligned}
& \frac{A\left(\sigma_{1}\right)}{\operatorname{vol}\left(\sigma_{1}\right)}=\left[\begin{array}{cccc}
\frac{1}{y_{12} y_{13}}-\frac{1}{y_{12} y_{13}}-\frac{1}{y_{12} y_{13}{ }^{2}} & \frac{1}{y_{21} y_{23}} & \frac{1}{y_{31} y_{32}} & 0 \\
\frac{1}{y_{12} y_{13}} & \frac{1}{y_{21} y_{23}}-\frac{1}{y_{21} y_{13}}-\frac{1}{y_{21} y_{23}{ }^{2}} & \frac{1}{y_{31} y_{32}{ }^{2}} & 0 \\
\frac{1}{y_{12} y_{13}{ }^{2}} & \frac{1}{y_{21} y_{23}{ }^{2}} & \frac{1}{y_{31} y_{32}}-\frac{1}{y_{31} y_{32}}-\frac{1}{y_{31} y_{32^{2}}{ }^{2}} & 0 \\
0 & 0 & 0 & 0
\end{array}\right], \\
& \frac{A\left(\sigma_{2}\right)}{\operatorname{vol}\left(\sigma_{2}\right)}=\left[\begin{array}{cccc}
0 & 0 & 0 & 0 \\
0 & \frac{1}{y_{23} y_{24}}-\frac{1}{y_{23} y_{24}}-\frac{1}{y_{23} y_{24} 4^{2}} & \frac{1}{y_{32} y_{34}} & \frac{1}{y_{42^{2}} y_{43}} \\
0 & \frac{1}{y_{23} y^{2} y_{24}} & \frac{1}{y_{32} y_{34}}-\frac{1}{y_{32} y_{3} y_{34}}-\frac{1}{y_{32} y_{34}{ }^{2}} & \frac{1}{y_{42} y_{43}{ }^{2}} \\
0 & \frac{1}{y_{23} y_{24^{2}}} & \frac{1}{y_{32} y_{34^{2}}} & \frac{1}{y_{42} y_{43}}-\frac{1}{y_{42} y_{43}}-\frac{1}{y_{42} y_{43^{2}}{ }^{2}}
\end{array}\right] .
\end{aligned}
$$

We define matrix $B$ to be the matrix $A$ with its $j$-th column multiplied by $\prod_{i \in \mathcal{N}(j)} y_{j i}^{2}$, for all $j$ from 1 to 4 . We obtain the following matrices

$$
\begin{aligned}
& \frac{B\left(\sigma_{1}\right)}{\operatorname{vol}\left(\sigma_{1}\right)}=\left[\begin{array}{cccc}
y_{13} y_{12}-y_{12}-y_{13} & y_{24}{ }^{2} y_{23} & y_{34}{ }^{2} y_{32} & 0 \\
y_{13} & y_{21} y_{23} y_{24}{ }^{2}-y_{24}{ }^{2} y_{21}-y_{23} y_{24}{ }^{2} & y_{34}{ }^{2} y_{31} & 0 \\
y_{12} & y_{24}{ }^{2} y_{21} & y_{31} y_{32} y_{34}{ }^{2}-y_{34}^{2} y_{31}-y_{34}^{2} y_{32} & 0 \\
0 & 0 & 0 & 0
\end{array}\right], \\
& \frac{B\left(\sigma_{2}\right)}{\operatorname{vol}\left(\sigma_{2}\right)}=\left[\begin{array}{cccc}
0 & 0 & 0 & 0 \\
0 & y_{21}{ }^{2} y_{23} y_{24}-y_{21}{ }^{2} y_{23}-y_{21}^{2} y_{24} & y_{31}{ }^{2} y_{34} & y_{43} \\
0 & y_{21}{ }^{2} y_{24} & y_{32} y_{31}{ }^{2} y_{34}-y_{31}{ }^{2} y_{32}-y_{31}^{2} y_{34} & y_{42} \\
0 & y_{21}{ }^{2} y_{23} & y_{31}{ }^{2} y_{32} & y_{43} y_{42}-y_{42}-y_{43}
\end{array}\right] .
\end{aligned}
$$

The product of the diagonal entries of $B=B\left(\sigma_{1}\right)+B\left(\sigma_{2}\right)$ is a polynomial of degree 12 . Whereas a term in the expansion of the determinant of $B$ with off-diagonal entries has at most degree 10 .

Proof of Theorem 3.3.1. The score equation matrix $A$ associated to a maximal regular triangulation $\Delta$ can be written as

$$
A=\sum_{\sigma \in \Delta} A(\sigma),
$$

where the entries of $A(\sigma)$ for $i \neq j$ are

$$
\begin{aligned}
& A(\sigma)_{i, j}=\operatorname{vol}(\sigma)\left(\prod_{\alpha \in \sigma \backslash\{j\}} \frac{1}{\left(y_{j}-y_{\alpha}\right)}\right)\left(\frac{1}{y_{j}-y_{i}}\right), \\
& A(\sigma)_{j, j}=\operatorname{vol}(\sigma)\left(\prod_{\alpha \in \sigma \backslash\{j\}} \frac{1}{\left(y_{j}-y_{\alpha}\right)}\right)\left(1-\sum_{\alpha \in \sigma \backslash\{j\}} \frac{1}{\left(y_{j}-y_{\alpha}\right)}\right) .
\end{aligned}
$$

The matrix $A(\sigma)$ is sparse: If $i$ or $j$ does not belong to $\sigma$ then $A_{i, j}(\sigma)=0$.
Let $B$ (resp. $B(\sigma)$ ) be the matrix that is obtained by multiplying the $j$-th column of $A$ (resp. $A(\sigma)$ )
by $\left(\prod_{\alpha \in \mathcal{N}(j)}\left(y_{j}-y_{\alpha}\right)^{2}\right)$ for $j=1, \ldots, n$ :

$$
\begin{equation*}
B_{., j}=A_{., j}\left(\prod_{\alpha \in \mathcal{N}(j)}\left(y_{j}-y_{\alpha}\right)^{2}\right)=\sum_{\sigma \in \Delta} A(\sigma)_{., j}\left(\prod_{\alpha \in \mathcal{N}(j)}\left(y_{j}-y_{\alpha}\right)^{2}\right) \tag{3.8}
\end{equation*}
$$

Fix $\sigma \in \Delta$. We describe separately the off-diagonal and diagonal entries of $B(\sigma)$. For $i, j \in \sigma$ and $i \neq j$ we get

$$
\begin{aligned}
B(\sigma)_{i, j} & =A(\sigma)_{i, j}\left(\prod_{\alpha \in \sigma \backslash\{j\}}\left(y_{j}-y_{\alpha}\right)^{2}\right)\left(\prod_{\alpha \in \mathcal{N}(j) \backslash \sigma}\left(y_{j}-y_{\alpha}\right)^{2}\right) \\
& =\frac{\operatorname{vol}(\sigma)}{y_{j}-y_{i}}\left(\prod_{\alpha \in \sigma \backslash\{j\}} \frac{1}{\left(y_{j}-y_{\alpha}\right)} \prod_{\alpha \in \sigma \backslash\{j\}}\left(y_{j}-y_{\alpha}\right)^{2}\right)\left(\prod_{\alpha \in \mathcal{N}(j) \backslash \sigma}\left(y_{j}-y_{\alpha}\right)^{2}\right) \\
& =\operatorname{vol}(\sigma)\left(\prod_{\alpha \in \sigma \backslash\{i, j\}}\left(y_{j}-y_{\alpha}\right)\right)\left(\prod_{\alpha \in \mathcal{N}(j) \backslash \sigma}\left(y_{j}-y_{\alpha}\right)^{2}\right)
\end{aligned}
$$

And for the diagonal entries

$$
\begin{aligned}
B(\sigma)_{j, j} & =A(\sigma)_{j, j}\left(\prod_{\alpha \in \mathcal{N}(j)}\left(y_{j}-y_{\alpha}\right)^{2}\right) \\
& =\operatorname{vol}(\sigma)\left(\prod_{\alpha \in \sigma \backslash\{j\}} \frac{1}{\left(y_{j}-y_{\alpha}\right)}\right)\left(1-\sum_{\alpha \in \sigma \backslash\{j\}} \frac{1}{\left(y_{j}-y_{\alpha}\right)}\right)\left(\prod_{\alpha \in \mathcal{N}(j)}\left(y_{j}-y_{\alpha}\right)^{2}\right) \\
& =\operatorname{vol}(\sigma)\left(\prod_{\alpha \in \sigma \backslash\{j\}}\left(y_{j}-y_{\alpha}\right)-\sum_{k \in \sigma \backslash\{j\}} \prod_{\alpha \in \sigma \backslash\{j, k\}}\left(y_{j}-y_{\alpha}\right)\right)\left(\prod_{\alpha \in \mathcal{N}(j) \backslash \sigma}\left(y_{j}-y_{\alpha}\right)^{2}\right) .
\end{aligned}
$$

Given a polynomial $f \in \mathbb{R}\left[y_{1}, \ldots, y_{m}\right]$, we can rewrite $f=\sum_{i=0}^{d_{j}} f_{i} y_{j}^{i}$ as a univariate polynomial in $y_{j}$ of degree $d_{j}$, where $f_{i} \in \mathbb{R}\left[y_{i}: i \neq j\right]$ is a constant with respect to $y_{j}$. We then define the initial form of $f$ with respect to $j$ to be

$$
\operatorname{in}_{j}(f)=f_{d_{j}} y_{j}^{d_{j}}
$$

We observe that for the off-diagonal entries $B(\sigma)_{i, j}$, the initial form with respect to $j$ is

$$
\operatorname{in}_{j}\left(B(\sigma)_{i, j}\right)=y_{j}^{2 \gamma_{j}-d-1}
$$

where $\gamma_{j}=|\mathcal{N}(j)|$ is the number of vertices adjacent to $j$ in $\Delta$. Whereas for the diagonal entry $B(\sigma)_{j, j}$, the initial form is

$$
\operatorname{in}_{j}\left(B(\sigma)_{j, j}\right)=y_{j}^{2 \gamma_{j}-d}
$$

In both cases, the degree of the initial form is the degree of the polynomial. We sum the matrices $B(\sigma)$ for $\sigma \in \Delta$, to get $B$ and note that the coefficient of the monomial $y_{j}^{2 \gamma_{j}-d}$ in $B_{j, j}$ is the number of simplices
in $\Delta$ containing vertex $j$. Hence, using the Leibniz formula to compute the determinant of $B$, we get that the product of diagonal entries is a polynomial of degree $\left(\sum_{j=1}^{n} 2 \gamma_{j}-d\right)$. All off-diagonal entries in that column of $B$ are of degree one smaller, thus any monomial in the expanded form of the determinant with off-diagonal entries must have degree at least two smaller than the product of diagonal entries. The following equality is a direct consequence of (3.8)

$$
\operatorname{det}(B)=\operatorname{det}(A) \prod_{j=1}^{n}\left(\prod_{\alpha \in \mathcal{N}(j)}\left(y_{j}-y_{\alpha}\right)^{2}\right)
$$

Since $\operatorname{det}(B)$ is not identically $0, \operatorname{det}(A)$ is not identically zero, hence $A$ is invertible over the field of rational functions.

The proof of Theorem 3.3.1 inspires the following conjecture about the combinatorial properties of the determinant.

Conjecture 3.3.3. The highest degree component of the numerator of $\operatorname{det}(A)$ is

$$
\prod_{j=1, \ldots, n}\left(\sum_{\sigma \in \Delta \text { s.t. }} \operatorname{vol}(\sigma) \prod_{\alpha \in \sigma}\left(y_{j}-y_{\alpha}\right)\right)
$$

Since $A$ is invertible, (3.6) can be rewritten as

$$
e^{y}=A^{-1} w
$$

where entries of $A$ are rational functions in $\mathbb{R}\left(y_{1}, \ldots, y_{m}\right)$.
Corollary 3.3.4. Fix a maximal triangulation $\Delta$. Then the critical equations (3.5) can be written in the form

$$
\begin{gather*}
e^{y_{1}}=p_{1}\left(y_{1}, y_{2}, \ldots, y_{m}\right) \\
e^{y_{2}}=p_{2}\left(y_{1}, y_{2}, \ldots, y_{m}\right)  \tag{3.9}\\
\vdots \\
e^{y_{m}}=p_{n}\left(y_{1}, y_{2}, \ldots, y_{m}\right)
\end{gather*}
$$

where $p_{1}, \ldots, p_{n} \in \mathbb{R}\left(y_{1}, \ldots, y_{m}\right)$. If $x_{1}, \ldots, x_{m} \in \mathbb{Q}^{d}$, then $p_{1}, \ldots, p_{n} \in \mathbb{Q}\left(y_{1}, \ldots, y_{m}\right)$.
We will explore rational-exponential systems of the form (3.9) further in Sections 3.3.3-3.3.4. The following is a result from transcendental number theory, for a textbook reference see Theorem 1.4 of [8].

Theorem 3.3.5 (Lindemann-Weierstrass). If $y_{1}, \ldots, y_{r}$ are distinct algebraic numbers then the numbers $e^{y_{1}}, \ldots, e^{y_{r}}$ are linearly independent over the algebraic numbers.

A special case of the Lindemann-Weierstrass theorem is the Lindemann theorem which states that $e^{y}$ is transcendental for algebraic $y \neq 0$.

Theorem 3.3.6. Assume that $X \subset \mathbb{Q}^{d}$. If $\operatorname{vol}(\operatorname{conv}(X)) \neq 1$ then at least one coordinate of the optimal height vector $y^{*}$ is transcendental. If $\operatorname{vol}(\operatorname{conv}(X))=1$, then all coordinates of $y^{*}$ are algebraic if and only if $w$ is in the cone over the secondary polytope $\Sigma(X)$.

Proof. It follows from the proof of [79, Lemma 3.4] that any relevant $y^{*} \in \mathbb{R}^{m}$ such that $e^{h_{X, y^{*}}}$ is a density, is a critical point of $S_{\Delta}\left(y_{1}, \ldots, y_{m}\right)$ for a maximal regular triangulation $\Delta$ and some weight vector $w$. We consider the rational-exponential system (3.9) for this choice of $\Delta$ and $w$. Then we have $e^{y_{1}}=p_{1}\left(y_{1}, \ldots, y_{m}\right)$ where $p_{1}$ is a rational function in $\mathbb{Q}\left(y_{1}, \ldots, y_{m}\right)$. Assume that $y_{1}, \ldots, y_{m}$ are algebraic. By Lindemann's theorem $e^{y_{1}}$ is algebraic if and only if $y_{1}=0$.

However, $p\left(y_{1}, \ldots, y_{m}\right)$ is always algebraic, since $y_{1}, \ldots, y_{m}$ are algebraic and the algebraic numbers form a field. Hence $y_{1}=0$. We can argue similarly that $y_{i}=0$ for all $i$. The vector $y=(0, \ldots, 0)$ belongs to the boundary of the Samworth body if and only if the volume of the convex hull of $X$ is 1 . In this case, $y$ is the optimal solution for any $w$ in the cone over the secondary polytope $\Sigma(X)$ by $[79$, Corollary 3.9].

### 3.3.2 An Alternative Proof of Invertibility for the Simplex Case

In this section, we describe an alternative proof of theorem 3.3.1, for the special case when the convex hull of the point configuration $X$ forms a full-dimensional simplex $\Delta=\left\{x_{1}, \ldots, x_{m}\right\} \subset \mathbb{R}^{m-1}$. This proof was suggested to me by professor R. Liu.

Lemma 3.3.7. The critical equation matrix of a single full-dimensional simplex $\Delta=\operatorname{conv}\left(\left\{x_{1}, \ldots, x_{m}\right\}\right) \subset$ $\mathbb{R}^{m-1}$ is invertible for all $m \in \mathbb{N}$.

Proof. We specialize the objective function $S_{\Delta}$ to a simplex $\Delta$ in $\mathbb{R}^{m-1}$ of volume 1:

$$
f(y)=w \cdot y-\sum_{i=1}^{m} \frac{e^{y_{i}}}{\prod_{j \neq i}\left(y_{i}-y_{j}\right)} .
$$

We compute the $i$-th critical function, the partial of $f$ with respect to $y_{i}$. For the summands for $i$ from 1 to $m$, we have:

$$
\begin{aligned}
\frac{\partial}{\partial y_{i}}\left(\frac{e^{y_{i}}}{\prod_{j \neq i}\left(y_{i}-y_{j}\right)}\right) & =\left(\frac{\partial e^{y_{i}}}{\partial y_{i}}\right)\left(\frac{1}{\prod_{j \neq i}\left(y_{i}-y_{j}\right)}\right)+e^{y_{i}} \frac{\partial}{\partial y_{i}}\left(\frac{1}{\prod_{j \neq i}\left(y_{i}-y_{j}\right)}\right) \\
& =\left(\frac{e^{y_{i}}}{\prod_{j \neq i}\left(y_{i}-y_{j}\right)}\right)+e^{y_{i}}\left(\frac{-\sum_{k=1, k \neq j}^{m}\left(y_{i}-y_{1}\right) \ldots\left(\overline{y_{i}-y_{k}}\right) \ldots\left(y_{i}-y_{n}\right)}{\prod_{j \neq i}\left(y_{i}-y_{j}\right)^{2}}\right) \\
& =\left(\frac{e^{y_{i}}}{\prod_{j \neq i}\left(y_{i}-y_{j}\right)}\right)\left(1-\sum_{j=1, j \neq i}^{m} \frac{1}{\left(y_{i}-y_{j}\right)}\right) .
\end{aligned}
$$

The second term is

$$
\frac{\partial}{\partial y_{i}}\left(\sum_{j=1, j \neq i}^{m}\left(\frac{e^{y_{j}}}{\prod_{k \neq j}\left(y_{j}-y_{k}\right)}\right)\right)=-\sum_{j=1, j \neq i}^{m}\left(\frac{e^{y_{j}}}{\left(y_{j}-y_{i}\right) \prod_{k \neq j}\left(y_{j}-y_{k}\right)}\right) .
$$

From here we can construct the matrix $A$, we change the notation to emphasize that we pay particular attention the columns. We distinguish entries on the diagonal

$$
A_{j, j}=\left(\prod_{\alpha=1, \alpha \neq j}^{m} \frac{1}{\left(y_{j}-y_{\alpha}\right)}\right)\left(1-\sum_{\alpha=1, \alpha \neq j}^{m} \frac{1}{\left(y_{j}-y_{\alpha}\right)}\right)
$$

and off the diagonal for $i \neq j$

$$
\begin{equation*}
A_{i, j}=\left(\prod_{\alpha=1, \alpha \neq j}^{m} \frac{1}{\left(y_{j}-y_{\alpha}\right)}\right)\left(\frac{-1}{y_{j}-y_{i}}\right) \tag{3.10}
\end{equation*}
$$

We introduce a matrix $B$ that is obtained by multiplying the $j$-th column of $A$ by $\prod_{\alpha=1, \alpha \neq j}^{m}\left(y_{j}-y_{\alpha}\right)$ for all $j$ from 1 to $m$. Because scaling the column of a matrix scales the determinant by that same amount, we get

$$
\begin{align*}
\operatorname{det}(B) & =\prod_{j=1}^{m} \prod_{\alpha \neq j}^{m}\left(y_{j}-y_{\alpha}\right) \operatorname{det}(A)  \tag{3.11}\\
& =(-1)^{m} \prod_{i=1}^{m} \prod_{i<j}^{m}\left(y_{i}-y_{j}\right)^{2} \operatorname{det}(A) \tag{3.12}
\end{align*}
$$

The matrix $B$ we obtain has entries

$$
B_{j, j}=1-\sum_{\alpha \neq j}^{m} \frac{1}{y_{j}-y_{\alpha}}, \quad \quad B_{i, j}=\frac{-1}{y_{j}-y_{i}}
$$

We subtract the identity matrix from $B$ to get rid of the 1 's on the diagonal, $B^{\prime}=B-I d_{m}$.

$$
B_{j, j}=-\sum_{\alpha \neq j}^{m} \frac{1}{y_{j}-y_{\alpha}}
$$

We then observe that the sum of all rows of $B^{\prime}$ is the zero vector since, for all $j$ from 1 to $m$

$$
\begin{equation*}
B_{j, j}^{\prime}=\sum_{i=1, i \neq j}^{m} B_{i, j}^{\prime} \tag{3.13}
\end{equation*}
$$

We want the characteristic polynomial of $B^{\prime}$ to be zero. Let $L$ be the Vandermonde matrix associated to $\left(y_{1}, \ldots, y_{m}\right)$.

$$
L=\left[\begin{array}{ccccc}
1 & y_{1} & y_{1}^{2} & \ldots & y_{1}^{m-1} \\
1 & y_{2} & y_{2}^{2} & \ldots & y_{2}^{m-1} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
1 & y_{m} & y_{m}^{2} & \ldots & y_{m}^{m-1}
\end{array}\right]
$$

We want to show that $B^{\prime}$ has all eigenvalue equal to zero, in other words, the characteristic polynomial
associated to that matrix is $\lambda^{m}$.
First we look at $B^{\prime} L$.

$$
\begin{aligned}
\left(B^{\prime} L\right)_{i, j} & =\sum_{l=1}^{m}\left(B_{i, l}^{\prime}\right) y_{l}{ }^{j-1} \\
& =\left(B_{i, i}^{\prime}\right) y_{i}^{j-1}+\sum_{\substack{l=1 \\
l \neq k}}^{m}\left(B_{i, l}^{\prime}\right) y_{l}^{j-1} \\
& =\sum_{\substack{l=1 \\
l \neq i}}^{m} \frac{-y_{i}^{j-1}}{y_{i}-y_{l}}+\sum_{\substack{l=1 \\
l \neq i}}^{m}\left(\frac{-y_{l}^{j-1}}{y_{l}-y_{i}}\right) \\
& =\sum_{\substack{l=1 \\
l \neq i}}^{m} \frac{y_{l}^{j-1}-y_{i}^{j-1}}{y_{i}-y_{l}}
\end{aligned}
$$

In particular, we get $B^{\prime} L_{i, 1}=0$ and $B^{\prime} L_{i, 2}=-(m-1)$ for all $i \in[1, \ldots, m]$. For $j$ larger than 1 , we can re-write

$$
\begin{aligned}
\left(B^{\prime} L\right)_{i, j} & =\sum_{\substack{l=1 \\
l \neq i}}^{m} \frac{y_{l}^{j}-y_{i}^{j}}{y_{i}-y_{l}} \\
& =-\sum_{\substack{l=1 \\
l \neq i}}^{m} \frac{\left(y_{i}-y_{l}\right)\left(\sum_{k=1}^{j} y_{l}^{j-k} y_{i}^{k}\right)}{y_{i}-y_{l}} \\
& =-\sum_{\substack{l=1 \\
l \neq i}}^{m}\left(\sum_{k=1}^{j} y_{l}^{j-k} y_{i}^{k}\right) \\
& =-\sum_{\substack{l=1 \\
l \neq i}}^{m} \sum_{\substack{\alpha, \beta \geq 0 \\
\alpha+\beta=j-1}} y_{i}^{\alpha} y_{l}^{\beta} .
\end{aligned}
$$

Hence the matrix $B^{\prime} L$ is of the form

$$
B^{\prime} L=\left[\begin{array}{ccccc}
0 & -m & -m y_{1}-y_{2}-y_{3} \ldots & \cdots & \sum_{l=2}^{m}\left(\sum_{k=0}^{m-2} y_{l}^{m-2-k} y_{1}^{k}\right) \\
0 & -m & -m y_{2}-y_{1}-y_{3} \ldots & \cdots & \sum_{\substack{l=1 \\
l \neq 2}}^{m-2}\left(\sum_{k=0}^{m} y_{l}^{m-2-k} y_{2}^{k}\right) \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
0 & -m & -m y_{m}-y_{1}-y_{2} \ldots & \cdots & \sum_{l=1}^{m}\left(\sum_{k=0}^{m-2} y_{l}^{m-2-k} y_{m}^{k}\right)
\end{array}\right] .
$$

Let $P_{k}\left(y_{1}, \ldots, y_{m}\right)=\sum_{i=1}^{m} y_{i}^{k}$. Let $C$ be the $(m) \times(m)$ matrix with the following entries

$$
C_{k, j}=\left\{\begin{array}{ll}
-P_{j-(k+1)}\left(y_{1}, \ldots, y_{m}\right) & : j>k+1  \tag{3.14}\\
-(m-j) & : j=k+1 . \\
0 & : j \leq k
\end{array} .\right.
$$

We compute $L C$

$$
\begin{aligned}
(L C)_{i, j} & =\sum_{k=1}^{m} L_{i, k} C_{k j} \\
& =\sum_{k=1}^{j-2} L_{i, k} C_{k j}+L_{i, j-1} C_{j-1, j}+\sum_{k=j}^{m} L_{i, k} C_{k, j} \\
& =\sum_{k=1}^{j-2} y_{i}^{k} P_{j-(k+1)}\left(y_{1}, \ldots, y_{m}\right)-y_{i}^{j-1}(m-j)+\sum_{k=j}^{m} 0 \cdot C_{k, j} \\
& =-\left(\sum_{k=1}^{j-2} y_{i}^{k}\left(y_{0}^{(j-1)-k}+\ldots+y_{m}^{(j-1)-k}\right)+(m-j) y_{i}^{j-1}\right) \\
& =-\left(\sum_{l=1}^{m} \sum_{k=1}^{j-2} y_{i}^{k} y_{l}^{(j-1)-k}\right)-(m-j) y_{i}^{j-1} \\
& =-\left(\sum_{l=1}^{m} \sum_{\substack{\alpha \geq 0, \beta \geq 1 \\
\alpha+\beta=j-1}} y_{i}^{\alpha} y_{l}^{\beta}\right)-(m-j) y_{i}^{j-1} \\
& =-\left(\sum_{\substack{l=1 \\
l \neq i}}^{m} \sum_{\substack{\alpha \geq 0, \beta \geq 1 \\
\alpha+\beta=j-1}} y_{i}^{\alpha} y_{l}^{\beta}\right)-(j) y_{i}^{j-1}-(m-j) y_{i}^{j-1} \\
& =-\sum_{\substack{l=1 \\
l \neq i}}^{\substack{\alpha, \beta, \beta \geq 0 \\
\alpha+\beta=j-1}} y_{i}^{\alpha} y_{l}^{\beta} \\
& =\left(B^{\prime} L\right)_{i, j} .
\end{aligned}
$$

We showed that $L C=B^{\prime} L$, in other words $C=L^{-1} B^{\prime} L$ and thus the characteristic polynomial of $B^{\prime}$ is equal to $\lambda^{n}$, the characteristic polynomial of $C$. Therefore, the score equation matrix of $\Delta$ is invertible over the field of rational function $\mathbb{R}\left(y_{1}, \ldots y_{m}\right)$.

### 3.3.3 One cell in one dimension

In this section we apply the invertibility of the score equation matrix to give a closed form solution to log-concave maximum likelihood estimator in case the logarithm of the optimal density is a linear
function on the real line. If $X=\left(x_{1}, x_{2}\right) \subset \mathbb{R}$, then

$$
A=\operatorname{vol}(\sigma)\left[\begin{array}{cc}
\frac{1}{y_{1}-y_{2}}-\frac{1}{\left(y_{1}-y_{2}\right)^{2}} & \frac{1}{\left(y_{1}-y_{2}\right)^{2}} \\
\frac{1}{\left(y_{1}-y_{2}\right)^{2}} & -\frac{1}{y_{1}-y_{2}}-\frac{1}{\left(y_{1}-y_{2}\right)^{2}}
\end{array}\right]
$$

and

$$
A^{-1}=\frac{1}{\operatorname{vol}(\sigma)}\left[\begin{array}{cc}
1+y_{1}-y_{2} & 1 \\
1 & 1-y_{1}+y_{2}
\end{array}\right]
$$

Hence the polynomial-exponential system (3.9) has the form

$$
\begin{align*}
& e^{y_{1}}=\frac{1}{\operatorname{vol}(\sigma)}\left(\left(1+y_{1}-y_{2}\right) w_{1}+w_{2}\right)  \tag{3.15}\\
& e^{y_{2}}=\frac{1}{\operatorname{vol}(\sigma)}\left(w_{1}+\left(1-y_{1}+y_{2}\right) w_{2}\right) \tag{3.16}
\end{align*}
$$

Dividing (3.15) by (3.16) and setting $y_{12}=y_{1}-y_{2}$, gives

$$
\begin{equation*}
e^{y_{12}}=\frac{\left(1+y_{12}\right) w_{1}+w_{2}}{w_{1}+\left(1-y_{12}\right) w_{2}} \tag{3.17}
\end{equation*}
$$

In the rest of the section we will discuss how to solve Equation (3.17) using Lambert functions. The solutions for $y_{1}$ and $y_{2}$ can then be obtained from Equations (3.15) and (3.16) by solving for $y_{12}$.

Definition 3.3.8 (Section 2 in [66]). For $x, t_{i}, s_{j} \in \mathbb{R}$, consider the function

$$
e^{x} \frac{\left(x-t_{1}\right)\left(x-t_{2}\right) \ldots\left(x-t_{n}\right)}{\left(x-s_{1}\right)\left(x-s_{2}\right) \ldots\left(x-s_{m}\right)}
$$

We denote its (generally multi-valued) inverse function at the point $a \in \mathbb{R}$ by

$$
W\left(t_{1}, t_{2}, \ldots, t_{n} ; s_{1}, s_{2}, \ldots, s_{m} ; a\right)
$$

and call it the generalized $\boldsymbol{W}$-Lambert function. The function $W(a):=W(0 ; ; a)$ is called the usual W-Lambert function.

We have $W(; ; a)=\log (a)$.
Proposition 3.3.9. The tent poles corresponding to a single-cell triangulation in 1 dimension are given by:

$$
\begin{gathered}
y_{1}=\log \left(w_{1} W\left(\rho+1 ;-\rho^{-1}-1 ;-\rho\right)+w_{1}+w_{2}\right)-\log (\operatorname{vol}(\sigma)), \\
y_{2}=\log \left(-w_{2} W\left(\rho+1 ;-\rho^{-1}-1 ;-\rho\right)+w_{1}+w_{2}\right)-\log (\operatorname{vol}(\sigma))
\end{gathered}
$$

where $\rho=w_{1} / w_{2}$ and $W\left(\rho+1 ;-\rho^{-1}-1 ;-\rho\right)$ is a value of the multi-valued generalized Lambert $W$ function if $y_{1} \neq y_{2}$. Otherwise $y=(-\log (\operatorname{vol}(\sigma)),-\log (\operatorname{vol}(\sigma)))$.

Proof. Recall from Equation (3.17):

$$
e^{y_{12}}=\frac{w_{1} y_{12}+w_{1}+w_{2}}{-w_{2} y_{12}+w_{1}+w_{2}}
$$

or, by setting $\rho=w_{1} / w_{2}$, equivalently

$$
\frac{y_{12}-\rho-1}{y_{12}+\rho^{-1}+1} e^{y_{12}}=-\rho .
$$

Seen as an equation in $y_{12}$ this has solutions given by the generalized Lambert function $W\left(\rho+1 ;-\rho^{-1}-\right.$ $1 ;-\rho$ ). The solutions for $y_{1}$ and $y_{2}$ can then be obtained from (3.15) and (3.16) by solving $y_{12}$.

Remark 3.3.10. Proposition 3.3.9 generalizes to the case when we have $n$ points on a line and the optimal tent function is supported on one cell.


Figure 3.5 Generalized Lambert function $W\left(\rho+1 ;-\rho^{-1}-1 ;-\rho\right)$.

The generally multi-valued generalized $W$-Lambert function $W\left(\rho+1 ;-\rho^{-1}-1 ;-\rho\right)$ is plotted in Figure 3.5. We explore its branches, i.e., single-valued functions of $\rho$, using $r$-Lambert functions.

Definition 3.3.11 (Section 3.2 in [66]). If $r \in \mathbb{R}$, consider the function

$$
x e^{x}+r x .
$$

We denote its inverse function in the point $a \in \mathbb{R}$ by $W_{r}(a)$ and call it the $r$-Lambert function.
The following theorem makes the connection between the generalized Lambert function and the $r$-Lambert function:

Theorem 3.3.12 (Theorem 3 in [66]). If $t, s, a \in \mathbb{R}$, the following equality holds:

$$
\left.W(t ; s ; a)=t+W_{\left.-a e^{-t}\right)}\left(a e^{-t}\right)(t-s)\right) .
$$

Hence

$$
W\left(\rho+1 ;-\rho^{-1}-1 ;-\rho\right)=\rho+1+W_{\rho e^{-\rho-1}}\left(-\rho e^{-\rho-1}\left(\rho+\rho^{-1}+2\right)\right) .
$$

The number of branches of the $r$-Lambert function is classified in [66, Theorem 4] and [62, Theorem 4]. For $r=\rho e^{-\rho-1}$, it translates to

1. two branches, if $\rho e^{-\rho-1}<0$;
2. three branches, if $0<\rho e^{-\rho-1}<e^{-2}$;
3. one branch, if $\rho e^{-\rho-1} \geq e^{-2}$.

The second case happens when $\rho>0$, in which case we have the double branch of constant zero function and an additional branch. This is the branch that is relevant to us in the context of Proposition 3.3.9. The first case happens when $\rho<0$, in which case there exists a double branch of the constant zero function. This cannot appear for positive weights $w_{i}$. The third case does not happen.

The $r$-Lambert function can be computed with the C++ implementation [65]. Alternatively, one can use results about computing roots of polynomial-exponential equations. In [61], a symbolic-numeric algorithm is proposed for constructing explicitly an interval containing all the real roots of a single real polynomial-exponential equation, and counting how many roots are contained in a non-bounded interval. In [77], the decision problem of the existence of positive roots of such functions is discussed. This subject is strongly related to quantifier elimination [103], and to transcendentality problems [64, 22, 23]. The latter problem of the transcendence theory appears in our Theorem 3.3.6.

### 3.3.4 Two cells in one dimension

Let $X=\left(x_{1}, x_{2}, x_{3}\right) \subset \mathbb{R}$. Then

$$
A=\left[\begin{array}{ccc}
\frac{v_{1}}{\left(y_{1}-y_{2}\right)^{2}}-\frac{v_{1}}{y_{1}-y_{2}} & -\frac{v_{1}}{\left(y_{1}-y_{2}\right)^{2}} & 0 \\
-\frac{v_{1}}{\left(y_{1}-y_{2}\right)^{2}} & \frac{v_{1}}{\left(y_{1}-y_{2}\right)^{2}}-\frac{v_{1}}{y_{1}-y_{2}}+\frac{v_{2}}{\left(y_{2}-y_{3}\right)^{2}}-\frac{v_{2}}{y_{2}-y_{3}} & -\frac{v_{2}}{\left(y_{2}-y_{3}\right)^{2}} \\
0 & -\frac{v_{2}}{\left(y_{2}-y_{3}\right)^{2}} & \frac{v_{2}}{\left(y_{2}-y_{3}\right)^{2}}-\frac{v_{2}}{y_{2}-y_{3}}
\end{array}\right]
$$

Recall $y_{12}=y_{1}-y_{2}$ and $y_{23}=y_{2}-y_{3}$. Then

$$
A^{-1}=\frac{1}{v_{1}\left(1+y_{23}\right)+v_{2}\left(1-y_{12}\right)}\left[\begin{array}{ccc}
-\left(1+y_{12}\right)\left(1+y_{23}\right)+\frac{v_{2}}{v_{1}} y_{12}^{2} & -1-y_{23} & -1 \\
-1-y_{23} & \left(-1+y_{12}\right)\left(1+y_{23}\right) & -1+y_{12} \\
-1 & -1+y_{12} & -\left(-1+y_{12}\right)\left(-1+y_{23}\right)+\frac{v_{1}}{v_{2}} y_{23}^{2}
\end{array}\right]
$$

Consider the polynomial-exponential system $e^{y}=A^{-1} w$ as in (3.9). Dividing the first equality with
the second one and the second one with the third one gives:

$$
\left\{\begin{array}{l}
e^{y_{12}}=\frac{\left(-\left(1+y_{12}\right)\left(1+y_{23}\right)+\frac{v_{2}}{v_{1}} y_{12}^{2}\right) w_{1}+\left(-1-y_{23}\right) w_{2}-w_{3}}{\left(-1-y_{23}\right) w_{1}+\left(-1+y_{12}\right)\left(1+y_{23}\right) w_{2}+\left(-1+y_{12}\right) w_{3}},  \tag{3.18}\\
e^{y_{23}}=\frac{\left(-1-y_{23}\right) w_{1}+\left(-1+y_{12}\right)\left(1+y_{23}\right) w_{2}+\left(-1+y_{12}\right) w_{3}}{-w_{1}+\left(y_{12}-1\right) w_{2}-\left(\left(y_{12}-1\right)\left(y_{23}-1\right)+\frac{v_{1}}{v_{2}} y_{23}^{2}\right) w_{3}} .
\end{array}\right.
$$

Hence we could reduce a polynomial-exponential system with three equations and three variables to a polynomial-exponential system with two equations and two variables. Systems of two rational bivariate polynomial-exponential equations such as (3.18) are studied in [61]. An algorithm giving the number of solutions of such a system is provided, where all the solutions are contained in a generalized open rectangle of type $I_{1} \times I_{2} \subset \mathbb{R}^{2}$, under the hypothesis that at least one of the intervals $I_{1}$ or $I_{2}$ is bounded.

Remark 3.3.13. Let $X \subset \mathbb{R}$. If we consider tent functions $h_{X, y}$ that are supported on two cells such that $h_{X, y}$ is a constant function on one of the two cells, then one can use methods similar to the one cell case (see Section 3.3.3) to give the optimal solution using the Lambert function.

### 3.3.5 Optimization for the One Dimensional Case

Several algorithms have been developed to compute the log-concave MLE in one dimension [82] and in higher dimensions [27, 5, 76]. Software implementations include R packages such as logcondens [36] and cnmlcd [59] in one dimension, and LogConcDEAD [25] and fmlogcondens [75] in higher dimensions. Nevertheless, we formulate our own algorithm to the log-concave MLE of an arbitrary one-dimensional weighted sample using convex optimization.

For a given set of points $X=\left\{x_{1}, \ldots, x_{m}\right\} \subset \mathbb{R}$, there exists a unique induced maximal subdivision, which is $\Delta=\left\{\left\{x_{1}, x_{2}\right\}, \ldots,\left\{x_{m-1}, x_{m}\right\}\right\}$.

Lemm 3.3.14. Let $\Delta$ be the maximal subdivision $\Delta=\left\{\left\{x_{1}, x_{2}\right\}, \ldots,\left\{x_{m-1}, x_{m}\right\}\right\}$. The maximum loglikelihood estimate for the sample ( $X, w$ ) can be obtained by maximizing $S_{\Delta}$ subject to y being contained in the polyhedral cone

$$
\begin{equation*}
C=\left\{y_{i} \geq \frac{y_{i+1}-y_{i-1}}{x_{i+1}-x_{i-1}}\left(x_{i}-x_{i-1}\right)+y_{i-1}\right\}_{i \in[2, \ldots, m]} . \tag{3.19}
\end{equation*}
$$

Proof. Recall from [79] that finding the log-likelihood estimate is equivalent to solving the following optimization problem

$$
\begin{equation*}
\max _{y \in \mathbb{R}^{m}} w \cdot y-\int_{P} e^{\left(h_{X, y}(t)\right)} d t . \tag{3.20}
\end{equation*}
$$

where $h_{X, y}: \mathbb{R} \rightarrow \mathbb{R}$ is the least concave function satisfying $h_{X, y}\left(x_{i}\right) \geq y_{i}$ for all $i=1, \ldots, m$. We observe that the optimal solution must satisfy the inequalities

$$
\begin{equation*}
\left\{y_{i} \geq \frac{y_{i+1}-y_{i-1}}{x_{i+1}-x_{i-1}}\left(x_{i}-x_{i-1}\right)+y_{i-1}\right\}_{i \in[2, \ldots, m]}, \tag{3.21}
\end{equation*}
$$

thus adding that polyhedral cone as a constraint does not affect the optimal solution. In order to solve the original optimization problem, we use the equivalent formulation from [79] of maximizing 3.3

$$
\begin{equation*}
S\left(y_{1}, \ldots, y_{m}\right)=w \cdot y-\sum_{\sigma \in \Delta} \sum_{i \in \sigma} \frac{\operatorname{vol}(\sigma) \cdot \exp \left(y_{i}\right)}{\prod_{\alpha \in \sigma \backslash i}\left(y_{i}-y_{\alpha}\right)} \tag{3.22}
\end{equation*}
$$

where $y$ is contained in the secondary cone of $\Delta$.
In the one-dimensional case, $\Delta=\left\{\left\{x_{1}, x_{2}\right\}, \ldots,\left\{x_{m-1}, x_{m}\right\}\right\}$ is the unique maximal triangulation, all $y$ inducing a coarser subdivision $\Delta^{\prime}$ are contained in the boundary of the secondary cone over $\Delta$. Hence, adding the polyhedral constraints $C$ restricts the search space to only height vectors $y$ inducing a relevant tent function, and does not exclude any potentially valid solution to the original optimization problem.


Figure 3.6 The tent function and log-concave MLE on a sample of 100 points from a $\operatorname{Beta}(2,2)$ distribution.


Figure 3.7 The tent function and log-concave MLE on a sample of 100 points from a Laplace $(0,4)$ distribution.

Example 3.3.15. We test our optimization method of computing the log-concave MLE on small examples sampled from well known log-concave distributions. In figure 3.6, we sample 100 point from the Beta
distribution with parameters 2, 2 and in figure 3.7, we sample from the Laplace distribution with parameters 0,4 . In either cases, we set $\Delta$ to be the maximal subdivision on the interval of interest and we maximize $S\left(y_{1}, \ldots, y_{m}\right)$ using the built-in NLPSolve command in Maple. We display the tent function and associated log-concave probability density function associated to the numerical optimal solution $y$ *.

### 3.4 Certifying solutions with Smale's $\alpha$-theory

As explained in Section 3.2, our task is to maximize the objective function $S\left(y_{1}, \ldots, y_{m}\right)$ defined in Corollary 3.2.1. For a subdivision $\Delta$, we can find the optimal $y_{\Delta}^{*}$ by considering $S_{\Delta^{\prime}}\left(y_{1}, \ldots, y_{m}\right)$ for any maximal triangulation $\Delta^{\prime}$ that refines $\Delta$, substituting $y_{i}$ that can be expressed in terms of other $y$ 's for the subdivision $\Delta$ and solving the system of critical equations $\partial \widetilde{S}_{\Delta} / \partial y_{i}=0$ for the resulting function $\widetilde{S}_{\Delta}$. For maximal triangulations, we have $\widetilde{S}_{\Delta}=S_{\Delta}$ and the system of critical equations is given by (3.5). We will write $S_{\Delta}$ instead of $\widetilde{S}_{\Delta}$ also when talking about general subdivisions and for brevity we denote the system of critical equations by $\nabla S_{\Delta}(y)=0$. We say the system is square because we have $n$ equations $\partial S_{\Delta} / \partial y_{i}=0$ in $n$ variables $y_{1}, \ldots, y_{m}$. Usually it will be impossible to write down exact solutions to these systems, but there is a way forward. In what follows we discuss the computation of certified solutions to this system of equations. To do so, we discuss Smale's $\alpha$-theory, which makes mathematically rigorous the idea of approximate zeros in the sense of quadratic convergence of Newton iterations. The following influential definition was given in [18, 94].

Definition 3.4.1 (Chapter 8 of [18]). Let $D f(x)$ be the $n \times n$ Jacobian matrix of the square system of complex-analytic equations $f(x)=0 \in \mathbb{C}^{n}$, where $f: \mathbb{C}^{n} \rightarrow \mathbb{C}^{n}$ is written as a column vector of its component functions

$$
f(x)=\left[f_{1}\left(x_{1}, \ldots, x_{m}\right), \ldots, f_{n}\left(x_{1}, \ldots, x_{m}\right)\right]^{T} .
$$

A point $z \in \mathbb{C}^{n}$ is an approximate zero of $f$ if there exists a zero $z^{*} \in \mathbb{C}^{n}$ of $f$ such that the sequence of Newton iterates

$$
z_{k+1}=z_{k}-D f\left(z_{k}\right)^{-1} f\left(z_{k}\right)
$$

satisfies

$$
\left\|z_{k+1}-z^{*}\right\| \leq \frac{1}{2}\left\|z_{k}-z^{*}\right\|^{2}
$$

for all $k \geq 1$ where $z_{0}=z$. If this holds, then we call $z^{*}$ the associated zero of $z$. Here $\|x\|:=\left(\sum_{i=1}^{n} x_{i} \overline{x_{i}}\right)^{\frac{1}{2}}$ is the standard norm in $\mathbb{C}^{n}$, and the zero $z^{*}$ is assumed to be nonsingular, meaning that $\operatorname{det} D f\left(z^{*}\right) \neq 0$.

Therefore the problem becomes two-fold. Given a system of equations $f$, we need a way to (1) generate approximate solutions, and (2) certify their quadratic convergence under Newton iterations. The methods of Smale's $\alpha$-theory solve exactly this second problem. This is accomplished using the constants $\alpha(f, x), \beta(f, x)$ and $\gamma(f, x)$, which we will discuss in Section 3.4.1. Typically $\gamma$ is difficult to compute, since it is defined as the supremum of infinitely many quantities depending on higher-order derivatives of our system of equations. However, explicit upper bounds on $\gamma$ were calculated in [49] which we can specialize to the system required for log-concave density estimation. These upper bounds
have the advantage that they are easily computed from our system $\nabla S=0$, and can therefore be used to $\alpha$-certify approximate solutions coming from numerical software. In Section 3.4.1, we make this precise, discussing recent work on the subject $[49,50,91,94]$ and how it applies in our context.

Remark 3.4.2. One might wonder why we do not directly evaluate the equations in question to the approximate height values given by statistical packages. The reason is that we want to have a measure of how accurate this solution is, which is also very sensitive to the system. Consider for example the system consisting of the single polynomial $f(x)=x$. We would not accept $1 / 2$ as a solution. But if we consider the system $f(x)=x^{10}$ and we evaluate at $x=1 / 2$, we get a value that is less than 0.001 . This could have been tempting, but note that in both cases the difference between actual solution and approximation is the same.

Another example that illustrates the potential difficulties involved in judging a numerical solution based on its evaluation into the original system of equations comes from [11]. Consider the univariate polynomial

$$
f(z)=z^{10}-30 z^{9}+2 .
$$

A solution which is accurate within $9.4 \times 10^{-12}$ of the true solution is

$$
z^{*}=30.00000000000142-0.00000000000047 i
$$

but evaluating the polynomial at this solution yields a complex number $f\left(z^{*}\right)$ with norm $\left|f\left(z^{*}\right)\right|=31.371$, which certainly seems far from zero. However, refining the accuracy of this solution to

$$
z^{* *}=29.9999999999998983894731343124+0.0000000000000000000000062 i,
$$

we find that $\left|f\left(z^{* *}\right)\right|=0.00000000032$, which is much better.

### 3.4.1 Smale's $\alpha$-theory

The intuition behind $\alpha$-theory is as follows. The size of the initial Newton iteration step combined with the size of the derivatives control how quickly Newton iteration converges to a true solution. We can calculate the size of the Newton iteration step, so if we have some control over the higher order derivatives of $f$, then we should be able to certify whether a solution satisfies the criterion of Definition 3.4.1. This motivates the definition of the following constants $\alpha, \beta, \gamma \in \mathbb{R}$, associated to a system of equations $f$ at a point $x$. These constants measure quantities relevant to certifying approximate zeros.

Definition 3.4.3. Let $f: \mathbb{C}^{n} \rightarrow \mathbb{C}^{n}$ be a system of complex-analytic functions and let $x \in \mathbb{C}^{n}$. We define $\alpha(f, x)$ to be the product of $\beta(f, x)$ and $\gamma(f, x)$ :

$$
\alpha(f, x)=\beta(f, x) \gamma(f, x) .
$$

The constant $\beta(f, x)$ measures the size of the Newton iteration step applied at $x$, namely:

$$
\beta(f, x)=\left\|D f(x)^{-1} f(x)\right\|
$$

while $\gamma(f, x)$ bounds the sizes of the following quantities, involving the higher order derivatives:

$$
\gamma(f, x)=\sup _{k \geq 2}\left\|\frac{D f(x)^{-1} D^{k} f(x)}{k!}\right\|^{\frac{1}{k-1}}
$$

If we can compute these constants $\beta, \gamma$ for a candidate solution, then we can utilize the following
Theorem 3.4.4 (Chapter 8 of [18]). If $f: \mathbb{C}^{n} \rightarrow \mathbb{C}^{n}$ is a system of complex-analytic functions and $x \in \mathbb{C}^{n}$ satisfies

$$
\alpha(f, x)<\frac{13-3 \sqrt{17}}{4} \approx 0.157671
$$

then $x$ is an approximate zero of $f=0$.
For polynomial systems, all higher-order derivatives eventually vanish. Exactly this fact was used in [91] to derive an upper bound for $\gamma(f, x)$ which involves the degrees of the polynomials in the system $f$. This is highly convenient since, even for systems of polynomials, calculating $\gamma(f, x)$ purely based on the definition is quite a difficult task. Yet, if we are to certify candidate solutions to our system of equations, we need to calculate $\gamma$ and $\beta$ at our candidate $x$, multiply them, and hope they are below $\approx 0.157671$.

### 3.4.2 Polynomial-exponential systems

For polynomial-exponential systems $f$, calculating $\gamma(f, x)$ is even harder. However, in [49], an upper bound was computed for $\gamma$ involving quantities more readily apparent in a given system $f$ than what appears in the bare definition of $\gamma$. In fact, an upper bound for $\gamma$ is calculated which applies to a general class of systems, as well as upper bounds for several special cases. One of these special cases can be further specialized to the system of equations $\nabla S=0$ arising in log-concave density estimation (this is Lemma 3.4.9 below). In [49] an example is given where the bounds for the special cases allowed candidate solutions to be $\alpha$-certified despite failure using the more general bounds. In this section we summarize the results of [49] as they relate to log-concave density estimation. First we need a few definitions.

Definition 3.4.5. For a point $x \in \mathbb{C}^{n}$ define

$$
\|x\|_{1}^{2}=1+\|x\|^{2}=1+\sum_{i=1}^{n}\left|x_{i}\right|^{2}
$$

For a polynomial $g: \mathbb{C}^{n} \rightarrow \mathbb{C}$ given as $g(x)=\sum_{|\rho| \leq d} a_{\rho} x^{\rho}$ define

$$
\|g\|^{2}=\frac{1}{d!} \sum_{|\rho| \leq d} \rho!\cdot(d-|\rho|)!\cdot\left|a_{\rho}\right|^{2}
$$

For a polynomial system $f: \mathbb{C}^{n} \rightarrow \mathbb{C}^{n}$ with $f(x)=\left[f_{1}(x), \ldots, f_{n}(x)\right]^{T}$, we define

$$
\|f\|^{2}=\sum_{i=1}^{n}\left\|f_{i}\right\|^{2}
$$

We now define a quantity $\mu(f, x)$ associated to a polynomial system which will play a role in bounding $\gamma$ later.

Definition 3.4.6. Let $f: \mathbb{C}^{n} \rightarrow \mathbb{C}^{n}$ be a polynomial system with $\operatorname{deg} f_{i}=d_{i}$. Define

$$
\mu(f, x)=\max \left\{1,\|f\| \cdot\left\|D f(x)^{-1} C_{f}(x)\right\|\right\}
$$

where $C_{f}(x)$ is the diagonal matrix

$$
C_{f}(x)=\left[\begin{array}{lll}
d_{1}^{1 / 2} \cdot\|x\|_{1}^{d_{1}-1} & & \\
& \ddots & \\
& & d_{n}^{1 / 2} \cdot\|x\|_{1}^{d_{n}-1}
\end{array}\right]
$$

Following [49], we extend Definition 3.4.6 to certain polynomial-exponential systems.
Definition 3.4.7. Let $a \in \mathbb{Z}_{\geq 0}, \delta_{i} \in \mathbb{C}$, and $\sigma_{i} \in\{1, \ldots, n\}$. Consider the polynomial-exponential system

$$
G\left(x_{1}, \ldots, x_{m}, u_{1}, \ldots, u_{a}\right)=\left[\begin{array}{c}
P\left(x_{1}, \ldots, x_{m}, u_{1}, \ldots, u_{a}\right)  \tag{3.23}\\
u_{1}-e^{\delta_{1} x_{\sigma_{1}}} \\
u_{2}-e^{\delta_{2} x_{\sigma_{2}}} \\
\vdots \\
u_{a}-e^{\delta_{a} x_{\sigma_{a}}}
\end{array}\right]
$$

where $P: \mathbb{C}^{N} \rightarrow \mathbb{C}^{n}$ is a polynomial system with $N=n+a$ variables. Thus, the system $G$ is a square system of size $N$. We write $X:=(x, u)$. Define

$$
\mu(G, X)=\max \left\{1,\left\|D G(x, u)^{-1}\left[\begin{array}{cc}
C_{P}(x, u)\|P\| & \\
& I_{a}
\end{array}\right]\right\|\right\}
$$

The following specializes Corollary 2.6 of [49].
Theorem 3.4.8. Let $a \in \mathbb{Z}_{\geq 0}, \delta_{i} \in \mathbb{C}$, and $\sigma_{i} \in\{1, \ldots, n\}$ and consider the polynomial-exponential system (3.23). Let $d_{i}=\operatorname{deg} P_{i}$ and $D=\max d_{i}$. For any $\lambda, \theta \in \mathbb{C}$ define

$$
A(\lambda, \theta)=\max \left\{|\lambda|,\left|\frac{\lambda^{2} e^{\lambda \theta}}{2}\right|\right\}
$$

Then, for any $X=(x, u) \in \mathbb{C}^{N}$ such that the Jacobian of $G$ is invertible,

$$
\begin{equation*}
\gamma(G, X) \leq \mu(G, X)\left(\frac{D^{3 / 2}}{2\|X\|_{1}}+\sum_{i=1}^{a} A\left(\delta_{i}, x_{\sigma_{i}}\right)\right) \tag{3.24}
\end{equation*}
$$

Proof. This is a straight-forward specialization of Corollary 2.6 of [49]. We set to zero quantities that deal with functions not relevant to log-concave density estimation.

Therefore, reformulating our system of polynomial-exponential equations $\nabla S_{\Delta}=0$ in the format (3.23) will allow us to calculate an upper bound on $\gamma$, which will allow us to certify solutions to our critical equations.

Lemma 3.4.9. Fix a maximal regular triangulation $\Delta$. The polynomial-exponential system $\nabla S_{\Delta}=0$ can be reformulated as a system of equations of the form (3.23), demonstrating that Theorem 3.4.8 applies in the context of log-concave maximum likelihood estimation.

Proof. The partial derivatives $\partial S_{\Delta} / \partial y_{k}$ are rational functions of the $y_{i}$ and the $e^{y_{i}}$. Since we set each partial derivative to zero, we can clear denominators, creating a system of equations, each of which is a polynomial in the $y_{i}$ and the $e^{y_{i}}$. Setting each $\delta_{i}=1$ in (3.23), we can replace each occurrence of $e^{y_{i}}$ with $u_{i}$, creating the polynomial system $P\left(y_{1}, \ldots, y_{m}, u_{1}, \ldots, u_{n}\right)$, hence $a=n$ as well. Appending the equations $u_{i}-e^{y_{i}}$ to the system of polynomials $P$, we have a system of $2 n$ equations in $2 n$ unknowns. This system is of the required form in order to apply Theorem 3.4.8.

Thus, we have everything we need to compute the upper bound in (3.24) for a system of critical equations $\nabla S_{\Delta}=0$ when $\Delta$ is a maximal regular triangulation. By calculating this upper bound for a given system of equations, we can certify approximate numerical solutions obtained in any way. When $\Delta$ is not a maximal regular triangulation, one must impose further linear constraints on some of the $y_{i}$, as was the case in Example 3.2.5. After simplifications, one might still end up with terms involving exponentials of fractional convex combinations of the $y_{i}$. This poses no threat for the purposes of $\alpha$-certification, as one may in fact use products of exponentials of the form $e^{\beta y_{i}}$. In particular, a bound for $\gamma(G, X)$ also for these more general polynomial-exponential systems is given in [49, Corollary 2.6].

Question 3.4.10. In algebraic statistics, it is common to find algebraic invariants which characterize algebraic complexity. For example, the maximum likelihood degree of a statistical model gives information about the critical points of the likelihood function of a parametric model [1]. Similarly, in nonparametric algebraic statistics, it could be the case that the combinatorial complexity of the optimal subdivision gives us information about the computational complexity of finding a numerical solution. If this is true, then we would expect that increasing the combinatorial complexity will decrease the likelihood that the numerical output from LogConcDEAD is $\alpha$-certified. We test this hypothesis experimentally in the next section. In future work, one could hope to precisely describe this phenomenon, should it exist. Of course, higher degrees, more variables, more equations will always increase the bound on $\gamma$ we calculate, but the combinatorics should still play some role.

### 3.4.3 A procedure for $\alpha$-certifying

One of our motivating questions was to determine the correct subdivision for a given data set, as was the case in Example 3.1.1. In this section we describe a procedure based on Smale's $\alpha$-theory that in principle allows us to find the certifiably correct subdivision. Recall that the objective function $S\left(y_{1}, \ldots, y_{m}\right)$ depends on a subdivision of the convex hull of the data set $X$. If there are $m$ subdivisions, then there are $m$ different objective functions $S_{1}, \ldots, S_{m}$, and $m$ different possible systems of equations $\nabla S_{1}=0, \ldots, \nabla S_{m}=0$. Given an estimate of a solution $y^{*}$, perhaps computed numerically using existing software, we can attempt to $\alpha$-certify that solution using any of these systems as input to Lemma 3.4.9 and Theorem 3.4.8. As we collect $\alpha$-certified critical points for the various objective functions, we can use this data to determine the correct subdivision, helping to answer our motivating question.

In practice, we have found that numerically computed solutions $y^{*}$ are often not $\alpha$-certified, using any of the systems $\nabla S_{i}=0$. However, using a brute-force search over all possible additional digits, we often can find one system $\nabla S_{j}=0$ to which $y^{*}+\varepsilon$ is an $\alpha$-certified solution. Here, $\varepsilon=\left(\varepsilon_{1}, \ldots, \varepsilon_{n}\right)$ is a vector providing additional digits of precision to each component of $y^{*}$. As we compute $\alpha$-values for each $y^{*}+\varepsilon$, we move in the direction which causes a decrease in the computed $\alpha$-value, until we are able to find an $\alpha$-certified $y^{*}+\varepsilon$. We describe this in the following

Input: A system $\nabla S_{i}=0$ coming from the $i$ th candidate subdivision and a candidate approximate solution $y^{*}=\left(y_{1}, \ldots, y_{m}\right)$.
Result: A refinement of the heights $y^{*}+\varepsilon$ along with alpha certification of the system, or inability to certify.
1 Let $p$ be the number of trusted significant digits (in binary) of the approximate solution $y^{*}$.;
2 Expressing $y^{*}$ in binary, compute the $\alpha$-value for all $3^{n}$ points $y_{i}+\epsilon_{i} 2^{-p}, \epsilon_{i} \in\{-1,0,1\}$. Keep the point with the lowest alpha value, and set this as the new $y_{i}$;
3 If the alpha value is below 0.157671 stop and return the solution. If it has decreased between steps or remained the same, increase $p$ by 1 and go to step 2 . If there is no improvement for several loops in a row, stop and declare inability to certify the system. ;

Remark 3.4.11. Here we collect a few comments on Algorithm 3.

1. We note that this brute-force search over all possible digits could be replaced by any numerical procedure for finding solutions to a given set of equations. For example, Newton iteration could be used on the system of equations to produce more accurate solutions, which could then be $\alpha$-certified. However, to compare the outputs of LogConcDEAD for problems of increasing combinatorial complexity (see Table 3.1), we wanted to use a completely "blind" brute-force search as described above.
2. One does not need to stop at Step 3 once a solution is certified. Repeating the loop allows increasing the precision of the solution by moving to lower $\alpha$ values. This is in contrast to statistical software like LogConcDEAD which only allows up to 7 significant digits.
3. Although precision can be added, our (first) goal with Algorithm 3 is to find the correct subdivision induced by the heights. One can test several subdivisions here, therefore we say that we test the (approximate) solution against the corresponding system of equations.
4. It might happen that the $\alpha$-value does not immediately decrease from one loop to the next even if we have the correct system of equations. One reason is that if the next significant digit is a zero for all heights, we are computing an $\alpha$-value for the same point multiple times.
5. In step 1 of the above algorithm, we let $p$ be the number of trusted significant digits of the approximate solution $y^{*}$. We have found that several of the last digits of a solution computed with LogConcDEAD were incorrect, in the sense that if we start our search (in Algorithm 3) earlier in the significant digits of $y^{*}$ we are able to $\alpha$-certify some $y^{*}+\varepsilon$. In this way, we can correct for some of the imprecision of a numerical solver.

Example 3.4.12. Consider the data set $X=(2,5,7)$ with weights $w=\left(\frac{1}{3}, \frac{1}{2}, \frac{1}{6}\right)$. With this input, the package LogConcDEAD returns the heights

$$
y^{*}=\left(y_{1}, y_{2}, y_{3}\right)=(-1.454152,-1.605833,-1.888083)
$$

suggesting that there are two regions of linearity (Figure 3.8a). Let $\Delta=\{12,23\}$. We consider critical equations for

$$
S_{\Delta}\left(y_{1}, y_{2}, y_{3}\right)=\frac{y_{1}}{3}+\frac{y_{2}}{2}+\frac{y_{3}}{6}-3 \frac{\mathrm{e}^{y_{1}}-\mathrm{e}^{y_{2}}}{y_{1}-y_{2}}-2 \frac{\mathrm{e}^{y_{2}}-\mathrm{e}^{y_{3}}}{y_{2}-y_{3}}
$$

which lead to the polynomial-exponential system $\nabla S_{\Delta}: \mathbb{C}^{3} \rightarrow \mathbb{C}^{3}$ given by

$$
\begin{aligned}
\left(y_{1}-y_{2}\right)^{2} \frac{\partial S\left(y_{1}, y_{2}, y_{3}\right)}{\partial y_{1}} & =0 \\
\left(y_{1}-y_{2}\right)^{2}\left(y_{2}-y_{3}\right)^{2} \frac{\partial S\left(y_{1}, y_{2}, y_{3}\right)}{\partial y_{2}} & =0 \\
\left(y_{2}-y_{3}\right)^{2} \frac{\partial S\left(y_{1}, y_{2}, y_{3}\right)}{\partial y_{3}} & =0
\end{aligned}
$$

where we have cleared denominators. The numerical solution from LogConcDEAD is not immediately $\alpha$-certified, but after applying Algorithm 3 we obtain the $\alpha$-certified solution: $y^{*}+\varepsilon=\left(y_{1}, y_{2}, y_{3}\right)=$ ( $-1.45415181,-1.60583278,-1.88808307$ ) .

Example 3.4.13. We now consider the same sample $X=(2,5,7)$ with uniform weights. As discussed in Example 3.2.5, LogConcDEAD output suggests that the logarithm of the optimal density has a single region of linearity (Figure 3.8c). Can we certify this assessment? Recall that substituting $y_{2}=\frac{2}{5} y_{1}+\frac{3}{5} y_{3}$ to $S\left(y_{1}, y_{2}, y_{3}\right)=\frac{1}{3} y_{1}+\frac{1}{3} y_{2}+\frac{1}{3} y_{3}-3 \frac{e^{y_{1}}-e^{y_{2}}}{y_{1}-y_{2}}-2 \frac{e^{y_{2}}-e^{y_{3}}}{y_{2}-y_{3}}$ gives

$$
\widetilde{S}=\frac{7}{15} y_{1}+\frac{8}{15} y_{3}-5 \frac{e^{y_{1}}-e^{y_{3}}}{y_{1}-y_{3}}
$$

The system of equations $\nabla \widetilde{S}=0$ does have solutions, and we were able to check that the numerical
solution $y^{*}$ computed by LogConcDEAD is an $\alpha$-certified solution to this amended system of equations.


Figure 3.8 The height functions for (a) Example 3.4.12; (b) Example 3.4.13; (c) Example 3.4.15

Example 3.4.14. We used Algorithm 3 to certify the sample $X=(0,1,2, \ldots, n) \subset \mathbb{R}$ for weights given by the binomial distribution with $p=6 / 11$, i.e., $w_{i}=\binom{n}{i}(6 / 11)^{i}(5 / 11)^{n-i}$. Looking at the LogConcDEAD output, we suspect that the triangulation given by the points consists of all consecutive line segments $\{i-1, i\}$ for $i \in 1,2, \ldots, n$. We therefore compute $\alpha$-values using the system of equations corresponding to the full triangulation. In all cases tested, we were able to certify the system for some refinement of the original LogConcDEAD output. In Table 3.1, we summarize the number of binary digits required for certification in each case. This table suggests that the complexity of $\alpha$-certifying increases when the number of sample points increases.

Table 3.1 Number of binary digits needed to certify $m+1$ points with weights coming from an asymmetric binomial distribution.

| n | 3 | 4 | 5 | 6 | 7 |
| :---: | :---: | :---: | :---: | :---: | :---: |
| binary digits | 22 | 23 | 27 | 31 | 31 |

We now present an example in two dimensions that needs more significant digits than the previous cases.

Example 3.4.15. We consider the point configuration from [79, Example 1.1], given by

$$
X=((0,0),(0,100),(22,37),(36,41),(43,22),(100,0)) \subset \mathbb{R}^{2}
$$

and uniform weights. The package LogConcDEAD returns the heights

$$
\left(y_{1}, y_{2}, y_{3}, y_{4}, y_{5}, y_{6}\right)=(-8.789569,-8.772087,-8.253580,-8.217959,-8.236983,-8.756922)
$$

as the optimal solution. This gives rise to a triangulation of the convex hull of the data points with regions of linearity consisting of the triangles

$$
\{1,2,3\},\{1,3,5\},\{1,5,6\},\{2,3,4\},\{2,4,6\},\{3,4,5\},\{4,5,6\},
$$

in Figure 3.8b. This data gives an $\alpha$-value of $10^{26}$, which is much larger than the required 0.157671 . However, the system of equations it came from has a relatively high degree and the polynomial equations, when expanded, have between 929 and 1564 terms. We try to decrease the $\alpha$-value using the uniform sampling algorithm described above. We create a list of $729=3^{6}$ points in $\mathbb{R}^{6}$, consisting of all points whose i-th coordinate is

$$
y_{i}+\epsilon_{i} 2^{-14}, \epsilon_{i} \in\{-1,0,1\}
$$

After a few repetitions, this finds a point with a lower alpha value. We repeat this process, each time decreasing the exponent of 2 when creating the new test points. After 95 rounds we detect the refined point

$$
\left[\begin{array}{l}
-8.789570552675578322471018111262921 \\
-8.772086862481395608253513836856700 \\
-8.253580886913590521217040193671505 \\
-8.217957742357924329528595494315867 \\
-8.236983233544571734253428918807660 \\
-8.756919956247208359690046164738877
\end{array}\right]
$$

with alpha value 0.125519 . Therefore, this new solution is $\alpha$-certified. Note that this number has 34 decimal digits; we have rounded digits coming from the conversion from base 2 ( 109 digits) after this position. Our conclusion is that the triangulation obtained by the heights in the LogConcDEAD output is certifiably correct.

Example 3.4.16. We return to our motivating example 3.1.1 from the introduction, and consider two possible subdivisions of $P=\operatorname{conv}(X)$ for the regions of linearity of the optimal tent function:

$$
\Delta_{1}=\{\{1,2,3\},\{1,3,4\},\{2,3,4\},\{2,4,12\},\{1,4,8,11\},\{4,11,12\},\{8,11,12,13,14\}\}
$$

and

$$
\Delta_{2}=\{\{1,2,3\},\{1,3,4\},\{2,3,4,12\},\{1,4,8,12,13,14\}\} .
$$

The first subdivision $\Delta_{1}$ in Figure 3.9 a arises from the LogConcDEAD output with default parameters after using the "unique" function. The second subdivision $\Delta_{2}$ in Figure 3.9 b is given by the four regions of linearity in Figure 3.1 that we get by adjusting the precision in LogConcDEAD and then using the "unique" function. Unfortunately the objective functions involved have too many summands for $\alpha$-certification to be feasible.

As an alternative, we use the NMaximize command in Mathematica directly on the objective functions $S_{\Delta_{1}}$ and $S_{\Delta_{2}}$. The optimal $y_{\Delta_{1}}^{*}$ for the 7-cell subdivision gives a tent function whose regions of


Figure 3.9 Subdivisions in Example 3.4.16. (a) Subdivision $\Delta_{1}$ (b) Subdivision $\Delta_{2}$ (c) The subdivision induced by $y_{\Delta_{1}}^{*}$.
linearity are

$$
\{\{1,2,3\},\{1,8,13\},\{1,3,13\},\{2,3,14\},\{3,13,14\}\}
$$

which are depicted in Figure 3.9c. This triangulation is not refined by the subdivision $\Delta_{1}$ : For example, the triangle $\{1,3,4\}$ in the subdivision $\Delta_{1}$ intersects the interiors of triangles $\{1,3,13\},\{2,3,14\},\{3,13,14\}$. Thus the 7-cell subdivision $\Delta_{1}$ is not the subdivision that we are looking for. In fact, the vector $y_{\Delta_{1}}^{*}$ is not relevant, i.e. there exists $x_{i}$ such that $h_{X, y_{\Delta_{1}}^{*}}\left(x_{i}\right)>y_{i}$, and as a result $\int_{P} e^{h_{X, y_{\Delta_{1}}^{*}}(t)} \neq 1$.

The command NMaximize gives for the 4-cell subdivision

$$
\begin{aligned}
y_{\Delta_{2}}^{*}= & (-4.32285,-4.7141,-4.2737,-4.14495,-4.26961,-4.10156,-3.94188 \\
& -3.91671,-3.94162,-3.80042,-3.76397,-3.68413,-3.69541,-3.62252)
\end{aligned}
$$

In comparison, the optimal height vector that we obtain using LogConcDEAD is

$$
\begin{aligned}
y^{*}= & (-4.322797,-4.714126,-4.273678,-4.144934,-4.269616,-4.101524,-3.941869 \\
& -3.916668,-3.941666,-3.800423,-3.764006,-3.684179,-3.695395,-3.622560)
\end{aligned}
$$

A computation in Polymake verifies that $y_{\Delta_{2}}^{*}$ gives a tent function whose regions of linearity are exactly the cells of $\Delta_{2}$. This suggests that the 4-cell subdivision $\Delta_{2}$ is indeed the subdivision induced by the optimal $y^{*}$ in Example 3.1.1. We conclude with a haiku.

Approximate heights,
subdivisions inexact.
A long road ahead.

## Chapter 4

## Conic Hulls of Moment Curves and Connections to Coalescence Manifolds

The content of this chapter is the result of a joint project with Z. Rosen and C. Vinzant. The preprint has been submitted for peer review and is accessible at: https://arxiv.org/abs/2012.14467. We investigate a univariate moment problem of piecewise-constant density functions on the interval $[0,1]$ and its consequences for an inference problem in population genetics. We show that, up to closure, any collection of $n$ moments is achieved by a step function with at most $n-1$ breakpoints and that this bound is tight. We use this to show that any point in the $n$th coalescence manifold in population genetics can be attained by a piecewise constant population history with at most $n-2$ changes. Both the moment cones and the coalescence manifold are projected spectrahedra and we describe the problem of finding a nearest point on them as a semidefinite program.

### 4.1 Background and Motivation

Given a finite collection $A \subset \mathbb{N}$, we consider the convex cone $M(A)$ of all moments $\left(m_{a}\right)_{a \in A}$ of the form $m_{a}=\int x^{a} d \mu$ where $\mu$ is a nonnegative Borel measure on the unit interval $[0,1]$. For consecutive moments $A=\{0,1,2 \ldots, d\}$, this is a classical object in analysis and real algebraic geometry. The problem of determining membership in the cone $M(A)$ is known as the truncated Haussdorff moment problem. See, for example, [51, 17, 58, 63, 87].

We study moments coming from piecewise-constant density functions with the idea of minimizing the number of pieces needed. Formally, we consider the set $M_{k}(A)$ as the closure of the set of moments $\left(m_{a}\right)_{a \in A}$ where $m_{a}=\int_{0}^{1} x^{a} f(x) d x$ and $f$ is a nonnegative step function with at most $k$ discontinuities.

One of our motivations for studying this problem came from its relation to the coalescence manifold studied in [80]. The coalescence manifold $C_{n, k}$, formally defined in Section 4.4, is a set of summary statistics in population genetics, derived from observing $n$ genomes with a population history consisting of $k+1$ different population sizes.

The authors in [80] show that the manifold $C_{n, k}$ stabilizes at $k=2 n-2$, i.e. $C_{n, 2 n-2}=C_{n, k}$ for all
$k \geq 2 n-2$. The main theorem of section 4.4 improves this bound by a factor of two, showing that the coalescence manifolds stabilize at $k=n-2$ and this bound is tight.

The content of section 4.6 relies on the connections to the classical moment problem to provide a description of $C_{n, n-2}$ as the projection of a spectrahedron. The problem of finding the nearest point in $C_{n, n-2}$ to a given point in $\mathbb{R}^{n-1}$ can then be formulated as a semidefinite program.

### 4.2 Moments of step functions

For $k \in \mathbb{N}$, let $S_{k}$ denote the set of nonnegative step functions on $[0,1]$ of the form

$$
\begin{equation*}
f=y_{1} \mathbf{1}_{\left[0, s_{1}\right]}+\sum_{i=2}^{k+1} y_{i} \mathbf{1}_{\left(s_{i-1}, s_{i}\right]} \tag{4.1}
\end{equation*}
$$

where $0=s_{0}<s_{1}<\ldots<s_{k}<s_{k+1}=1$ and $y_{1}, \ldots, y_{k+1} \in \mathbb{R}_{\geq 0}$. Note that:

1. $S_{k}$ is invariant under nonnegative scaling,
2. $S_{k} \subseteq S_{\ell}$ when $k \leq \ell$, and
3. $S_{k}+S_{\ell}$, defined as $\left\{f+g \mid f \in S_{k}, g \in S_{\ell}\right\}$, is a subset of $S_{k+\ell}$.

Elements of $S_{k}$ define nonnegative measures on $[0,1]$. We will be interested in the possible moments of these measures. Given a finite collection $A \subset \mathbb{N}$, we define to be the Euclidean closure of the set moments given by density functions in $S_{k}$ :

$$
M_{k}(A)=\overline{\left\{\left(\int_{0}^{1} x^{a} f(x) d x\right)_{a \in A}: f \in S_{k}\right\}}
$$

One important case is that of consecutive moments $A=\{0,1, \ldots, d\}$. For any finite collection $A \subset \mathbb{N}$, the moment cone $M_{k}(A)$ can be expressed, up to closure, as the image of $M_{k}(\{0,1, \ldots, \max (A)\})$ under the coordinate projection $\pi_{A}: \mathbb{R}^{\max (A)+1} \rightarrow \mathbb{R}^{A}$ given by $\pi_{A}\left(m_{0}, \ldots, m_{\max (A)}\right)=\left(m_{a}\right)_{a \in A}$.

Remark 4.2.1. By linearity of the integral, we see that $M_{k}(A)$ inherits many properties of $S_{k}$. That is, $M_{k}(A)$ is invariant under nonnegative scaling, $M_{k}(A) \subseteq M_{\ell}(A)$ when $k \leq \ell$ and

$$
M_{k}(A)+M_{\ell}(A) \subseteq M_{k+\ell}(A)
$$

(here, in the sense of the Minkowski sum), as desired.
We will be interested in comparing this to the full moment cone:

$$
M(A)=\overline{\left\{\left(\int_{0}^{1} x^{a} d \mu\right)_{a \in A}: \mu \text { is a nonnegative Borel measure on }[0,1]\right\}}
$$

The cone $M(A)$ is dual to the convex cone of univariate polynomials supported on $A$ that are nonnegative on $[0,1]$, as will be discussed below in Proposition 4.2.6.

When $0 \in A$, the closure in the definition of $M(A)$ is not necessary, and the extreme rays of $M(A)$ are come from point evaluations. That is, we can write $M(A)$ as the conical hull of the image of $[0,1]$ under the corresponding moment map:

$$
M(A)=\operatorname{conicalHull}\left\{v_{A}(t): t \in[0,1]\right\} \quad \text { where } v_{A}(t)=\left(t^{a}\right)_{a \in A} .
$$

See, for example, [87, Prop. 10.5].
When $0 \notin A$, this equality only holds up to closure, as the curve parameterized by $v_{A}(t)$ includes the origin. In this case, $M(A)=\overline{\text { conicalHull }\left\{v_{A}(t): t \in[0,1]\right\}}$. As we will see below, then we can still write $M(A)$ as the conical hull of a curve segment. Specifically, $M(A)=$ conicalHull $\left\{v_{B}(t): t \in[0,1]\right\}$ where $B=\{a-\min (A): a \in A\}$.

Lemma 4.2.2. If $A \subset \mathbb{N}$ is finite and $B=\{a-\min (A): a \in A\}$, then $M(A)=M(B)$.
Proof. For $t \in(0,1]$, the point $v_{A}(t)$ can be rewritten as $t^{\min (A)} v_{B}(t)$, a scalar multiple of $v_{B}(t)$. It follows that the conical hulls of $\left\{v_{A}(t): t \in(0,1]\right\}$ and $\left\{v_{B}(t): t \in(0,1]\right\}$ are equal. We observe that the extreme ray $v_{B}(0)$ of $M(B)$ can be attained in the closure of $M(A)$ as the limit of the moment of the step function $f=\epsilon^{-(\min (A)+1)} \mathbf{1}_{[0, \epsilon]}$ as $\epsilon$ goes to zero:

$$
\lim _{\epsilon \rightarrow 0} \int_{0}^{1} x^{a} f(x) d x=\lim _{\epsilon \rightarrow 0} \epsilon^{-(\min (A)+1)} \int_{0}^{\epsilon} x^{a} d x=\lim _{\epsilon \rightarrow 0} \frac{\epsilon^{a-\min (A)}}{a+1}= \begin{cases}\frac{1}{a+1} & \text { if } a=\min (A) \\ 0 & \text { otherwise }\end{cases}
$$

It follows that $v_{B}(0)$ belongs to $M(A)$. Since $M(B)$ can be written as the union of the cone over $v_{A}(t)$ for $t \in(0,1]$ and the ray over $v_{B}(0)$, the equality between the two cones ensues:

$$
M(A)=\overline{\text { conicalHull }\left\{v_{A}(t): t \in[0,1]\right\}}=M(B) .
$$

Example 4.2.3. Consider $A=\{1,2\}$ and $B=\{0,1\}$. Then

$$
M(B)=\operatorname{conicalHull}\{(1, t): t \in[0,1]\}=\overline{\operatorname{conicalHull}\left\{\left(t, t^{2}\right): t \in[0,1]\right\}}=M(A) .
$$

Here we see the need for taking closures when $0 \notin A$. The point $(1,0)=v_{B}(0)$ is not contained in the the conical hull of the curve segment $\left\{\left(t, t^{2}\right): t \in[0,1]\right\}$ but is contained in its closure. See Figure 4.1. In this case, the boundary of $M(A)=M(B)$ consists of scalar multiples of $v_{B}(0)$ and $v_{B}(1)$, both of which belong to $M_{1}(A)$, by Proposition 4.2.4 below. Arguments below will then show that $M(A)=M_{1}(A)$.

Proposition 4.2.4. Let $A \subset \mathbb{N}$ be finite and let $B=\{a-\min (A): a \in A\}$. The points $v_{B}(0)$ and $v_{B}(1)$ belong to $M_{1}(A)$ and for every $t \in(0,1), v_{B}(t)$ belongs to $M_{2}(A)$.

Proof. For $0<t<1$, and $0<\epsilon<1-t$ consider the step function $f=\epsilon^{-1} t^{-\min (A)} \mathbf{1}_{(t, t+\epsilon]}$ in $S_{2}$. By continuity, the integral $\int_{0}^{1} x^{a} f(x) d x$ converges to $t^{a-\min (A)}$ as $\epsilon \rightarrow 0$, thus $M_{2}(A)$ contains the limit point $v_{B}(t)=\left(t^{b}\right)_{b \in B}$. Similarly, the limit as $\epsilon \rightarrow 0$ of the $A$-moment vectors of step functions


Figure 4.1 The cone $M(A)$ along with the curve segments $v_{A}([0,1])$ and $v_{B}([0,1])$ for $A=\{1,2\}$ and $B=\{0,1\}$.
$f=(\min (A)+1) \epsilon^{-1-\min (A)} \mathbf{1}_{[0, \epsilon]}$ and $f=\epsilon^{-1} \mathbf{1}_{(1-\epsilon, 1]}$ in $S_{1}$ are $v_{B}(0)$ and $v_{B}(1)$, respectively. Therefore these vectors belong to $M_{1}(A)$.

A corollary of this statement is that $M_{k}(A)=M(A)$ for $k=2|A|$. By Carathéodory's Theorem, any point in $M(A)$ is in the conical hull of at most $|A|$ points of the form $\left(t^{a}\right)_{a \in A}$ where $t \in[0,1]$, each of which belongs to $M_{2}(A)$ by Proposition 4.2.4. By Remark 4.2.1, the sum of $|A|$ elements from $M_{2}(A)$ belongs to $M_{2|A|}(A)$, giving $M(A) \subseteq M_{2|A|}(A)$. In fact, $M_{k}(A)$ fills out the whole moment cone much sooner:

Theorem 4.2.5. If $k \geq|A|-1, M_{k}(A)=M(A)$.

The proof of this theorem relies on understanding the points on the boundary of $M(A)$.
Proposition 4.2.6. Let $A \subset \mathbb{N}$ be finite with $0 \in A$. If $\mathbf{m}=\left(m_{a}\right)_{a \in A}$ belongs to the Euclidean boundary of $M(A)$, then any representing measure $\mu$ on $[0,1]$ with $m_{a}=\int x^{a} d \mu$ has finite support. Specifically, the support of $\mu$ is a subset of the roots contained in $[0,1]$ of a polynomial nonnegative on $[0,1]$ and of the form $p(x)=\sum_{a \in A} p_{a} x^{a}$. The vector $\mathbf{m}$ is a conic combination of the vectors $v_{A}(r)$ where ranges over the roots of $p$.

Proof. Let $\ell: \mathbb{R}^{A} \rightarrow \mathbb{R}$ be a linear function $\ell(\mathbf{v})=\sum_{a \in A} p_{a} v_{a}$ defining a supporting hyperplane of $M(A)$ at $\mathbf{m}$. That is, $\ell(\mathbf{v}) \geq 0$ for all $\mathbf{v} \in M(A)$ and $\ell(\mathbf{m})=0$. Consider the polynomial $p(x)=$ $\ell\left(v_{A}(x)\right)=\sum_{a \in A} p_{a} x^{a}$. Since $v_{A}(t) \in M(A)$ for all $t \in[0,1], p$ is nonnegative on [0, 1]. Furthermore, for any measure $\mu$ with moments $\mathbf{m}$,

$$
\int p(x) d \mu=\sum_{a \in A} p_{a} m_{a}=\ell(\mathbf{m})=0
$$

The measure $\mu$ is nonnegative and the polynomial $p$ is nonnegative on $[0,1]$. From this we see that the support of the measure $\mu$ must be contained in the (finite) set of roots $R$ of $p(x)$. Specifically, $\mu=\sum_{r \in R} w_{r} \delta_{r}$ for some $w_{r} \in \mathbb{R}_{\geq 0}$; therefore, $\mathbf{m}=\sum_{r \in R} w_{r} v_{A}(r)$.

Proof of Theorem 4.2.5. First, consider a point $\mathbf{m}$ in the boundary of $M(A)$. By Lemma 4.2.2, $M(A)=$ $M(B)$ where $B=\{a-\min (A): a \in A\}$, and so $\mathbf{m}$ also belongs to the boundary of $M(B)$. By Proposition 4.2.6, $\mathbf{m}$ is the vector of $B$-moments of a measure $\mu$ supported on the roots of a nonnegative polynomial on $[0,1]$ of the form $p(x)=\sum_{a \in B} p_{a} x^{a}$. Let $b$ be the number of distinct roots of $p$ in the set $\{0,1\}$ and $i$ be the number of distinct roots in of $p$ in the open interval $(0,1)$. Then $\mathbf{m}$ is in the conical hull of the $b+i$ points given by $v_{B}(r)$ where $r$ ranges over these roots. By Proposition 4.2.4, $\mathbf{m}$ belongs to $M_{k}(A)$ for $k=b+2 i$.

By Descartes' rule of signs, the number of positive roots of $p$, counting multiplicity, is at most the number of sign changes in the list of coefficients $\left\{p_{a}\right\}_{a \in B}$. If $p_{0} \neq 0$, then $p$ has at most $|B|-1$ roots in $\mathbb{R}_{>0}$. If $p_{0}=0$, then $p$ is the sum of at most $|B|-1$ nonzero terms and its number of roots in $\mathbb{R}_{>0}$ must be smaller or equal to $|B|-2$. Note that every root of $p$ in $(0,1)$ must have even multiplicity greater or equal to 2 . All together this gives $b+2 i \leq|B|-1=|A|-1$.

Now consider $\mathbf{m}$ in the interior of $M(A)$. Let $\mathbf{c}=(1 /(a+1))_{a \in A} \in M_{0}(A)$ denote the vector obtained by integrating against the constant step function of height one. Let $\lambda^{*}$ be the maximum value of $\lambda \in \mathbb{R}$ for which $\mathbf{m}-\lambda \mathbf{c}$ belongs to $M(A)$. From $\mathbf{m} \in M(A)$, we see that $\lambda^{*} \geq 0$. Moreover, since $M(A)$ is pointed, $-\mathbf{c}$ does not belong to $M(A)$, meaning that for sufficiently large $\lambda, \mathbf{m}-\lambda \mathbf{c}$ does not belong to $M(A)$. Since $M(A)$ is closed, it follows that such a maximum $\lambda^{*}$ must exist.

The point $\mathbf{m}-\lambda^{*} \mathbf{c}$ belongs to the boundary of $M(A)$. By the arguments above, $\mathbf{m}-\lambda^{*} \mathbf{c}$ belongs to $M_{k}(A)$ for $k \geq|A|-1$. Since $\mathbf{c} \in M_{0}(A)$ and

$$
\mathbf{m}=\left(\mathbf{m}-\lambda^{*} \mathbf{c}\right)+\lambda^{*} \mathbf{c},
$$

the point $\mathbf{m}$ also belongs to $M_{k}(A)$ for $k \geq|A|-1$.
Remark 4.2.7. It follows from the proof of Theorem 4.2 .5 that for all $k \geq 0, M_{k}(A)$ is star convex with respect to the point $\mathbf{c}=(1 /(a+1))_{a \in A}$, the $A$-moment of the constant function. Indeed, since $\mathbf{c}$ belongs to $M_{0}(A), \lambda \mathbf{c}+M_{k}(A) \subseteq M_{k}(A)$ for all $\lambda \geq 0$.

We can go further in characterizing the facial structure of the boundary of $M(A)$. Through a connection to Schur polynomials, we can deduce linear independence among sets of points from the curve of the correct size.

Proposition 4.2.8. For a collection $A$ of integers $0=a_{1}<a_{2}<\ldots<a_{n}$ and any real values $0 \leq r_{1}<r_{2}<\ldots<r_{n} \leq 1$, the determinant of the matrix $\mathcal{S}_{A}$ is strictly positive, where

$$
\mathcal{S}_{A}(\mathbf{r})=\left(\begin{array}{cccc}
1 & 1 & \ldots & 1 \\
r_{1}^{a_{2}} & r_{2}^{a_{2}} & \ldots & r_{n}^{a_{2}} \\
\vdots & \vdots & & \vdots \\
r_{1}^{a_{n}} & r_{2}^{a_{n}} & \ldots & r_{n}^{a_{n}}
\end{array}\right) .
$$

Proof. By the bialternant formula for Schur polynomials, the determinant of the matrix $\mathcal{S}_{A}(\mathbf{r})$ can be
expressed as

$$
\begin{equation*}
\operatorname{det}\left(\mathcal{S}_{A}\right)=\left(\prod_{1 \leq i<j \leq n}\left(r_{j}-r_{i}\right)\right) \mathfrak{s}_{\lambda}\left(r_{1}, \ldots, r_{n}\right) \text { for } \lambda=\left(a_{n}-(n-1), a_{n-1}-(n-2), \ldots, a_{1}\right), \tag{4.2}
\end{equation*}
$$

where $\mathfrak{s}_{\lambda}\left(x_{1}, \ldots, x_{n}\right)$ denotes the Schur polynomial associated to the partition $\lambda$. By definition, the Schur polynomial $\mathfrak{s}_{\lambda}\left(x_{1}, \ldots, x_{n}\right)$ is the sum of monomials $\mathbf{x}^{T}$ over all semistandard Young tableaux $T$ of shape $\lambda$. One can observe, either from expanding the determinant of $\mathcal{S}_{A}(r)$ along the first column, or by filling out a semistandard Young Tableau of shape $\lambda$ without using the number 1, that $x_{1}$ does not appear in all the monomials of the determinant of $\mathcal{S}_{A}$. It follows that $\operatorname{det}\left(\mathcal{S}_{A}\right)$ is strictly positive for any $0 \leq r_{1}<r_{2}<\ldots<r_{n} \leq 1$.

Corollary 4.2.9. All proper faces of $M(A)$ are simplicial.
Proof. By Lemma 4.2.2, we can assume that $0 \in A$. Recall that $M(A)$ is the conical hull over the curve segment $\left\{v_{A}(t): t \in[0,1]\right\}$ and any proper face $F$ of this cone can be expressed as the conical hull of some points $v_{A}\left(r_{1}\right), \ldots, v_{A}\left(r_{k}\right)$ where $0 \leq r_{1}<r_{2}<\ldots<r_{k} \leq 1$. If $k>\operatorname{dim}(F)$, then there is a subset of these points of size $\operatorname{dim}(F)+1 \leq n$, which necessarily lie in $F$ and are therefore linearly dependent, contradicting Proposition 4.2.8. Therefore $k=\operatorname{dim}(F)$ and $F$ is simplicial.

This lemma lets us assign an index to points on the boundary of $M(A)$, following [87, Ch. 10.2]. Let $\mathbf{m}$ be a point on the boundary of $M(A)$. By Corollary 4.2.9, there is a unique representation of $\mathbf{m}$ as $\sum_{j=1}^{k} w_{j} v_{A}\left(r_{j}\right)$ where $0 \leq r_{1}<\ldots<r_{k} \leq 1$ and $w_{1}, \ldots, w_{k} \in \mathbb{R}_{>0}$. We define the index of $\mathbf{m}$, denoted $\operatorname{ind}(\mathbf{m})$, to be $b+2 i$ where $b=\#\left\{j: r_{j} \in\{0,1\}\right\}$ and $i=\#\left\{j: r_{j} \in(0,1)\right\}$. By Proposition 4.2.4, any point $\mathbf{m}$ on the boundary of $M(A)$ belongs to $M_{\mathrm{ind}(\mathbf{m})}(A)$.

To prove the converse, we must rule out the possibility that $\mathbf{m} \in M_{k}(A)$ for $k<\operatorname{ind}(\mathbf{m})$. In other words, it is impossible to approach a point $\mathbf{m}$ on the boundary of $M(A)$ with moment vectors of step functions with fewer breakpoints than expected.

Lemma 4.2.10. Let $\mathbf{m}$ be a point on the boundary of $M(A)$. For $k<\operatorname{ind}(\mathbf{m}), \mathbf{m} \notin M_{k}(A)$. That is, if $\mathbf{m} \in M_{k}(A)$, then $\operatorname{ind}(\mathbf{m}) \leq k$.

Proof. Note that for any non-zero point $\mathbf{m}$ in $M(A), m_{0}>0$ and so we can rescale $\mathbf{m}$ to have $m_{0}=1$. We will write $M_{k}(A) \cap\left\{m_{0}=1\right\}$ as the image of a compact polytope under a polynomial map and check that any point $\mathbf{m}$ in the image of this map and the boundary of $M(A)$ has index $\leq k$.

Any function $f \in S_{k}$ can be written as $f=y_{1} \mathbf{1}_{\left[0, s_{1}\right]}+\sum_{i=2}^{k+1} y_{i} \mathbf{1}_{\left(s_{i-1}, s_{i}\right]}$ for some values

$$
0=s_{0}<s_{1}<\ldots<s_{k}<s_{k+1}=1
$$

and $y_{i} \geq 0$ for all $i$. We now introduce transformed $w$-coordinates by letting $w_{i}=y_{i}\left(s_{i}-s_{i-1}\right)$ denote the area $\int_{s_{i-1}}^{s_{i}} f(x) d x$. The corresponding moment in $M_{k}(A)$ is given by the image of the point
$(\mathbf{s}, \mathbf{w})=\left(s_{1}, \ldots, s_{k}, w_{1}, \ldots, w_{k+1}\right)$ under the polynomial map

$$
\begin{equation*}
\mu_{A}(\mathbf{s}, \mathbf{w})=\left(\sum_{i=1}^{k+1} y_{i} \frac{s_{i}^{a+1}-s_{i-1}^{a+1}}{a+1}\right)_{a \in A}=\left(\sum_{i=1}^{k+1} w_{i} \frac{\left(s_{i}^{a}+s_{i}^{a-1} s_{i-1}+\cdots+s_{i-1}^{a}\right)}{a+1}\right)_{a \in A} \tag{4.3}
\end{equation*}
$$

Note that the constraint that $m_{0}=1$ translates into $\sum_{i} w_{i}=1$. Consider the polytope

$$
\begin{equation*}
P=\left\{(\mathbf{s}, \mathbf{w}) \in \mathbb{R}^{k} \times \mathbb{R}^{k+1} \text { such that } 0 \leq s_{1} \leq \ldots \leq s_{k} \leq 1, w_{i} \geq 0, \sum_{i=1}^{k+1} w_{i}=1\right\}, \tag{4.4}
\end{equation*}
$$

which is a product of two simplices of dimension $k$. The moments of step functions $f \in S_{k}$ with $\int f(x) d x=1$ is the image under $\mu_{A}$ of the set of points $(\mathbf{s}, \mathbf{w}) \in P$ with distinct $0<s_{1}<\ldots<s_{k}<1$. Its closure is $M_{k}(A) \cap\left\{m_{0}=1\right\}$, which necessarily coincides with the image of $P$ under $\mu_{A}$, as the image of a compact set under a continuous map is closed.

If $w_{i}>0$ and $s_{i-1}<s_{i}$ for some $i$, then $\mu_{A}(\mathbf{s}, \mathbf{w})$ has a representing measure whose support includes the interval $\left(s_{i-1}, s_{i}\right]$ and is therefore not finite. Then by Proposition 4.2.6, $\mathbf{m}$ belongs to the interior of $M(A)$.

Suppose the point $\mathbf{m}$ belongs to $M_{k}(A)$. Then $\mathbf{m}=\mu_{A}(\mathbf{s}, \mathbf{w})$ for some $(\mathbf{s}, \mathbf{w}) \in P$. Let $I$ denote the collection of indices $1 \leq i \leq k$ for which $w_{i}>0$. If $\mathbf{m}$ belongs to the boundary of $M(A), s_{i-1}=s_{i}$ for all $i \in I$. Then

$$
\mathbf{m}=\mu_{A}(\mathbf{s}, \mathbf{w})=\sum_{i \in I} w_{i} v_{A}\left(s_{i}\right) .
$$

We can bound $\operatorname{ind}(\mathbf{m})$ by bounding the number of distinct values of $s_{i}$ that appear. For each $i \in I$ with $s_{i} \in(0,1), s_{i}$ equals $s_{i-1}$, hence there are at least two indices $j$ in $\{1, \ldots, k\}$ for which $s_{j}=s_{i}$. Trivially, if $s_{i} \in\{0,1\}$, there is at least one $j \in\{1, \ldots, k\}$ such that $s_{j}=s_{i}$. Together, these show that

$$
\operatorname{ind}(\mathbf{m})=\#\left\{s_{i} \in\{0,1\}: i \in I\right\}+2 \cdot\left(\#\left\{s_{i} \in(0,1): i \in I\right\}\right) \leq k .
$$

Lemma 4.2.11. The intersection of $M_{k}(A)$ with the Euclidean boundary of $M(A)$ is a semialgebraic set of dimension $\leq k$.

Proof. By Lemma 4.2.10, the intersection of $M_{k}(A)$ with the Euclidean boundary of $M(A)$ is the set of boundary points of index $\leq k$. We can parameterize this as the union of the semialgebraic sets:

$$
\bigcup_{\sigma \in\{0,1\}^{2}}\left\{\sum_{j=1}^{\ell} w_{j} v_{A}\left(r_{j}\right)+w_{\ell+\sigma_{1}} v_{A}(0)+w_{\ell+\sigma_{1}+\sigma_{2}} v_{A}(1): \mathbf{r} \in(0,1)^{\ell}, \mathbf{w} \in\left(\mathbb{R}_{>0}\right)^{\ell+\sigma_{1}+\sigma_{2}}\right\},
$$

where in each set, $\ell$ is chosen so that $2 \ell+\sigma_{1}+\sigma_{2} \leq k$. Here we use $\mathbf{r}$ to denote the vector $\left(r_{j}\right)_{j}$ and $\mathbf{w}$ for the vector $\left(w_{j}\right)_{j}$. Note that each set is the image of $(0,1)^{n} \times\left(\mathbb{R}_{>0}\right)^{m}$ under a polynomial map where $n+m \leq k$ and therefore has dimension $\leq k$.

Corollary 4.2.12. If $k<|A|-1, M_{k}(A) \neq M(A)$.

Proof. The cone $M(A)$ is full-dimensional in $\mathbb{R}^{|A|}$, in consequence, the cone's boundary is a hypersurface of dimension $|A|-1$. By Lemma 4.2.11, the dimension of the intersection of $M_{k}(A)$ with the boundary of $M(A)$ has dimension $\leq k$, so for $k<|A|-1$, this cannot be the entire boundary of $M(A)$.

Example 4.2.13. For $A=\{0,2,5\}$, Theorem 4.2.5 and Corollary 4.2.12 imply that $M_{k}(A)=M_{2}(A)$ for all $k \geq 2$ but not for $k=1$. Affine transformations of their intersections with the affine hyperplane $\left\{m_{0}=1\right\}$ are shown in Figure 4.2. See also [80, Figure 6]. The intersection of $M_{1}(A)$ with the boundary of $M_{2}(A)$ consists of just two rays, which appear as points in the hyperplane $\left\{m_{0}=1\right\}$. The set $M_{1}(A)$ consists of moments of functions with just one breakpoint. Step functions with one breakpoint and total mass one can be parameterized by $f=\frac{w}{s} \mathbf{1}_{[0, s]}+\frac{1-w}{1-s} \mathbf{1}_{(s, 1]}$ for $s \in(0,1)$ and $w \in[0,1]$. Note that fixing $w$ and taking the limit as $s \rightarrow 0$ gives a weighted sum of a point mass at zero and the constant function $w \delta_{0}+(1-w) \mathbf{1}_{(0,1]}$. Similarly $s \rightarrow 1$ gives $w \mathbf{1}_{[0,1)}+(1-w) \delta_{1}$.

For $s=w \in[0,1]$, the corresponding step function is constant, i.e. $f=\mathbf{1}_{[0,1]}$ and the moment map sends this line segment to a single point. However, away from this line, the moment map is a homeomorphism to its image in $M_{1}(\{0,2,5\})$.


Figure 4.2 The parameter space of $M_{1}(A)$ and $M_{1}(A), M_{2}(A)$ for $A=\{0,2,5\}$.

### 4.3 Increasing and decreasing step functions

In this section, we study the moment cones of non-negative monotone functions on the unit interval $[0,1]$. They form two convex full dimensional cones contained in the full moment cone $M(A)$ and only require roughly half the number of steps to be generated compared to $M(A)$. So far, the cones of moments of monotone step functions are the only full dimensional subsets of the moment cone whose geometric structure we were able to describe. We define the increasing and decreasing moment cones

$$
\begin{aligned}
& M^{\uparrow}(A)=\overline{\left\{\left(\int_{0}^{1} x^{a} f(x) d x\right)_{a \in A}: f \text { is nonnegative and increasing on }[0,1]\right\}} \text { and } \\
& M^{\downarrow}(A)=\overline{\left\{\left(\int_{0}^{1} x^{a} f(x) d x\right)_{a \in A}: f \text { is nonnegative and decreasing on }[0,1]\right\}} .
\end{aligned}
$$

Recall that if a function $f:[0,1] \rightarrow \mathbb{R}$ is monotone, then it is automatically Borel-measurable. As in the non-monotone case, all of these moment vectors can be achieved as a limit of moments of step functions with a bounded number of steps. For $k \in \mathbb{N}$, let $S_{k}^{\uparrow}$ denote the set of nonnegative, increasing step functions on $[0,1]$ with at most $k$ discontinuities. Similarly, let $S_{k}^{\downarrow}$ denote the analogous set of decreasing step functions. This corresponds to requiring $y_{1} \leq y_{2} \leq \ldots \leq y_{k+1}$ or $y_{1} \geq y_{2} \geq \ldots \geq y_{k+1}$ in (4.1).

Similarly, for finite $A \subset \mathbb{N}$, we consider the $A$-moments of these step functions,

$$
M_{k}^{\square}(A)=\overline{\left\{\left(\int_{0}^{1} x^{a} f(x) d x\right)_{a \in A}: f \in S_{k}^{\square}\right\}} \text { for } \quad \square \in\{\uparrow, \downarrow\} .
$$

Just as with $M_{k}(A)$, we see that the set $M_{k}^{\square}(A)$ is invariant under nonnegative scaling, $M_{k}^{\square}(A) \subseteq M_{\ell}^{\square}(A)$ when $k \leq \ell$ and $M_{k}^{\square}(A)+M_{\ell}^{\square}(A) \subseteq M_{k+\ell}^{\square}(A)$.

As in the non-monotone case, we can understand the cones $M^{\square}(A)$ as the conical hull of curve segments.

Definition 4.3.1. We define maps $\gamma_{A}^{\uparrow}$ and $\gamma_{A}^{\downarrow}$ from $[0,1]$ to $\mathbb{R}^{A}$ where, for $t \in[0,1], \gamma_{A}^{\uparrow}(t)$ and $\gamma_{A}^{\downarrow}(t)$ are the $A$-moment vectors of the step functions $(1 /(1-t)) \mathbf{1}_{(t, 1]}$ and $\left(1 / t^{\min (A)+1}\right) \mathbf{1}_{[0, t]}$, respectively. For every $a \in A$, the $a$ th coordinate of these maps are given by

$$
\begin{aligned}
\left(\gamma_{A}^{\uparrow}(t)\right)_{a} & =\frac{1}{1-t} \int_{t}^{1} x^{a} d x=\frac{1}{a+1} \sum_{i=0}^{a} t^{i} \\
\text { and } \quad\left(\gamma_{A}^{\downarrow}(t)\right)_{a} & =\frac{1}{t^{\min (A)+1}} \int_{0}^{t} x^{a} d x=\frac{1}{a+1} t^{a-\min (A)} .
\end{aligned}
$$

We observe that $\gamma_{A}^{\uparrow}(0)=\gamma_{A}^{\downarrow}(1)=(1 /(a+1))_{a \in A}$ corresponds to the moment vector of constant function $\mathbf{1}_{[0,1]}$. The other end points correspond to point masses. Specifically, $\gamma_{A}^{\uparrow}(1)=v_{A}(1)$ is the moment vector of a point mass at $t=1$ and $\gamma_{A}^{\downarrow}(0)=\frac{1}{\min (A)+1} v_{B}(0)$ for $B=\{a-\min (A): a \in A\}$ corresponds to a point mass at $t=0$.

Remark 4.3.2. The conical hull over $\left\{\gamma_{A}^{\square}(t): t \in[0,1]\right\}$ is closed because this curve is compact and does not contain the origin. Indeed, for $\square=\uparrow$, the $a$ th coordinate of $\gamma_{A}^{\uparrow}(t)$ is $\geq(1 / a+1)$ for all $t$. For $\square=\downarrow$, the $\min (A)$-th coordinate of $\gamma_{A}^{\downarrow}(t)$ is identically $1 /(\min (A)+1)$.

Lemma 4.3.3. For $\square \in\{\uparrow, \downarrow\}$, the cone $M^{\square}(A)$ equals the conical hull of $\left\{\gamma_{A}^{\square}(t): t \in[0,1]\right\}$.
Proof. Since $M^{\square}(A)$ is a convex cone containing the point $\gamma_{A}^{\square}(t)$ for all $t$, it automatically contains the conical hull of this curve.

For the other direction, consider a monotone function $f:[0,1] \rightarrow \mathbb{R}$. We can construct a sequence of step functions $f_{n}$ converging uniformly to $f$ on $[0,1]$. For example, we may take $f_{n}=\sum_{i=1}^{n} \frac{M}{n} \mathbf{1}_{T_{i}}$ where $M \in\{f(0), f(1)\}$ is the maximal value of $f$ on $[0,1]$ and $\mathbf{1}_{T_{i}}$ is the indicator function of $T_{i}=\{x \in[0,1]: f(x) \geq i M / n\}$. That is $f_{n}(x)=\frac{M}{n} \cdot\left\lfloor\frac{n}{M} f(x)\right\rfloor$. Note that $\left|f_{n}-f\right| \leq M / n$ and so $f_{n}$
converges uniformly to $f$ on $[0,1]$. It follows that for any $a, x^{a} f_{n}$ converges uniformly to $x^{a} f$ and so the integral $\int_{0}^{1} x^{a} f_{n}(x) d x$ converges to $\int_{0}^{1} x^{a} f(x) d x$.

Note that the set $T_{i}$ defined above has the form $\left(s_{i}, 1\right]$ or $\left[s_{i}, 1\right]$ if $f$ is increasing and $\left[0, s_{i}\right]$ or $\left[0, s_{i}\right)$ if $f$ is decreasing for some $s_{i} \in[0,1]$. The moment vector of $f_{n}$ therefore is a conic combination of the points $\gamma_{A}^{\mathrm{\square}}\left(s_{i}\right)$ for the appropriate $\square \in\{\uparrow, \downarrow\}$. Taking $n \rightarrow \infty$ shows that the moment vector of $f$ belongs to the closure of the conical hull of $\left\{\gamma_{A}^{\square}(t): t \in[0,1]\right\}$.

Therefore the moment cone $\left\{\left(\int_{0}^{1} x^{a} f(x) d x\right)_{a \in A}: f\right.$ nonnegative and increasing on $\left.[0,1]\right\}$ belongs to the closure of the conical hull of $\left\{\gamma_{A}^{\uparrow}(t): t \in[0,1]\right\}$. By definition, $M^{\uparrow}(A)$ is the closure of this set and so also belongs to the closure of this conical hull. Similarly $M^{\downarrow}(A)$ belongs to the closure of the conical hull of $\left\{\gamma_{A}^{\downarrow}(t): t \in[0,1]\right\}$. By Remark 4.3.2, both of these conical hulls are already closed.

Proposition 4.3.4. If $k \geq\left\lfloor\frac{|A|}{2}\right\rfloor$, then we have $M_{k}^{\uparrow}(A)=M^{\uparrow}(A)$ and $M_{k}^{\downarrow}(A)=M^{\downarrow}(A)$.
Proof. Our proof proceeds similarly to that of Theorem 4.2.5. Let $\mathbf{m}$ be a point of the boundary of $M^{\square}(A)$. We want to express $\mathbf{m}$ as the $A$-moment of an increasing step function of the fewest steps possible. Let $\ell: \mathbb{R}^{A} \rightarrow \mathbb{R}$ define a supporting hyperplane of $M^{\square}(A)$ at $\mathbf{m}$, so that $\ell \geq 0$ on $M^{\square}(A)$ and $\ell(\mathbf{m})=0$. By Lemma 4.3.3, $M^{\square}(A)$ is the conical hull of a curve, hence $\mathbf{m}$ will lie in the conical hull of points on this curve with $\ell=0$. We use this to show that $\mathbf{m}$ belongs to $M_{k}^{\square}(A)$ for $k \geq\left\lfloor\frac{|A|}{2}\right\rfloor$.
$(\downarrow)$ Let $p(x)=\ell\left(\gamma_{A}^{\downarrow}(x)\right)=\sum_{a \in A} \frac{p_{a}}{a+1} x^{a-\min (A)}$. The polynomial $p$ is nonnegative on [0, 1]. By Descartes' rule of signs, $p$ has at most $|A|-1$ positive roots, counting multiplicity, and if $p_{\min (A)}=0$, then it has at most $|A|-2$. Let $i$ denote the number of distinct roots of $p$ in $(0,1)$ and $b=1$ if $p(0)=0$ and 0 otherwise. Since each interior root of $p$ must have multiplicity $\geq 2$, this gives $2 i+b \leq|A|-1$. Note that $\gamma_{A}^{\downarrow}(t) \in M_{1}^{\downarrow}(A)$ for all $t \in[0,1)$ and belongs to $M_{0}^{\downarrow}(A)$ for $t=1$. Therefore $\mathbf{m}$ belongs to $M_{k}^{\downarrow}(A)$ for $k=i+b \leq \frac{1}{2}(|A|-1+b)$. The bound follows from the integrality of $i+b$ and $b \in\{0,1\}$.
( $\uparrow$ ) Let $p(x)=\ell\left(\gamma_{A}^{\uparrow}(x)\right)=\sum_{a \in A} \frac{p_{a}}{a+1} \sum_{i=0}^{a} x^{i}$, which is a polynomial nonnegative on [0,1]. Again, by Descartes' rule of signs, $p$ has at most $|A|-1$ positive roots, counting multiplicity. If $i$ is the number of distinct roots of $p$ in $(0,1)$ and $b=0$ if $p(1)=0$ and 0 otherwise, this gives that $2 i+b \leq|A|-1$. As before, $\gamma_{A}^{\uparrow}(t) \in M_{1}^{\uparrow}(A)$ for all $t \in(0,1]$ and belongs to $M_{0}^{\uparrow}(A)$ for $t=0$. Therefore $\mathbf{m}$ belongs to $M_{k}^{\downarrow}(A)$ for $k=i+b \leq \frac{1}{2}(|A|-1+b) \leq \frac{1}{2}|A|$.

Now consider $\mathbf{m}$ in the interior of $M^{\square}(A)$ and let $\mathbf{c}$ be the moment vector of the constant function $\mathbf{1}_{[0,1]}$. Let $\lambda^{*}$ be the maximum value of $\lambda \in \mathbb{R}$ for which $\mathbf{m}-\lambda \mathbf{c}$ belongs to $M^{\square}(A)$. Since $\mathbf{m} \in M^{\square}(A)$, we know that $\lambda^{*} \geq 0$, and for sufficiently large $\lambda, \mathbf{m}-\lambda \mathbf{c} \notin M^{\square}(A)$. Thus $\mathbf{m}-\lambda^{*} \mathbf{c}$ belongs to the boundary of $M^{\square}(A)$, which is equal to the boundary of $M_{k}^{\square}(A)$ by the argument above. Hence, $\mathbf{m}$ also belongs to $M_{k}^{\square}(A)$.

Proposition 4.3.5. For all $k<\left\lfloor\frac{|A|}{2}\right\rfloor$, the cone $M_{k}^{\square}(A)$ is a proper subset of $M^{\square}(A)$.
Proof. The cone $M_{k}^{\square}(A) \subset \mathbb{R}^{|A|}$ is a conic combination of $k$ points on the boundary curve $\gamma_{A}^{\square}$, each contributing two degrees of freedom, and the point corresponding to the image of the constant step
function $\gamma_{A}^{\uparrow}(0)=\gamma_{A}^{\downarrow}(1)$, contributing a single degree of freedom. Therefore, the semialgebraic set $M_{k}^{\square}(A)$ has dimension at most $\min \{2 k+1,|A|\}$. The cone $M^{\square}(A)$ is full-dimensional in $\mathbb{R}^{|A|}$. Let $n=\lfloor|A| / 2\rfloor$ so that $|A|$ is $2 n$ or $2 n+1$. In either case, we observe that for $k \leq n-1$, the dimension of $M_{k}^{\square}(A)$ is less than or equal to $2 n-1$, hence it cannot fill up all of $M^{\square}(A)$.

Example 4.3.6. For $A=\{0,2,5,9\}, M_{1}(A)$ is a union of $M_{1}^{\uparrow}(A)$ and $M_{1}^{\downarrow}(A)$, shown on the left in Figure 4.4. Since $1<2=\lfloor|A| / 2\rfloor$, these sets are not full dimensional and so cannot fill up $M^{\uparrow}(A)$ or $M^{\downarrow}(A)$. For $k=2=\lfloor|A| / 2\rfloor, M_{2}^{\uparrow}(A)=M^{\uparrow}(A)$ and $M_{2}^{\downarrow}(A)=M^{\downarrow}(A)$. These form parts of the full dimensional set $M_{2}(A)$ shown in the middle of Figure 4.4.

### 4.4 Connection with coalescence manifold

The motivation for studying moments of step functions comes from the field of population genetics. A central problem in this area is:

Question 4.4.1. Given a sample of $n$ genomes from a present-day population, what inferences can be drawn regarding the history of that population?

Our approach to the problem is to fix a function p describing effective population size at time $t$ before the present. We then compute, as a function of $p$, a vector of invariants $\mathbf{c}$ associated to the genome sample. Understanding the relationship between $p$ and $\mathbf{c}$ will allow us to infer likely values of $p$ based on measured data.

Following [12], we model the natural process of the production of a sample of $n$ genomes as follows:

- The genealogical tree connecting $n$ individuals will be formed by taking coalescence of each pair of lineages as a Poisson point process with rate parameter $1 / \mathrm{p}(t)$, where $\mathrm{p}(t)$ is the effective population size at time $t$ before present. (Heuristically, looking at the previous generation and picking parents at random, there is a $1 / \mathrm{p}(t)$ chance that two lineages will pick the same parent.)
- After the tree is specified, mutations are distributed on the tree as a Poisson point process with constant rate relative to branch length. The infinite-sites model is used, so that repeated mutation at a given site is disallowed, which is a good model for large genomes.

Definition 4.4.2. Fixing a population history, and defining the random process as above, we define random variables:

- The sample frequency spectrum (also known as the site or allele frequency spectrum), abbreviated SFS, is the vector of random variables $\left(X_{n, b}\right)_{b=1, \ldots, n-1}$ where $X_{n, b}$ denotes the number of mutations that are shared by exactly $b$ out of the $n$ individuals.
- The coalescence vector is the vector $\left(T_{i, i}\right)_{i=2, \ldots, n-1}$ of the time at which a sample of size $i$ has exactly $i$ distinct lineages, i.e. the time until the first coalescence.

For a fixed population function p , taking expectations gives the population invariants $\xi_{n, b}=\mathbb{E}\left[X_{n, b}\right]$ and $c_{i}=\mathbb{E}\left[T_{i, i}\right]$.

In practice, the SFS is more frequently discussed as a summary statistic, but the coalescence vector is simpler to use in computations. Fortunately, Polanski and Kimmel [72] proved that they are related by a linear transformation $A_{n}$, a matrix entirely determined by sample size $n$. Therefore, we focus on the coalescence vectors $\left(c_{i}\right)$.

Fact 4.4.3. We make the reasonable assumption that $\mathrm{p}(t)$ is bounded below by 0 and bounded above by a fixed $P$. Applying integration by parts and change of variables to the expected value of an exponential distribution yields the following expression for $c_{i}$ in terms of $\mathrm{p}(t)$ :

$$
\begin{equation*}
c_{i}(\mathrm{p})=\int_{0}^{\infty} \tilde{\mathrm{p}}(\tau) e^{-\binom{i}{2} \tau} \mathrm{~d} \tau \tag{4.5}
\end{equation*}
$$

where $\tilde{\mathrm{p}}(\tau)=\mathrm{p}\left(R_{\mathrm{p}}^{-1}(\tau)\right)$ and $R_{\mathrm{p}}(t)=\int_{0}^{t} \frac{1}{\mathrm{p}(x)} d x$. Because $0<\mathrm{p}(t)<P$, the function $R_{\mathrm{p}}$ is strictly increasing and unbounded; thus, it is a bijection from $\mathbb{R}_{\geq 0} \rightarrow \mathbb{R}_{\geq 0}$, so the inverse is well-defined. We call $\tilde{\mathrm{p}}(\tau)$ the transformed population history.

The coalescence vector can thus be considered a function from the space of (bounded) population history functions to $\mathbb{R}^{n-1}$. Since the former space is infinite-dimensional and the latter is finite-dimensional, it is natural to restrict our attention to a finite-dimensional space of population history functions. A common choice for this, motivated by injectivity considerations in [12], is

$$
\tilde{S}_{k}=\left\{\text { nonnegative step functions on } \mathbb{R}_{\geq 0} \text { with at most } k \text { breakpoints }\right\} .
$$

Definition 4.4.4. Let $n, k$ be integers with $n \geq 2$ and $k \geq 0$. The coalescence manifold $C_{n, k}$ is the Euclidean closure of the set of vectors $\tilde{\mathbf{c}}(\mathrm{p})=\mathbf{c}(\mathrm{p}) /\|\mathbf{c}(\mathrm{p})\|_{1}$ for all $\mathrm{p} \in \tilde{S}_{k}$. Here, $\mathbf{c}(\mathrm{p})=\left(c_{2}(\mathrm{p}), \ldots, c_{n}(\mathrm{p})\right)$ where $c_{i}(\mathrm{p})$ is defined as in Equation 4.5.

Because the vectors are normalized to have sum one, the coalescence manifold lives in the simplex $\Delta^{n-1}$. Note that this definition deviates slightly from the definition in [80] by allowing $k$ breakpoints instead of $k$ epochs (i.e. constant intervals). This shifts the index down by one. We now connect back to the moment cones studied above.

Theorem 4.4.5. Let $A=\left\{\binom{i}{2}-1: i=2, \ldots, n\right\}$. The coalescence manifold $C_{n, k}$ equals the intersection of the cone $M_{k}(A)$ with the affine hyperplane of points with coordinate sum equal to one:

$$
C_{n, k}=\left\{\mathbf{m} \in M_{k}(A): \sum_{a \in A} m_{a}=1\right\}
$$

Before we prove the theorem, we demonstrate two lemmas that will simplify the proof.
Lemma 4.4.6. Define $\tilde{\mathrm{p}}(\tau)$ as in Equation 4.5. Then $\mathrm{p}(t) \in \tilde{S}_{k}$ if and only if if $\tilde{\mathrm{p}}(\tau) \in \tilde{S}_{k}$.

Proof. Let $0=s_{0}<\cdots<s_{k-1}<s_{k}$ be the sequence of breakpoints of $\mathrm{p}(t)$. The function $R_{\mathrm{p}}(t)$ is a monotone increasing function, so the conditions below are equivalent:

$$
s_{j}<t \leq s_{j+1} \Longleftrightarrow R_{\mathrm{p}}\left(s_{j}\right)<R_{\mathrm{p}}(t) \leq R_{\mathrm{p}}\left(s_{j+1}\right)
$$

Since p is constant on $\left(s_{j}, s_{j+1}\right]$, the transformed history $\tilde{\mathrm{p}}(\tau)=\mathrm{p}\left(R_{\mathrm{p}}^{-1}(\tau)\right)$ is constant on $\left(R_{\mathrm{p}}\left(s_{j}\right), R_{\mathrm{p}}\left(s_{j+1}\right)\right]$. This implies that there are still at most $k$ breakpoints.

For the reverse direction, repeat the argument with $R_{\mathrm{p}}^{-1}$ in place of $R_{\mathrm{p}}$.
Lemma 4.4.7. Let $q$ be a strictly positive step function in $\tilde{S}_{k}$. Then, there exists p in $\tilde{S}_{k}$ such that $q(\tau)=\mathrm{p}\left(R_{\mathrm{p}}^{-1}(\tau)\right)$ where $R_{\mathrm{p}}(t)=\int_{0}^{t} \frac{1}{\mathrm{p}(x)} \mathrm{d} x$ as above.

Proof. Let $Q(t)=\int_{0}^{t} q(x) \mathrm{d} x$. We claim the desired function is $\mathrm{p}(t)=q\left(Q^{-1}(t)\right)$. First, note that because $q$ is strictly positive and takes only finitely many values, it is bounded away from zero. Therefore $Q$ is strictly increasing and takes all values in $[0, \infty)$. Its inverse $Q^{-1}$ therefore exists and is also increasing with range $[0, \infty)$. It follows that p takes the same values in the same order as $q$. In particular, $\mathrm{p} \in \tilde{S}_{k}$.

To check that $q(t)=\mathrm{p}\left(R_{\mathrm{p}}^{-1}(t)\right)$, we first show that $R_{\mathrm{p}}(Q(t))=t$ for all $t \geq 0$. By definition,

$$
R_{\mathrm{p}}(Q(t))=\int_{0}^{Q(t)} \frac{1}{\mathrm{p}(x)} d x=\int_{0}^{Q(t)} \frac{1}{q\left(Q^{-1}(x)\right)} d x=\int_{0}^{t} \frac{1}{q(w)} q(w) d w=t
$$

where the penultimate equation comes from substituting $x=Q(w)$ and $d x=q(w) d w$. Since both $Q$ and $R_{\mathrm{p}}$ are invertible, we see that $t=Q^{-1}\left(R_{\mathrm{p}}^{-1}(t)\right)$ for all $t$. Applying $q$ to both sides then gives the claim.

Proof of Theorem 4.4.5. We show that the set of coalescence vectors coming from population histories in $\tilde{S}_{k}$ is equal to the set of moments in $M_{k}(A)$ summing to 1 . The equality of the two closures is then automatic.

Assume $\mathrm{p} \in \tilde{S}_{k}$. From Lemma 4.4.6, $\tilde{\mathrm{p}}$ is also in $\tilde{S}_{k}$. Starting with Equation 4.5, we substitute $u=e^{-\tau}$ to obtain:

$$
c_{i}(\mathrm{p})=\int_{0}^{1} \tilde{\mathrm{p}}^{*}(u) u^{\left(\frac{i}{2}\right)-1} \mathrm{~d} u \text {, where } \tilde{\mathrm{p}}^{*}(u)=\mathrm{p}\left(R_{\mathrm{p}}^{-1}(-\ln (u))\right) .
$$

The function $\tilde{\mathrm{p}}^{*}$ is piecewise-constant on $[0,1]$ with at most $k$ breakpoints, so is in $S_{k}$; therefore, the quantity $c_{i}$ is the $\left(\binom{i}{2}-1\right)$-th moment of $\tilde{\mathrm{p}}^{*}$. This implies that $\mathbf{c}$ is in $M_{k}(A)$ where $A=\left\{\binom{i}{2}-1: i=2, \ldots, n\right\}$. Normalizing $\mathbf{c}$ is equivalent to scaling $\tilde{\mathrm{p}}^{*}$ so we may assume its sum is already equal to 1 .

Conversely, up to closure, any moment vector in $M_{k}(A)$ summing to 1 comes from some $f \in S_{k}$. Changing our domain to $\mathbb{R}_{\geq 0}$ gives $q(\tau)=f\left(e^{-\tau}\right)$ in $\tilde{S}_{k}$. By Lemma 4.4.7, we can produce $\mathrm{p} \in \tilde{S}_{k}$ that gives transformed population history $q$.

Example 4.4.8. Consider the population function $\mathrm{p}(t)=p_{1} \cdot \mathbf{1}_{\left[0, b_{1}\right)}+p_{2} \cdot \mathbf{1}_{\left[b_{1}, b_{2}\right)}+p_{3} \cdot \mathbf{1}_{\left[b_{2}, \infty\right)}$ where $p_{1}, p_{2}, p_{3}, b_{1}, b_{2} \in \mathbb{R}_{>0}$ with $b_{1}<b_{2}$. The function $R_{\mathrm{p}}(t)$ is piecewise linear, given by

$$
R_{\mathrm{p}}(t)=\int_{0}^{t} \frac{1}{\mathrm{p}(x)} d x=\frac{t}{p_{1}} \mathbf{1}_{\left[0, b_{1}\right)}+\left(\frac{t-b_{1}}{p_{2}}+\frac{b_{1}}{p_{1}}\right) \mathbf{1}_{\left[b_{1}, b_{2}\right)}+\left(\frac{t-b_{2}}{p_{3}}+\frac{b_{2}-b_{1}}{p_{2}}+\frac{b_{1}}{p_{1}}\right) \mathbf{1}_{\left[b_{2}, \infty\right)} .
$$



Figure 4.3 The functions p, $\tilde{\mathrm{p}}^{*}$, and $R_{\mathrm{p}}$ from Example 4.4.8.

This function is unbounded and strictly increasing with $R_{\mathrm{p}}(0)=0$, so it has an inverse $R_{\mathrm{p}}^{-1}$ that is also increasing and unbounded on $\mathbb{R}_{\geq 0}$. The function $\tilde{\mathrm{p}}(\tau)=\mathrm{p}\left(R_{\mathrm{p}}^{-1}(\tau)\right)$ is still piecewise constant with two break points $R_{\mathrm{p}}\left(b_{1}\right)=b_{1} / p_{1}$ and $R_{\mathrm{p}}\left(b_{2}\right)=\left(b_{2}-b_{1}\right) / p_{2}+b_{1} / p_{1}$, obtained by solving $R_{\mathrm{p}}^{-1}(\tau)=b_{i}$. The $i$ th entry of the coalescence vector is then

$$
c_{i}=\int_{0}^{\infty} \tilde{\mathrm{p}}(\tau) e^{\binom{i}{2} \tau} d \tau=\int_{0}^{1} \tilde{\mathrm{p}}^{*}(u) u^{\binom{i}{2}-1} d u \text { where } \tilde{\mathrm{p}}^{*}(u)=\tilde{\mathrm{p}}(-\ln (u)) .
$$

The second equality comes from the change of coordinates $u=e^{-\tau}$. Note that $\tilde{\mathrm{p}}^{*}$ is the step function given by

$$
\tilde{\mathrm{p}}^{*}=p_{3} \cdot \mathbf{1}_{\left(0, s_{1}\right]}+p_{2} \cdot \mathbf{1}_{\left(s_{1}, s_{2}\right]}+p_{1} \cdot \mathbf{1}_{\left(s_{2}, 1\right]} \text { where } s_{1}=e^{-R_{\mathrm{p}}\left(b_{2}\right)} \text { and } s_{2}=e^{-R_{\mathrm{p}}\left(b_{1}\right)} .
$$

The graphs of p and $\tilde{\mathrm{p}}^{*}$ for the values $\left(p_{1}, p_{2}, p_{3}\right)=(2,3,1)$ and $\left(b_{1}, b_{2}\right)=(2,5)$ are shown in Figure 4.3. In this case, the break points of $\tilde{\mathrm{p}}^{*}$ are $e^{-R_{\mathrm{p}}\left(b_{2}\right)}=e^{-2}$ and $e^{-R_{\mathrm{p}}\left(b_{1}\right)}=e^{-1}$.

Remark 4.4.9. Note that because $\mathrm{p}(t)$ denote the population size at time $t$ before the present, a population increasing over time corresponds to the function $\mathrm{p}(t)$ decreasing as a function of $t$, i.e. $p_{1}>p_{2}>p_{3}$ in the example above. Note that $\mathrm{p}(t)$ is decreasing in $t$ if and only if $\tilde{\mathrm{p}}(\tau)$ is decreasing in $\tau$. The parametrization $u=e^{-\tau}$ reverses direction and so the function $\tilde{\mathrm{p}}^{*}(u)$ is then increasing as a function of $u$. In these coordinates, $u=0$ corresponds "infinitely long ago" $(t=\infty)$ and $u=1$ corresponds to the present $(t=0)$. Therefore coalescence vectors of populations growing over time are moments of increasing step functions on $[0,1]$.

Theorem 4.4.5 allows us to apply our results from $M_{k}(A)$ to $C_{n, k}$.
Corollary 4.4.10. $C_{n, n-2}=C_{n, k}$ for all $k \geq n-2$ and $C_{n, n-3} \subsetneq C_{n, n-2}$.
Proof. For $A=\left\{\binom{i}{2}: i=2, \ldots, n\right\},|A|$ equals $n-1$. By Theorem 4.2.5, $M_{k}(A)=M(A)$ for all $k \geq n-2$. In particular, $M_{n-2}(A)=M_{k}(A)$ for all $k \geq n-2$. Intersecting with the hyperplane $\left\{\mathbf{m}: \sum_{a \in A} m_{a}=1\right\}$ gives that $C_{n, n-2}=C_{n, k}$ for all $k \geq n-2$. By Corollary 4.2.12, $M_{k}(A) \neq M(A)$ for $k<|A|-1=n-2$. Hence $M_{n-3}(A) \neq M(A)$. Since $M(A)=M_{n-2}(A)$, intersecting with the hyperplane $\left\{\mathbf{m}: \sum_{a \in A} m_{a}=1\right\}$ gives that $C_{n, n-2} \neq C_{n, n-3}$.

Affine transformations the sets $C_{5,1}, C_{5,2}$ and $C_{5,3}$ are show in Figure 4.4. As promised, $C_{5,3}$ is convex and $C_{5, k}$ is a strict subset for $k<3$.

### 4.5 The Coalescence Manifold in 3-Space

We illustrate the content of previous sections by focusing our attention on the case $A=\{0,2,5,9\}$, when $C_{5, k}$ is an affine slice of the cone $M_{k}(A)$. We begin by visualizing the moment sets $M_{k}(A)$ intersected with the affine hyperplane $\left\{m_{0}=1\right\}$, for $k=1,2$ and 3 . Affine transformations of these intersections are depicted in Figure 4.4. Note that the step functions with at most one breakpoint and total mass one can be written as $\lambda \mathbf{1}_{[0,1]}+(1-\lambda) \frac{1}{s} \mathbf{1}_{[0, s]}$ or $\lambda \mathbf{1}_{[0,1]}+(1-\lambda) \frac{1}{1-s} \mathbf{1}_{(s, 1]}$ where $\lambda \in[0,1]$. The result is a two-dimensional surface in the plane $\left\{m_{0}=1\right\}$. The set $M_{2}(A)$ is full-dimensional, but does not fill up all of $M(A)$. As promised by Lemma 4.2.11, the intersection $M_{2}(A)$ with the boundary of $M(A)$ has dimension $\leq 2$, so its image in $\left\{m_{0}=1\right\}$ has dimension $\leq 1$. Indeed, we see this intersection is given by the curve parameterized by $\left(t^{2}, t^{5}, t^{9}\right)$ for $t \in[0,1]$ and the line segment between its end points $(0,0,0)$ and $(1,1,1)$. Finally, by Theorem 4.2.5, $M_{3}(A)$ is the full cone $M(A)$. Points on the boundary of $M(A)$ have index $\leq 3$, and so have one of the two forms $w_{0} v_{A}(0)+w_{r} v_{A}(r)$ or $w_{1} v_{A}(1)+w_{r} v_{A}(r)$ where $r \in[0,1], w_{0}, w_{1}, w_{r} \in \mathbb{R}_{\geq 0}$.


Figure 4.4 The sets $M_{1}(A), M_{2}(A), M_{3}(A)$ in $\left\{m_{0}=1\right\}$ for $A=\{0,2,5,9\}$.

Now we take a closer look at $M_{2}(A)$ and analyze the structure of its boundaries. This example is of particular interest to us, since it is an illustration of what happens when $1<k<|A|-1$; cases where, for larger $n, C_{n, k}$ is not yet well understood. The domain of $M_{2}(A)$ is the set of step functions with two breakpoints, so it lives in the polytope $P$, the product of two simplices

$$
S=\left\{\left(s_{1}, s_{2}\right): 0 \leq s_{1} \leq s_{2} \leq 1\right\} \quad \text { and } \quad W=\left\{\left(w_{1}, w_{2}, w_{3}\right) \in \mathbb{R}_{\geq 0}^{3}: w_{1}+w_{2}+w_{3}=1\right\} .
$$

Here the $s$ variables parameterize the two breakpoints and the $w$ variables parameterize the proportion of mass in each piece. The coalescence manifold $C_{5,2}(A)$ is the image of the polytope $P=S \times W$ under the rational map $\mu_{A}: \mathbb{R}^{4} \rightarrow \mathbb{R}^{3}$. So the domain is four-dimensional while the image has dimension three. Therefore the generic fiber of the moment map has dimension one.

Proposition 4.5.1. Let p a point in the interior of $P$ that does not represent the constant function, the image $\mu_{A}(p)$ maps to the interior of $C_{5,2}(A)$.

To prove this proposition, we check where on $P$ does $J$, the Jacobian of the map $\mu_{A}$, drop rank.


Figure 4.5 The $S$ and $W$ simplex.

Definition 4.5.2. The Jacobian of the map $\mu_{A}$ is the $4 \times 3$ matrix

$$
J=\left[\begin{array}{ccc}
- & \frac{\partial}{\partial s_{1}}\left(\mu_{A}\right) & - \\
- & \frac{\partial}{\partial s_{2}}\left(\mu_{A}\right) & - \\
- & \frac{\partial}{\partial w_{1}}\left(\mu_{A}\right) & - \\
- & \frac{\partial}{\partial w_{2}}\left(\mu_{A}\right) & -
\end{array}\right]
$$

The $4 \times 4$ matrix $\tilde{J}(v)$ is the matrix $J$ augmented with a 4 th column $v$, recording which variables $s_{1}, s_{2}, w_{1}, w_{2}$ are set as constants.

$$
\tilde{J}(v)=\left[\begin{array}{cccc}
- & \frac{\partial}{\partial s_{1}}\left(\mu_{A}\right) & - & v_{1} \\
- & \frac{\partial}{\partial s_{2}}\left(\mu_{A}\right) & - & v_{2} \\
- & \frac{\partial}{\partial w_{1}}\left(\mu_{A}\right) & - & v_{3} \\
- & \frac{\partial}{\partial w_{2}}\left(\mu_{A}\right) & - & v_{4}
\end{array}\right]
$$

Proof. Let $p=\left(s_{1}, s_{2}, w_{1}, w_{2}\right)$ be in the interior of $S \times W$. For a fixed pair $\left(s_{1}, s_{2}\right)$ in the interior of $S$, if the image $\mu_{A}(p)$ is on the boundary of $C_{5,2}$, the tangent space at $p$ must be of dimension 2 at most, meaning the determinant of $\tilde{J}$ has to vanish for both $v=(1,0,0,0)$ and $v=(0,1,0,0)$. If only one of them vanishes, we still have 3 degrees of freedom to move around $\mu_{A}(p)$.

For $v=(1,0,0,0):$

$$
\begin{aligned}
& \operatorname{det}(\tilde{J})=-s_{2}^{3}\left(s_{1}-1\right)\left(w_{2}\left(1-s_{2}\right)+\left(1-w_{1}-w_{2}\right)\left(s_{1}-s_{2}\right)\right)\left(3 s_{1}{ }^{6} s_{2}^{3}+6 s_{1}{ }^{5} s_{2}^{4}+\right. \\
& 9 s_{1}{ }^{4} s_{2}{ }^{5}+12 s_{1}{ }^{3} s_{2}{ }^{6}+8 s_{1}{ }^{2} s_{2}{ }^{7}+4 s_{1} s_{2}{ }^{8}+6 s_{1}{ }^{6} s_{2}{ }^{2}+15 s_{1}{ }^{5} s_{2}{ }^{3}+24 s_{1}{ }^{4} s_{2}{ }^{4}+ \\
& 33 s_{1}{ }^{3} s_{2}{ }^{5}+28 s_{1}{ }^{2} s_{2}{ }^{6}+16 s_{1} s_{2}{ }^{7}+4 s_{2}{ }^{8}+4 s_{1}{ }^{6} s_{2}+14 s_{1}{ }^{5} s_{2}{ }^{2}+27 s_{1}{ }^{4} s_{2}{ }^{3}+ \\
& 40 s_{1}{ }^{3} s_{2}{ }^{4}+41 s_{1}{ }^{2} s_{2}{ }^{5}+28 s_{1} s_{2}{ }^{6}+8 s_{2}^{7}+2 s_{1}{ }^{6}+8 s_{1}{ }^{5} s_{2}+20 s_{1}{ }^{4} s_{2}^{2}+35 s_{1}{ }^{3} s_{2}{ }^{3}+ \\
& 40 s_{1}{ }^{2} s_{2}^{4}+33 s_{1} s_{2}^{5}+12 s_{2}^{6}+2 s_{1}^{5}+8 s_{1}{ }^{4} s_{2}+20 s_{1}{ }^{3} s_{2}^{2}+27 s_{1}^{2} s_{2}^{3}+24 s_{1} s_{2}^{4}+ \\
& \left.9 s_{2}^{5}+2 s_{1}^{4}+8 s_{1}^{3} s_{2}+14 s_{1}^{2} s_{2}^{2}+15 s_{1} s_{2}^{3}+6 s_{2}^{4}+2 s_{1}^{3}+4 s_{1}^{2} s_{2}+6 s_{1} s_{2}^{2}+3 s_{2}^{3}\right)
\end{aligned}
$$

and for $v=(0,1,0,0)$

$$
\begin{align*}
& \operatorname{det}(\tilde{J})=-s_{1} s_{2}\left(s_{1}-1\right)^{2}\left(w_{1}\left(s_{2}-s_{1}\right)-w_{2} s_{1}\right)\left(4 s_{1}{ }^{8} s_{2}+8 s_{1}{ }^{7} s_{2}{ }^{2}+12 s_{1}{ }^{6} s_{2}{ }^{3}+9 s_{1}{ }^{5} s_{2}{ }^{4}+\right. \\
& 6 s_{1}{ }^{4} s_{2}^{5}+3 s_{1}{ }^{3} s_{2}{ }^{6}+4 s_{1}{ }^{8}+16 s_{1}{ }^{7} s_{2}+28 s_{1}{ }^{6} s_{2}^{2}+33 s_{1}{ }^{5} s_{2}{ }^{3}+24 s_{1}{ }^{4} s_{2}^{4}+ \\
& 15 s_{1}{ }^{3} s_{2}{ }^{5}+6 s_{1}{ }^{2} s_{2}{ }^{6}+8 s_{1}{ }^{7}+28 s_{1}{ }^{6} s_{2}+41 s_{1}{ }^{5} s_{2}{ }^{2}+40 s_{1}{ }^{4} s_{2}{ }^{3}+27 s_{1}{ }^{3} s_{2}{ }^{4}+  \tag{4.6}\\
& 14 s_{1}{ }^{2} s_{2}{ }^{5}+4 s_{1} s_{2}{ }^{6}+12 s_{1}{ }^{6}+33 s_{1}{ }^{5} s_{2}+40 s_{1}^{4} s_{2}^{2}+35 s_{1}^{3} s_{2}{ }^{3}+20 s_{1}^{2} s_{2}^{4}+ \\
& 8 s_{1} s_{2}{ }^{5}+2 s_{2}{ }^{6}+9 s_{1}{ }^{5}+24 s_{1}^{4} s_{2}+27 s_{1}{ }^{3} s_{2}^{2}+20 s_{1}{ }^{2} s_{2}^{3}+8 s_{1} s_{2}^{4}+2 s_{2}^{5}+6 s_{1}^{4}+ \\
& \left.15 s_{1}{ }^{3} s_{2}+14 s_{1}{ }^{2} s_{2}^{2}+8 s_{1} s_{2}^{3}+2 s_{2}^{4}+3 s_{1}^{3}+6 s_{1}^{2} s_{2}+4 s_{1} s_{2}^{2}+2 s_{2}^{3}\right) \text {. }
\end{align*}
$$

Note that in either case, the large factor of degree 9 does not vanish on the interior of $S \times W$. Recall that under the change of coordinates to $w_{i}=y_{i}\left(s_{i}-s_{i-1}\right)$, the constraint $y_{1}=y_{2}$ is given by $w_{1}\left(s_{2}-s_{1}\right)-w_{2} s_{1}$ and $y_{2}=y_{3}$ is given by $w_{2}\left(1-s_{2}\right)+\left(1-w_{1}-w_{2}\right)\left(s_{1}-s_{2}\right)$. For both these polynomials to vanish, $p$ either needs to be on the boundary of $S \times W$ or describe the constant function.

The boundary of the polytope $P$ is composed of six triangular prisms given by $s_{1}=0, s_{2}=1, s_{1}=s_{2}$, $w_{1}=0, w_{2}=0$, and $w_{3}=0$. We can visualize this by way of a Schlegel diagram via one of its facets, seen in Figure 4.6.


Figure 4.6 Schlegel diagram for the boundary of $P$, together with the image of each facet under the moment map.

We can however obtain a generically finite-to-one map by restricting to facets of the polytope $P$.
Proposition 4.5.3. The boundary of $C_{5,2}$ is contained in the image of the 2 -dimensional faces of $P$ under the map $\mu_{A}$.

Proof. We look at the six 3-dimensional facets of $P$. For each face, we assign a vector $v \in \mathbb{R}^{4}$ describing the linear constraint on $\left(s_{1}, s_{2}, w_{1}, w_{2}\right)$ and compute the determinant of $\tilde{J}$.

1. If $s_{1}=0$, we set $v=(1,0,0,0)$,

$$
\operatorname{det}(\tilde{J})=\left(4 s_{2}^{5}+8 s_{2}^{4}+12 s_{2}^{3}+9 s_{2}^{2}+6 s_{2}+3\right) s_{2}^{6}\left(s_{2}\left(w_{1}-1\right)+w_{2}\right)
$$

This polynomial vanishes either when $s_{2}=0$ or $s_{2} w_{1}-s_{2}+w_{2}=0$.
2. If $s=s_{1}=s_{2}$, we set $v=(1,-1,0,0)$,

$$
\operatorname{det}(\tilde{J})=w_{2} s^{6}\left(14 s^{8}+14 s^{7}-16 s^{6}-16 s^{5}-16 s^{4}+5 s^{3}+5 s^{2}+5 s+5\right)
$$

This polynomial vanishes when $w_{2}=0$ and has 4 real roots in $s$, two of them negative, $s=0$ and $s=1$.
3. If $s_{2}=1$, we set $v=(0,1,0,0)$,

$$
\begin{aligned}
\operatorname{det}(\tilde{J})= & \left(s_{1}-1\right)^{2}\left(\left(w_{1}+w_{2}\right) s_{1}-w_{1}\right) s_{1} \\
& \left(8 s_{1}{ }^{8}+32 s_{1}{ }^{7}+80 s_{1}{ }^{6}+125 s_{1}^{5}+70 s_{1}^{4}+125 s_{1}^{3}+80 s_{1}^{2}+32 s_{1}+8\right)
\end{aligned}
$$

This polynomial vanishes if $s_{1}=0, s_{1}=1$ or $\left(w_{1}+w_{2}\right) s_{1}-w_{1}=0$.
4. If $w_{1}=0$, we set $v=(0,0,1,0)$

$$
\begin{aligned}
& \operatorname{det}(\tilde{J})=\left(s_{1}-1\right)^{2}\left(s_{1}-s_{2}\right) w_{2}\left(w_{2}\left(s_{2}-1\right)+\left(w_{2}-1\right)\left(s_{1}-s_{2}\right)\right)\left(8 s_{2} s_{1}{ }^{9}+32 s_{1}{ }^{8} s_{2}{ }^{2}+\right. \\
& 80 s_{1}{ }^{7} s_{2}^{3}+125 s_{2}^{4} s_{1}{ }^{6}+140 s_{1}^{5} s_{2}^{5}+125 s_{2}^{6} s_{1}^{4}+80 s_{1}{ }^{3} s_{2}{ }^{7}+32 s_{2}{ }^{8} s_{1}{ }^{2}+ \\
& 8 s_{2}{ }^{9} s_{1}+4 s_{1}{ }^{9}+32 s_{1}{ }^{8} s_{2}+104 s_{1}{ }^{7} s_{2}{ }^{2}+212 s_{1}{ }^{6} s_{2}{ }^{3}+278 s_{1}{ }^{5} s_{2}{ }^{4}+ \\
& 278 s_{1}{ }^{4} s_{2}{ }^{5}+212 s_{1}{ }^{3} s_{2}{ }^{6}+104 s_{1}{ }^{2} s_{2}{ }^{7}+32 s_{1} s_{2}{ }^{8}+4 s_{2}{ }^{9}+8 s_{1}{ }^{8}+56 s_{2} s_{1}{ }^{7}+ \\
& 155 s_{1}{ }^{6} s_{2}^{2}+260 s_{1}{ }^{5} s_{2}^{3}+302 s_{2}{ }^{4} s_{1}^{4}+260 s_{2}{ }^{5} s_{1}^{3}+155 s_{2}{ }^{6} s_{1}{ }^{2}+56 s_{2}{ }^{7} s_{1}+ \\
& 8 s_{2}{ }^{8}+12 s_{1}{ }^{7}+66 s_{1}{ }^{6} s_{2}+150 s_{1}{ }^{5} s_{2}^{2}+222 s_{2}^{3} s_{1}^{4}+222 s_{2}{ }^{4} s_{1}^{3}+150 s_{2}{ }^{5} s_{1}^{2}+ \\
& 66 s_{2}{ }^{6} s_{1}+12 s_{2}{ }^{7}+9 s_{1}{ }^{6}+48 s_{1}^{5} s_{2}+102 s_{1}^{4} s_{2}^{2}+132 s_{1}{ }^{3} s_{2}^{3}+102 s_{2}{ }^{4} s_{1}^{2}+ \\
& 48 s_{2}^{5} s_{1}+9 s_{2}{ }^{6}+6 s_{1}^{5}+30 s_{1}^{4} s_{2}+54 s_{1}^{3} s_{2}^{2}+54 s_{1}^{2} s_{2}^{3}+30 s_{1} s_{2}^{4}+6 s_{2}^{5}+ \\
& \left.3 s_{1}^{4}+12 s_{2} s_{1}^{3}+15 s_{1}^{2} s_{2}^{2}+12 s_{2}^{3} s_{1}+3 s_{2}^{4}\right) \text {. }
\end{aligned}
$$

This polynomial vanishes when $s_{1}=1, s_{1}=s_{2}, w_{2}=0$ or $w_{2}\left(s_{2}-1\right)+\left(w_{2}-1\right)\left(s_{1}-s_{2}\right)=0$.
5. If $w_{2}=0$ and we set $v=(0,0,0,1)$

$$
\begin{aligned}
& \operatorname{det}(\tilde{J})=w_{1} s_{1}\left(w_{1}-1\right)\left(s_{1}-s_{2}\right)\left(8 s_{2} s_{1}^{11}+8 s_{1}{ }^{10} s_{2}{ }^{2}+8 s_{2}{ }^{3} s_{1}{ }^{9}-27 s_{2}{ }^{4} s_{1}{ }^{8}-\right. \\
& 27 s_{2}{ }^{5} s_{1}{ }^{7}+27 s_{2}{ }^{8} s_{1}{ }^{4}+27 s_{2}{ }^{9} s_{1}{ }^{3}-8 s_{1}{ }^{2} s_{2}{ }^{10}-8 s_{2}{ }^{11} s_{1}-8 s_{2}{ }^{12}+4 s_{1}{ }^{11}+ \\
& 4 s_{2} s_{1}{ }^{10}+4 s_{2}{ }^{2} s_{1}{ }^{9}-24 s_{2}^{3} s_{1}{ }^{8}-24 s_{2}{ }^{4} s_{1}{ }^{7}+30 s_{2} s_{1}{ }^{6}+30 s_{1}{ }^{5} s_{2}{ }^{6}+54 s_{2}{ }^{7} s_{1}{ }^{4}+ \\
& 54 s_{2}{ }^{8} s_{1}{ }^{3}-16 s_{2}{ }^{9} s_{1}{ }^{2}-16 s_{1} s_{2}{ }^{10}-16 s_{2}{ }^{11}-21 s_{1}{ }^{8} s_{2}{ }^{2}-21 s_{1}{ }^{7} s_{2}{ }^{3}+60 s_{2}{ }^{4} s_{1}{ }^{6}+ \\
& 60 s_{1}{ }^{5} s_{2}^{5}+81 s_{2}{ }^{6} s_{1}^{4}+81 s_{1}{ }^{3} s_{2}{ }^{7}-24 s_{2}{ }^{8} s_{1}{ }^{2}-24 s_{2}{ }^{9} s_{1}-24 s_{2}{ }^{10}-14 s_{1}{ }^{8} s_{2}- \\
& 14 s_{1}{ }^{7} s_{2}{ }^{2}+40 s_{1}{ }^{6} s_{2}{ }^{3}+40 s_{1}{ }^{5} s_{2}{ }^{4}+58 s_{1}{ }^{4} s_{2}{ }^{5}+58 s_{1}{ }^{3} s_{2}{ }^{6}-32 s_{1}{ }^{2} s_{2}{ }^{7}-32 s_{1} s_{2}{ }^{8}- \\
& 32 s_{2}{ }^{9}-7 s_{1}{ }^{8}-7 s_{2} s_{1}{ }^{7}+20 s_{1}{ }^{6} s_{2}{ }^{2}+20 s_{1}^{5} s_{2}{ }^{3}+35 s_{2}^{4} s_{1}^{4}+35 s_{2}^{5} s_{1}^{3}- \\
& 40 s_{2}{ }^{6} s_{1}{ }^{2}-40 s_{2}{ }^{7} s_{1}-40 s_{2}{ }^{8}+12 s_{2}{ }^{3} s_{1}{ }^{4}+12 s_{2}{ }^{4} s_{1}{ }^{3}-48 s_{2}^{5} s_{1}{ }^{2}-48 s_{2}{ }^{6} s_{1}- \\
& 48 s_{2}{ }^{7}+9 s_{1}^{4} s_{2}^{2}+9 s_{1}{ }^{3} s_{2}{ }^{3}-36 s_{2}{ }^{4} s_{1}{ }^{2}-36 s_{2}{ }^{5} s_{1}-36 s_{2}{ }^{6}+6 s_{1}{ }^{4} s_{2}+6 s_{1}{ }^{3} s_{2}{ }^{2}- \\
& \left.24 s_{1}{ }^{2} s_{2}^{3}-24 s_{1} s_{2}{ }^{4}-24 s_{2}{ }^{5}+3 s_{1}{ }^{4}+3 s_{2} s_{1}^{3}-12 s_{1}{ }^{2} s_{2}^{2}-12 s_{2}{ }^{3} s_{1}-12 s_{2}^{4}\right) \text {. }
\end{aligned}
$$

This polynomial vanishes when $w_{1}=0, w_{1}=1, s_{1}=0$ or $s_{1}=s_{2}$, but it is not directly clear whether the last factor of $\operatorname{det}(\tilde{J})$, which we denote $p$, does not change signs on the region $0 \leq s_{1} \leq s_{2} \leq 1$. Hence, we want to verify that the polynomial $p$ of degree 12 in $s_{1}, s_{2}$ is non-positive on $S=\left\{\left(s_{1}, s_{2}\right): 0 \leq s_{1} \leq s_{2} \leq 1\right\}$ :

$$
\begin{align*}
& p=8 s_{2} s_{1}{ }^{11}+8 s_{1}{ }^{10} s_{2}{ }^{2}+8 s_{2}{ }^{3} s_{1}{ }^{9}-27 s_{2}{ }^{4} s_{1}{ }^{8}-27 s_{2}{ }^{5} s_{1}{ }^{7}+27 s_{2}{ }^{8} s_{1}{ }^{4}+27 s_{2}{ }^{9} s_{1}{ }^{3}- \\
& 8 s_{1}{ }^{2} s_{2}{ }^{10}-8 s_{2}{ }^{11} s_{1}-8 s_{2}{ }^{12}+4 s_{1}{ }^{11}+4 s_{2} s_{1}{ }^{10}+4 s_{1}{ }^{9} s_{2}{ }^{2}-24 s_{2}{ }^{3} s_{1}{ }^{8}- \\
& 24 s_{2}{ }^{4} s_{1}{ }^{7}+30 s_{2}{ }^{5} s_{1}{ }^{6}+30 s_{2}{ }^{6} s_{1}{ }^{5}+54 s_{2}{ }^{7} s_{1}{ }^{4}+54 s_{2}{ }^{8} s_{1}{ }^{3}-16 s_{2}{ }^{9} s_{1}{ }^{2}- \\
& 16 s_{1} s_{2}^{10}-16 s_{2}{ }^{11}-21 s_{1}{ }^{8} s_{2}{ }^{2}-21 s_{2}{ }^{3} s_{1}{ }^{7}+60 s_{2}{ }^{4} s_{1}{ }^{6}+60 s_{2}{ }^{5} s_{1}{ }^{5}+ \\
& 81 s_{2}{ }^{6} s_{1}{ }^{4}+81 s_{2}{ }^{7} s_{1}{ }^{3}-24 s_{2}{ }^{8} s_{1}{ }^{2}-24 s_{2}{ }^{9} s_{1}-24 s_{2}{ }^{10}-14 s_{1}{ }^{8} s_{2}-14 s_{1}{ }^{7} s_{2}{ }^{2}+  \tag{4.7}\\
& 40 s_{1}{ }^{6} s_{2}{ }^{3}+40 s_{1}{ }^{5} s_{2}{ }^{4}+58 s_{1}{ }^{4} s_{2}{ }^{5}+58 s_{1}{ }^{3} s_{2}{ }^{6}-32 s_{1}{ }^{2} s_{2}{ }^{7}-32 s_{1} s_{2}{ }^{8}-32 s_{2}{ }^{9}- \\
& 7 s_{1}{ }^{8}-7 s_{2} s_{1}{ }^{7}+20 s_{1}{ }^{6} s_{2}{ }^{2}+20 s_{2}{ }^{3} s_{1}{ }^{5}+35 s_{2}{ }^{4} s_{1}^{4}+35 s_{2}{ }^{5} s_{1}{ }^{3}-40 s_{2}{ }^{6} s_{1}{ }^{2}- \\
& 40 s_{2}{ }^{7} s_{1}-40 s_{2}{ }^{8}+12 s_{2}{ }^{3} s_{1}^{4}+12 s_{2}^{4} s_{1}^{3}-48 s_{2}{ }^{5} s_{1}^{2}-48 s_{2}{ }^{6} s_{1}-48 s_{2}{ }^{7}+ \\
& 9 s_{1}^{4} s_{2}^{2}+9 s_{2}^{3} s_{1}^{3}-36 s_{2}^{4} s_{1}^{2}-36 s_{2}{ }^{5} s_{1}-36 s_{2}{ }^{6}+6 s_{1}{ }^{4} s_{2}+6 s_{1}{ }^{3} s_{2}^{2}- \\
& 24 s_{1}^{2} s_{2}^{3}-24 s_{1} s_{2}^{4}-24 s_{2}^{5}+3 s_{1}^{4}+3 s_{2} s_{1}^{3}-12 s_{1}^{2} s_{2}^{2}-12 s_{2}^{3} s_{1}-12 s_{2}^{4} \text {. }
\end{align*}
$$

We compute both partials $\partial p / \partial s_{1}$ and $\partial p / \partial s_{2}$ and use the Maple package Msolve to check that both partial do not vanish on the interior of $S$. Through symbolic computations, we observe that $p$ is non-positive on all of the boundary of $S$ and only vanishes at the points $\{(0,0),(1,1)\}$.
6. if $w_{1}+w_{2}=1$, we set $v=(0,0,1,1)$

$$
\begin{align*}
& \operatorname{det}(\tilde{J})=w_{2}\left(s_{1}-s_{2}\right) s_{2}^{3} s_{1}\left(\left(w_{2}-1\right) s_{2}+s_{1}\right)\left(8 s_{1}^{8}+32 s_{1}^{7}+80 s_{1}^{6}+\right.  \tag{4.8}\\
& \left.125 s_{1}^{5}+70 s_{1}^{4}+125 s_{1}^{3}+80 s_{1}^{2}+32 s_{1}+8\right)
\end{align*}
$$

This polynomial vanishes when $w_{2}=0, s_{1}=0, s_{1}=s_{2}, s_{2}=0$ and $\left(w_{2}-1\right) s_{2}+s_{1}=0$.

To summarize, there are four 2-faces that the moment map collapses to a curve, namely the faces given by $0=s_{1}=s_{2}, s_{1}=s_{2}=1, w_{1}=w_{2}=0$, and $w_{2}=w_{3}=0$. In addition to these 2 -faces, the intersections
of the facets of $P$ with the hypersurfaces given by $y_{1}=y_{2}$ and $y_{2}=y_{3}$ sometimes drop dimension under the moment map $\mu_{A}$. In the $(\mathbf{s}, \mathbf{w})$ variables, these correspond to surfaces $\left(s_{2}-s_{1}\right) w_{1}=s_{1} w_{2}$ and $\left(1-s_{2}\right) w_{2}=s_{2} w_{3}$, respectively. For example, in each of the facets $s_{1}=0$ and $w_{1}=0$, the equation $\left(1-s_{2}\right) w_{2}=s_{2} w_{3}$ cuts out a surface whose image under $\mu_{A}$ is a curve. For the face $s_{1}=0$, the moments of this surface collapse to the line segment connecting the constant population and the point mass at 0 and for $w_{1}=0$, the image of this surface collapses to the curve segment of moments of step functions of a single step with $w_{1}=0$. Similarly, the faces $s_{2}=1$ and $w_{3}=0$ each contain a two-dimensional surface cut out by $s_{1} w_{2}=\left(1-s_{1}\right) w_{1}$ whose moments collapse to a line segment-from the constant population to the point mass at 1 for $s_{2}=1$ and a curve segment of moments of a single step with $w_{3}=0$. Aside from these subsurfaces, the map on the boundary $\partial P$ is locally nondegenerate. Interestingly, the images of these facets can overlap in full-dimensional sets. One consequence is that the fibers of the moment map can be disconnected.


Figure 4.7 A disconnected fiber of $\mu_{A}$ for $A=\{0,2,5,9\}$.

For example, the point $\left(m_{0}, m_{2}, m_{5}, m_{9}\right)=(1,0.164,0.054,0.031)$ belongs to $M_{2}(A)$ for $A=$ $\{0,2,5,9\}$. Its fiber under $\mu_{A}$ is a curve in the four-dimensional polytope $P$ from (4.4). Figure 4.7 shows the $\left(s_{1}, s_{2}\right)$-coordinates of this curve. In particular, this fiber has at least two connected components. In a lighter shade is the two-dimensional fiber of the point $\left(m_{0}, m_{2}, m_{5}\right)=(1,0.164,0.054)$ under the corresponding map for $\{0,2,5\}$.

### 4.6 Connections with semidefinite programming

In this section, we describe how to write the moment cone $M(A)$ and coalescence manifold $C_{n, n-2}$ as projections of spectrahedra. This gives rise to natural algorithms for testing membership and finding nearest points in these sets based on semidefinite programming. Formally, a spectrahedron is a set of the form $\left\{\mathbf{x} \in \mathbb{R}^{n}: A_{0}+\sum_{i=1}^{n} x_{i} A_{i} \geq 0\right\}$ where $A_{0}, \ldots, A_{n}$ are real symmetric matrices and $X \geq 0$ denotes that the matrix $X$ is positive semidefinite. These are the feasible sets of semidefinite programs. See e.g. [17, Ch. 5 and 6]. Python code for computing the nearest point in $C_{n, n-2}$ to an arbitrary point in $\mathbb{R}^{n-1}$ is available at: https://github.com/gescholt/DistanceToCoalescenceManifold

Theorem 4.6.1 (Theorems 10.1 and 10.2 [87]). For any $d \in \mathbb{Z}_{+}$, the cone $M(\{0,1, \ldots, d\})$ is a spectrahedron. If $d=2 e$ is even, then

$$
M(\{0,1, \ldots, d\})=\left\{\mathbf{m} \in \mathbb{R}^{d+1}:\left(m_{i+j}\right)_{0 \leq i, j \leq e} \geq 0 \text { and }\left(m_{i+j+1}-m_{i+j+2}\right)_{0 \leq i, j \leq e-1} \geq 0\right\},
$$

and if $d=2 e+1$ is odd, then

$$
M(\{0,1, \ldots, d\})=\left\{\mathbf{m} \in \mathbb{R}^{d+1}:\left(m_{i+j+1}\right)_{0 \leq i, j \leq e} \geq 0 \text { and }\left(m_{i+j}-m_{i+j+1}\right)_{0 \leq i, j \leq e} \geq 0\right\} .
$$

Corollary 4.6.2. For any finite set of integers $A \subset \mathbb{N}$, the convex cones $M(A), M^{\uparrow}(A)$ and $M^{\downarrow}(A)$ are projections of the spectrahedron $M(\{0,1, \ldots, \max (A)\})$.

Proof. Let $d=\max (A)$. Note that by definition, $M(A)$ equals the closure of the projection of $M(\{0,1, \ldots, d\})$ under the map $\left(m_{0}, m_{1}, \ldots, m_{d}\right) \mapsto\left(m_{a}\right)_{a \in A}$. For $0 \in A$, this projection is closed and otherwise, we replace $A$ with $B=\{a-\min (A): a \in A\}$ as in Lemma 4.2.2. By Theorem 4.6.1, $M(\{0,1, \ldots, d\})$ is a spectrahedron.

More generally, consider any finite collection of polynomials $p_{1}, \ldots, p_{n} \in \mathbb{R}[x]_{\leq d}$. We claim that the conical hull of the curve parameterized by $\mathbf{p}(t)=\left(p_{1}(t), \ldots, p_{n}(t)\right)$ for $t \in[0,1]$ is the image of $M(\{0,1, \ldots, d\})$ under a linear map. Specifically, consider the linear map $\pi: \mathbb{R}^{d+1} \rightarrow$ $\mathbb{R}^{n}$ taking $\left(m_{0}, m_{1}, \ldots, m_{d}\right)$ to $\left(\sum_{j=0}^{d} p_{i j} m_{j}\right)_{i \in[n]}$ where $p_{i}(x)=\sum_{j=0}^{d} p_{i j} x^{j}$. For any $t \in[0,1]$, $\mathbf{p}(t)$ equals $\pi\left(v_{d}(t)\right)$ where $v_{d}(t)=\left(1, t, t^{2}, \ldots, t^{d}\right)$. Since $M(\{0,1, \ldots, d\})$ is the conical hull of $\left\{v_{d}(t): t \in[0,1]\right\}$, the conical hull of $\{\mathbf{p}(t): t \in[0,1]\}$ is the image of $M(\{0,1, \ldots, d\})$ under $\pi$.

Note that the coordinates of both $\gamma_{A}^{\uparrow}(t)$ and $\gamma_{A}^{\downarrow}(t)$ are given by polynomials in $t$ of degree $\leq d$. Then by Lemma 4.3.3 and the arguments above, both $M^{\uparrow}(A)$ and $M^{\downarrow}(A)$ can be written as the image of $M(\{0,1, \ldots, d\})$ under a linear map.

Example 4.6.3. For $A=\{0,2,5,9\}$, we write $M(A), M^{\uparrow}(A)$ and $M^{\downarrow}(A)$ are projections of the spectrahedron $M(\{0,1, \ldots, 9\})$. By Theorem 4.6.1, this is given by the set of $\mathbf{m}=\left(m_{0}, \ldots, m_{9}\right)$ in $\mathbb{R}^{10}$ for which the matrices

$$
\left(\begin{array}{lllll}
m_{1} & m_{2} & m_{3} & m_{4} & m_{5} \\
m_{2} & m_{3} & m_{4} & m_{5} & m_{6} \\
m_{3} & m_{4} & m_{5} & m_{6} & m_{7} \\
m_{4} & m_{5} & m_{6} & m_{7} & m_{8} \\
m_{5} & m_{6} & m_{7} & m_{8} & m_{9}
\end{array}\right) \text { and }\left(\begin{array}{lllll}
m_{0}-m_{1} & m_{1}-m_{2} & m_{2}-m_{3} & m_{3}-m_{4} & m_{4}-m_{5} \\
m_{1}-m_{2} & m_{2}-m_{3} & m_{3}-m_{4} & m_{4}-m_{5} & m_{5}-m_{6} \\
m_{2}-m_{3} & m_{3}-m_{4} & m_{4}-m_{5} & m_{5}-m_{6} & m_{6}-m_{7} \\
m_{3}-m_{4} & m_{4}-m_{5} & m_{5}-m_{6} & m_{6}-m_{7} & m_{7}-m_{8} \\
m_{4}-m_{5} & m_{5}-m_{6} & m_{6}-m_{7} & m_{7}-m_{8} & m_{8}-m_{9}
\end{array}\right)
$$

are positive semidefinite. We obtain $M(A)$ as the image of this cone under the linear map

$$
\mathbf{m} \mapsto\left(m_{0}, m_{2}, m_{5}, m_{9}\right)
$$

Similarly, the cones $M^{\uparrow}(A)$ and $M^{\downarrow}(A)$ are the images of $M(\{0,1, \ldots, 9\})$ under the (respective) maps

$$
\mathbf{m} \mapsto\left(m_{0}, \frac{m_{0}+m_{1}+m_{2}}{3}, \frac{1}{6} \sum_{i=0}^{5} m_{i}, \frac{1}{10} \sum_{i=0}^{9} m_{i}\right) \text { and } \mathbf{m} \mapsto\left(m_{0}, \frac{m_{2}}{3}, \frac{m_{5}}{6}, \frac{m_{9}}{10}\right)
$$

Corollary 4.6.4. Testing membership any of the cones $M(A), M^{\uparrow}(A)$ or $M^{\downarrow}(A)$ is equivalent to testing the feasibility of a semidefinite program in $\leq d+1$ variables with two matrix constraints, each of size $\leq d / 2+1$, where $d=\max (A)$.

Corollary 4.6.5. For $k \geq n-2$, the coalescence manifold $C_{n, k}$ is the projection of a spectrahedron. Testing membership in $C_{n, k}$ is equivalent to testing the feasibility of a semidefinite program in $\leq n^{2} / 2$ variables with two matrix constraints, each of size $\leq n^{2} / 4$.

Proof. By Theorem 4.4.5 and Corollary 4.4.10, for all $k \geq n-2$, coalescence manifold $C_{n, k}$ equals in the intersection of $M(A)$ with the affine hyperplane given by $\sum_{a \in A} m_{a}=1$ where $A=\left\{\binom{i}{2}-1: i=2, \ldots, n\right\}$. By Corollary 4.6.2, $M(A)$ is the projection of $M(\{0,1, \ldots, d\})$ where $d=\binom{n}{2}-1$. It follows that $C_{n, k}$ is the projection of the points in $M(\{0,1, \ldots, d\})$ satisfying the affine linear equation $\sum_{a \in A} m_{a}=1$. The intersection of a spectrahedron with an affine linear space is again a spectrahedron and so $C_{n, k}$ is the projection of a spectrahedron.

The spectrahedron $M(\{0,1, \ldots, d\})$ is defined by two linear matrix inequalities of size $\leq d / 2+1 \leq$ $n^{2} / 4$. There are at most $d+1=\binom{n}{2} \leq n^{2} / 2$ variables.

Similarly, given a point $\mathbf{p} \in \mathbb{R}^{n-1}$, we can use a semidefinite program to find the nearest point in $C_{n, k}$ for sufficiently large $k$. This comes from the description of $C_{n, k}$ above and the fact that distance minimization can be phrased as a semidefinite program (see, e.g. [20]). Specifically, given $\mathbf{x} \in \mathbb{R}^{n-1}$, the matrix $n \times n$ matrix $\left(\begin{array}{cc}\lambda & (\mathbf{x}-\mathbf{p})^{T} \\ \mathbf{x}-\mathbf{p} & \operatorname{Id}_{n-1}\end{array}\right)$ is positive semidefinite if and only if $\|\mathbf{x}-\mathbf{p}\|_{2}^{2} \leq \lambda$, where $\operatorname{Id}_{n-1}$ denotes the $(n-1) \times(n-1)$ identity matrix. Given a set $S \subset \mathbb{R}^{n-1}$, suppose that $\lambda^{*}$ and $\mathbf{x}^{*}$ attain the minimum

$$
\min _{\lambda \in \mathbb{R}, \mathbf{x} \in S} \lambda \text { such that }\left(\begin{array}{cc}
\lambda & (\mathbf{x}-\mathbf{p})^{T} \\
\mathbf{x}-\mathbf{p} & \operatorname{Id}_{n-1}
\end{array}\right) \geq 0
$$

Then $\mathbf{x}^{*}$ is (one of) the nearest points in $S$ to $\mathbf{p}$ and the distance $\left\|\mathbf{x}^{*}-\mathbf{p}\right\|_{2}$ is $\sqrt{\lambda^{*}}$. In particular, if the set $S$ is the projection of a spectrahedron, then this minimization problem is a semidefinite program.

Corollary 4.6.6. Given $\mathbf{p} \in \mathbb{R}^{n-1}$, the problem of finding the closest point to $\mathbf{p}$ in $C_{n, k}$ for sufficiently large $k$ is equivalent to solving a semidefinite program in $\leq n^{2} / 2$ variables with three matrices of size $\leq n^{2} / 4$.

Example 4.6.7. For $n=5$ and $k \geq 3, C_{5, k}$ equals the set of points in $M(\{0,2,5,9\})$ with $m_{0}+m_{2}+$
$m_{5}+m_{9}=1$. Projecting from $M(\{0,1, \ldots, 9\})$, we see that

$$
\begin{aligned}
C_{5, k}=\{ & \left(m_{0}, m_{2}, m_{5}, m_{9}\right) \in \mathbb{R}^{4}: m_{0}+m_{2}+m_{5}+m_{9}=1 \\
& \text { and } \left.\exists\left(m_{1}, m_{3}, m_{4}, m_{6}, m_{7}, m_{8}\right) \in \mathbb{R}^{6} \text { such that }\left(m_{j}\right)_{j=0, \ldots, 9} \in M(\{0,1, \ldots, 9\})\right\}
\end{aligned}
$$

Let $\mathcal{A}(\mathbf{m})$ and $\mathcal{B}(\mathbf{m})$ denote the two $5 \times 5$ matrices appearing in Example 4.6.3. Then $M(\{0,1, \ldots, 9\})$ is the set of points $\mathbf{m} \in \mathbb{R}^{10}$ for which $\mathcal{A}(\mathbf{m}) \geq 0$ and $\mathcal{B}(\mathbf{m}) \geq 0$. Given a point $\mathbf{p}=(a, b, c, d) \in \mathbb{R}^{4}$, we can find the closest point in $C_{5, k}$ by solving the following semidefinite program with 10 parameters and three $5 \times 5$ linear matrix constraints:

$$
\begin{aligned}
& \min _{\lambda, m_{0}, \ldots, m_{9}} \lambda \text { such that } m_{0}+m_{2}+m_{5}+m_{9}=1, \mathcal{A}(\mathbf{m}) \geq 0, \mathcal{B}(\mathbf{m}) \geq 0, \\
& \text { and }\left(\begin{array}{cccccc}
\lambda & m_{0}-a & m_{2}-b & m_{5}-c & m_{9}-d \\
m_{0}-a & 1 & 0 & 0 & 0 \\
m_{2}-b & 0 & 1 & 0 & 0 \\
m_{5}-c & 0 & 0 & 1 & 0 \\
m_{9}-d & 0 & 0 & 0 & 1
\end{array}\right) \geq 0 .
\end{aligned}
$$

If $\left(\lambda^{*}, \mathbf{m}^{*}\right)$ denotes the points achieving this minimum, then $\left(m_{0}^{*}, m_{2}^{*}, m_{5}^{*}, m_{9}^{*}\right)$ is the closest point in $C_{5, k}$ to $\mathbf{p}$ with distance $\sqrt{\lambda^{*}}$.

### 4.7 Discussion and open questions

One takeaway from Section 4.2 is that the points on the boundary of $C_{n, k}$ for $k \geq n-2$ correspond to moment vectors of point evaluations on $[0,1]$. However these do not correspond to biologically meaningful population functions! Similarly, a point in the interior of $C_{n, k}$ can come from several different population functions, some of which are more biologically plausible than others. One natural question from this standpoint is how to pick the right population history from the fiber of a coalescence vector.

Question 4.7.1. Given a point $\mathbf{m}$ in the interior of $M_{k}(A)$, how can we find the "best" step function $f \in S_{k}$ with moment vector $\mathbf{m}$ ?

Here there is some natural flexibility in the notion of "best". Ideally it should be biologically plausible and also easy to compute. For plausibility, it might be reasonable to try to bound or minimize the ratios $y_{i+1} / y_{i}$ of consecutive population sizes. One step towards this would be to understand the structure of the fibers of the moment map $\mu_{A}$.

For $k=2$ and $A=\{0,2,5\}$, the $\left(s_{1}, s_{2}\right)$-coordinates of the fibers of some points in $M_{2}(A)$ are shown below.

To understand the fibers, it may also help to relate the combinatorial structure of the polytope $P$ (which is a product of two $k$-dimensional simplices) to the semi-algebraic and combinatorial structure


Figure 4.8 The central image depicts $M_{2}(A)$ in yellow. The orange region is $M^{\uparrow}(A)$ and the green region $M^{\downarrow}(A)$; their union is $M_{1}(A)$. The triangle above each point depicts the fiber as a subset of the $\left(s_{1}, s_{2}\right)$-simplex.
of $M_{k}(A)$. For example, the boundary of $M_{2}(\{0,2,5,9\})$, seen in Figure 4.4, comes from some of the two-dimensional faces of the four-dimensional polytope $P$.

Question 4.7.2. How does the facial structure of $P$ relate to the algebraic boundary of $M_{k}(A)$ ?
Finally, Section 4.6 gives an algorithm for testing membership in $M(A)$, which coincides with $M_{k}(A)$ for $k \geq|A|-1$. It would be desirable to be able to test membership for smaller $k$ as well.

Question 4.7.3. Is there an effective method to test membership in $M_{k}(A)$ for $k<|A|-1$ ?
These sets are not convex and may have complicated semialgebraic structure (Figure 4.4). One possibility is the following connection to low rank matrix completion would be interesting to explore further.

Consider a step function $f=y_{1} \mathbf{1}_{\left[0, s_{1}\right]}+\sum_{i=2}^{k+1} y_{i} \mathbf{1}_{\left(s_{i-1}, s_{i}\right]}$ in $S_{k}$. In a slight abuse of notation, we define its derivative to be $f^{\prime}=\sum_{i=1}^{k}\left(y_{i+1}-y_{i}\right) \delta_{s_{i}}$, which is a signed weighted sum of delta functions. For $j \in A$, let $m_{j}^{\prime}$ denote the $j$ th moment of the signed measure given by $f^{\prime}$ :

$$
m_{j}^{\prime}=\int_{0}^{1} x^{j} f^{\prime}(x) d x=\sum_{i=1}^{k}\left(y_{i+1}-y_{i}\right)\left(s_{i}\right)^{j} .
$$

One can check that for any $j, m_{j}^{\prime}=f(1)-j m_{j-1}$. In particular, we can write differences of consecutive moments of $f^{\prime}$ in terms of moments of $f$, namely $m_{j}^{\prime}-m_{j+1}^{\prime}=(j+1) m_{j}-j m_{j-1}$.

In the case of full moments $A=\{0,1, \ldots, d\}$, this lets us bound the value of $k$ by the rank of the moment matrix corresponding to the moments of $\left(x-x^{2}\right) f^{\prime}(x)$. Specifically, for $\mathbf{m} \in \mathbb{R}^{A}$, define the matrix

$$
\begin{equation*}
\mathcal{M}(\mathbf{m})=\left((j+\ell+2) m_{j+\ell+1}-(j+\ell+1) m_{j+\ell}\right)_{0 \leq j, \ell \leq\lfloor(d-1) / 2\rfloor} \tag{4.9}
\end{equation*}
$$

Proposition 4.7.4. If for some $f \in S_{k}, m_{j}=\int_{0}^{1} x^{j} f(x) d x$ for all $j$, then $\operatorname{rank}(\mathcal{M}(\mathbf{m})) \leq k$.
Proof. As noted above, we can rewrite the $(j, \ell)$ th entry of $\mathcal{M}(\mathbf{m})$ as

$$
\mathcal{M}(\mathbf{m})_{j, \ell}=(j+\ell+2) m_{j+\ell+1}-(j+\ell+1) m_{j+\ell}=m_{j+\ell+1}^{\prime}-m_{j+\ell+2}^{\prime}=\sum_{i=1}^{k}\left(y_{i+1}-y_{i}\right)\left(s_{i}-s_{i}^{2}\right) s_{i}^{j+\ell}
$$

This shows that $\mathcal{M}(\mathbf{m})$ is a sum of $k$ rank-one matrices $\sum_{i=1}^{k}\left(y_{i+1}-y_{i}\right)\left(s_{i}-s_{i}^{2}\right) v_{e}\left(s_{i}\right) v_{e}\left(s_{i}\right)^{T}$, where $e=\lfloor(d-1) / 2\rfloor$ and $v_{e}(t)=\left(1, t, t^{2}, \ldots, t^{e}\right)^{T}$. Therefore $\mathcal{M}(\mathbf{m})$ has rank $\leq k$.

Note that if the values of $y_{i}$ are increasing then this is a sum of positive semidefinite rank one matrices, in which case the rank of $\mathcal{M}(\mathbf{m})$ will equal $k$, but if the values $y_{i+1}-y_{i}$ have different signs, this might not be the case. Regardless, this suggests the following approach.

Question 4.7.5. Given $\left(m_{a}\right)_{a \in A} \in M(A)$, when does the following low-rank matrix completion find the minimum $k$ for which $\left(m_{a}\right)_{a \in A}$ belongs to $M_{k}(A) ?$ :

$$
\text { Minimize } \operatorname{rank}(\mathcal{M}(\mathbf{m})) \quad \text { such that } \mathcal{A}(\mathbf{m}) \geq 0, \mathcal{B}(\mathbf{m}) \geq 0
$$

Here $\mathcal{A}$ and $\mathcal{B}$ are the matrices introduced in Theorem 4.6.1 and the minimization is taken over all $\mathbf{m} \in \mathbb{R}^{\{0,1, \ldots, \max (A)\}}$ for which $\mathbf{m}_{a}=m_{a}$ for all $a \in A$.

While it seems unlikely that this will always give the correct value, it would be interesting to know how far off this value might be from the true minimal value of $k$. We conclude this chapter by a haiku.

Truncated moments, numerous pre-images.
Fewest steps we seek.

## BIBLIOGRAPHY

[1] Carlos Améndola, Mathias Drton, and Bernd Sturmfels. Maximum likelihood estimates for Gaussian mixtures are transcendental. In Int. Conf. on Math. Asp. of Comput. and Inf. Sci., pages 579-590. Springer, 2015.
[2] Federico Ardila. Algebraic and geometric methods in enumerative combinatorics. arXiv: Combinatorics, 2015.
[3] Federico Ardila and Adam Boocher. The closure of a linear space in a product of lines. J. Algebraic Combin., 43(1):199-235, 2016.
[4] Federico Ardila, Federico Castillo, and José Alejandro Samper. The topology of the external activity complex of a matroid. Electron. J. Combin., 23(3):Paper 3.8, 20, 2016.
[5] Brian Axelrod, Ilias Diakonikolas, Alistair Stewart, Anastasios Sidiropoulos, and Gregory Valiant. A polynomial time algorithm for log-concave maximum likelihood via locally exponential families. In Adv. Neural Inform. Process. Syst. 32, pages 7723-7735. Curran Associates, Inc., 2019.
[6] Miriam Ayer, H. D. Brunk, G. M. Ewing, W. T. Reid, and Edward Silverman. An empirical distribution function for sampling with incomplete information. Ann. Math. Statist., 26:641-647, 1955.
[7] Mark Bagnoli and Ted Bergstrom. Log-concave probability and its applications. Econom. Theory, 26(2):445-469, 2005.
[8] Alan Baker. Transcendental number theory. Cambridge Mathematical Library. Cambridge University Press, Cambridge, second edition, 1990.
[9] Fadoua Balabdaoui and Jon A. Wellner. Estimation of a $k$-monotone density: limit distribution theory and the spline connection. Ann. Statist., 35(6):2536-2564, 2007.
[10] Fadoua Balabdaoui and Jon A. Wellner. Estimation of a $k$-monotone density: characterizations, consistency and minimax lower bounds. Stat. Neerl., 64(1):45-70, 2010.
[11] Daniel J. Bates, Jonathan D. Hauenstein, Andrew J. Sommese, and Charles W. Wampler. Numerically solving polynomial systems with Bertini, volume 25 of Software, Environments, and Tools. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2013.
[12] Anand Bhaskar and Yun S. Song. Descartes' rule of signs and the identifiability of population demographic models from genomic variation data. Ann. Statist., 42(6):2469-2493, 2014.
[13] Anders Björner. Oriented matroids. Cambridge ; New York : Cambridge University Press, Cambridge, 1999.
[14] Anders Björner and Michelle L. Wachs. Shellable nonpure complexes and posets. II. Trans. Amer. Math. Soc., 349(10):3945-3975, 1997.
[15] Grigoriy Blekherman. Convexity properties of the cone of nonnegative polynomials. Discrete \& Computational Geometry, 32(3):345-371, Sep 2004.
[16] Grigoriy Blekherman. Nonnegative polynomials and sums of squares. In Semidefinite optimization and convex algebraic geometry, volume 13 of MOS-SIAM Ser. Optim., pages 159-202. SIAM, Philadelphia, PA, 2013.
[17] Grigoriy Blekherman, Pablo A. Parrilo, and Rekha R. Thomas, editors. Semidefinite optimization and convex algebraic geometry, volume 13 of MOS-SIAM Series on Optimization. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA; Mathematical Optimization Society, Philadelphia, PA, 2013.
[18] Lenore Blum, Felipe Cucker, Michael Shub, and Steve Smale. Complexity and real computation. Springer-Verlag, New York, 1998.
[19] Jacek Bochnak, Michel Coste, and Marie-Françoise Roy. Géométrie algébrique réelle. Berlin ; New York : Springer-Verlag, Berlin, 1987.
[20] Stephen Boyd and Lieven Vandenberghe. Convex optimization. Cambridge University Press, Cambridge, 2004.
[21] Timothy Carpenter, Ilias Diakonikolas, Anastasios Sidiropoulos, and Alistair Stewart. Nearoptimal sample complexity bounds for maximum likelihood estimation of multivariate logconcave densities. In Sébastien Bubeck, Vianney Perchet, and Philippe Rigollet, editors, Proc. 31 Conf. Learn. Theory, volume 75 of Proceedings of Machine Learning Research, pages 1234-1262. PMLR, 06-09 Jul 2018.
[22] Ventsislav Chonev, Joël Ouaknine, and James Worrell. On recurrent reachability for continuous linear dynamical systems. In Proc. of the 31st Annual ACM-IEEE Symposium on Logic in Computer Science (LICS 2016), page 10. ACM, New York, 2016.
[23] Ventsislav Chonev, Joël Ouaknine, and James Worrell. On the Skolem problem for continuous linear dynamical systems. In 43rd International Colloquium on Automata, Languages, and Programming, volume 55 of LIPIcs. Leibniz Int. Proc. Inform., pages Art. No. 100, 13. Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern, 2016.
[24] David A. Cox, John Little, and Donal O'Shea. Ideals, varieties, and algorithms. Undergraduate Texts in Mathematics. Springer, Cham, fourth edition, 2015. An introduction to computational algebraic geometry and commutative algebra.
[25] Madeleine Cule, Robert Gramacy, and Richard Samworth. LogConcDEAD: An R package for maximum likelihood estimation of a multivariate log-concave density. J. Stat. Software, 29(2), 2009.
[26] Madeleine Cule and Richard Samworth. Theoretical properties of the log-concave maximum likelihood estimator of a multidimensional density. Electron. J. Stat., 4:254-270, 2010.
[27] Madeleine Cule, Richard Samworth, and Michael Stewart. Maximum likelihood estimation of a multi-dimensional log-concave density. J. R. Stat. Soc. Ser. B Stat. Methodol., 72(5):545-607, 2010.
[28] Anindya De, Philip M. Long, and Rocco A. Servedio. Density estimation for shift-invariant multidimensional distributions. In 10th Innovations in Theoretical Computer Science, volume 124 of LIPIcs. Leibniz Int. Proc. Inform., pages Art. No. 28, 20. Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern, 2019.
[29] 1966 De Loera, Jesùs A., Jörg Rambau, and 1968 Santos, Francisco. Triangulations : structures for algorithms and applications. Berlin ; New York : Springer, c2010., Berlin, 2010.
[30] Jesús A. De Loera, Bernd Sturmfels, and Cynthia Vinzant. The central curve in linear programming. Found. Comput. Math., 12(4):509-540, 2012.
[31] Ilias Diakonikolas, Daniel M. Kane, and Alistair Stewart. Learning multivariate log-concave distributions. In Satyen Kale and Ohad Shamir, editors, Proc. 2017 Conf. Learn. Theory, volume 65 of Proceedings of Machine Learning Research, pages 711-727, Amsterdam, Netherlands, 07-10 Jul 2017. PMLR.
[32] Alexandru Dimca and SpringerLink (Online service). Hyperplane Arrangements An Introduction. Cham : Springer International Publishing : Imprint: Springer, 2017., Cham, 2017.
[33] Charles R. Doss and Jon A. Wellner. Global rates of convergence of the MLEs of log-concave and $s$-concave densities. Ann. Statist., 44(3):954-981, 2016.
[34] Andreas W.M Dress and Walter Wenzel. Valuated matroids. Advances in Mathematics, 93(2):214 - 250, 1992.
[35] Lutz Dümbgen and Kaspar Rufibach. Maximum likelihood estimation of a log-concave density and its distribution function: basic properties and uniform consistency. Bernoulli, 15(1):40-68, 2009.
[36] Lutz Dümbgen and Kaspar Rufibach. logcondens: Computations related to univariate log-concave density estimation. J. Stat. Software, 39(6):1-28, 2011.
[37] Lutz Dümbgen, Richard Samworth, and Dominic Schuhmacher. Approximation by log-concave distributions, with applications to regression. Ann. Statist., 39(2):702-730, 2011.
[38] P. P. B. Eggermont and V. N. LaRiccia. Maximum likelihood estimation of smooth monotone and unimodal densities. Ann. Statist., 28(3):922-947, 2000.
[39] P. P. B. Eggermont and V. N. LaRiccia. Maximum penalized likelihood estimation. Vol. I. Springer Series in Statistics. Springer-Verlag, New York, 2001. Density estimation.
[40] David Eisenbud. Commutative algebra with a view toward algebraic geometry. New York : Springer-Verlag, [1995], New York, 1995.
[41] Alex Fink, David E. Speyer, and Alexander Woo. A Gröbner basis for the graph of the reciprocal plane. Journal of Commutative Algebra, 12(1):77-86, 2020.
[42] Evelyn Fix and J.L. Hodges Jr. Discriminatory analysis: nonparametric discrimination, consistency properties. USAF school of Aviation Medicine, Randolph Field, Texas, 1951.
[43] Ulf Grenander. On the theory of mortality measurement. II. Skand. Aktuarietidskr., 39:125-153 (1957), 1956.
[44] Piet Groeneboom and Geurt Jongbloed. Nonparametric estimation under shape constraints, volume 38 of Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, New York, 2014. Estimators, algorithms and asymptotics.
[45] Piet Groeneboom, Geurt Jongbloed, and Jon A. Wellner. Estimation of a convex function: characterizations and asymptotic theory. Ann. Statist., 29(6):1653-1698, 2001.
[46] Joe Harris. Algebraic geometry, volume 133 of Graduate Texts in Mathematics. Springer-Verlag, New York, 1995.
[47] Brendan Hassett and Cambridge University Press. Introduction to algebraic geometry. Cambridge : Cambridge University Press, 2007., Cambridge, 2007.
[48] Allen Hatcher. Algebraic topology. Cambridge ; New York : Cambridge University Press, 2002., Cambridge, 2002.
[49] Jonathan D. Hauenstein and Viktor Levandovskyy. Certifying solutions to square systems of polynomial-exponential equations. J. Symbolic Comput., 79(part 3):575-593, 2017.
[50] Jonathan D. Hauenstein and Frank Sottile. Algorithm 921: alphaCertified: certifying solutions to polynomial systems. ACM Trans. Math. Software, 38(4):Art. 28, 20, 2012.
[51] Felix Hausdorff. Summationsmethoden und momentfolgen i. Mathematische Zeitschrift, 9:74109, 1921.
[52] June Huh and Botong Wang. Enumeration of points, lines, planes, etc. Acta Math., 218(2):297317, 2017.
[53] 1972 Jonsson, Jakob. Simplicial complexes of graphs. Berlin ; New York : Springer, c2008., Berlin, 2008. Originally issued as the author's thesis (Ph. D.)-Royal Institute of Technology, Stockholm, 2005.
[54] Arlene K. H. Kim, Adityanand Guntuboyina, and Richard J. Samworth. Adaptation in log-concave density estimation. Ann. Statist., 46(5):2279-2306, 2018.
[55] Arlene K. H. Kim and Richard J. Samworth. Global rates of convergence in log-concave density estimation. Ann. Statist., 44(6):2756-2779, 2016.
[56] Mario Kummer and Cynthia Vinzant. The Chow Form of a Reciprocal Linear Space. Michigan Mathematical Journal, 68(4):831-858, 2019.
[57] Gil Kur, Yuval Dagan, and Alexander Rakhlin. Optimality of maximum likelihood for log-concave density estimation and bounded convex regression. arXiv preprint arXiv:1903.05315, 2019.
[58] Jean Bernard Lasserre. Moments, positive polynomials and their applications, volume 1 of Imperial College Press Optimization Series. Imperial College Press, London, 2010.
[59] Yu Liu and Yong Wang. cnmlcd: Maximum Likelihood Estimation of a Log-Concave Density Function, 2018. R package version 1.2-0.
[60] Diane Maclagan and Bernd Sturmfels. Introduction to tropical geometry, volume 161 of Graduate Studies in Mathematics. American Mathematical Society, Providence, RI, 2015.
[61] Aude Maignan. Solving one and two-dimensional exponential polynomial systems. In Proc. of the 1998 Int. Symp. Symb. and Algebr. Comput., ISSAC '98, pages 215-221, New York, NY, USA, 1998. ACM.
[62] Aude Maignan and Tony C. Scott. Fleshing out the generalized Lambert $W$ function. ACM Commun. Comput. Algebra, 50(2):45-60, 2016.
[63] Murray Marshall. Positive polynomials and sums of squares, volume 146 of Mathematical Surveys and Monographs. American Mathematical Society, Providence, RI, 2008.
[64] Scott McCallum and Volker Weispfenning. Deciding polynomial-transcendental problems. J. Symbolic Comput., 47(1):16-31, 2012.
[65] István Mező. r-Lambert-function. https://github.com/IstvanMezo/ r-Lambert-function, 2017.
[66] István Mező and Árpád Baricz. On the generalization of the Lambert $W$ function. Trans. Amer. Math. Soc., 369(11):7917-7934, 2017.
[67] Mateusz Michałek, Bernd Sturmfels, Caroline Uhler, and Piotr Zwiernik. Exponential varieties. Proc. Lond. Math. Soc. (3), 112(1):27-56, 2016.
[68] Ezra Miller and Bernd Sturmfels. Combinatorial commutative algebra, volume 227 of Graduate Texts in Mathematics. Springer-Verlag, New York, 2005.
[69] James Oxley. Matroid theory, volume 21 of Oxford Graduate Texts in Mathematics. Oxford University Press, Oxford, second edition, 2011.
[70] Jayanta Kumar Pal, Michael Woodroofe, and Mary Meyer. Estimating a Polya frequency function, volume 54 of IMS Lecture Notes Monogr. Ser. Inst. Math. Statist., Beachwood, OH, 2007.
[71] Emanuel Parzen. On estimation of a probability density function and mode. Ann. Math. Statist., 33:1065-1076, 1962.
[72] A. Polanski and M. Kimmel. New explicit expressions for relative frequencies of singlenucleotide polymorphisms with application to statistical inference on population growth. Genetics, 165(1):427-436, 2003.
[73] Nicholas Proudfoot and David Speyer. A broken circuit ring. Beiträge Algebra Geom., 47(1):161166, 2006.
[74] Nicholas Proudfoot, Max Wakefield, and Ben Young. Intersection cohomology of the symmetric reciprocal plane. J. Algebraic Combin., 43(1):129-138, 2016.
[75] Fabian Rathke. Fast multivariate log-concave density estimation in r. https://github.com/ FabianRathke/fmlogcondens, 2018.
[76] Fabian Rathke and Christoph Schnörr. Fast multivariate log-concave density estimation. Computational Statistics \& Data Analysis, 140:41-58, 2019.
[77] Daniel Richardson. Roots of real exponential functions. J. London Math. Soc. (2), 28(1):46-56, 1983.
[78] Elina Robeva, Bernd Sturmfels, Ngoc Tran, and Caroline Uhler. Maximum likelihood estimation for totally positive log-concave densities. Scandinavian Journal of Statistics, 2018.
[79] Elina Robeva, Bernd Sturmfels, and Caroline Uhler. Geometry of log-concave density estimation. Discrete Comput. Geom., 61(1):136-160, 2019.
[80] Zvi S. Rosen, Anand Bhaskar, Sébastien Roch, and Y. Song. Geometry of the sample frequency spectrum and the perils of demographic inference. Genetics, 210:665-682, 2018.
[81] Murray Rosenblatt. Remarks on some nonparametric estimates of a density function. Ann. Math. Statist., 27:832-837, 1956.
[82] Kaspar Rufibach. Computing maximum likelihood estimators of a log-concave density function. J. Stat. Comput. Simul., 77(7-8):561-574, 2007.
[83] Richard J. Samworth. Recent progress in log-concave density estimation. Statist. Sci., 33(4):493509, 2018.
[84] Raman Sanyal, Frank Sottile, and Bernd Sturmfels. Orbitopes. Mathematika, 57(2):275-314, 2011.
[85] Raman Sanyal, Bernd Sturmfels, and Cynthia Vinzant. The entropic discriminant. Adv. Math., 244:678-707, 2013.
[86] Konrad Schmüdgen. The Moment Problem. New York : Springer, New York, 2017.
[87] Konrad Schmüdgen. The moment problem, volume 277 of Graduate Texts in Mathematics. Springer, Cham, 2017.
[88] Georgy Scholten and Cynthia Vinzant. Semi-inverted linear spaces and an analogue of the broken circuit complex. Algebraic Combinatorics, 2(4):645-661, 2019.
[89] David W Scott. Multivariate density estimation: theory, practice, and visualization. Wiley Series in Probability and Statistics. John Wiley \& Sons, Hoboken, New Jersey, second edition, 2015.
[90] E. Shamovich and V. Vinnikov. Livsic-type determinantal representations and hyperbolicity. Advances in Mathematics, 329:487 - 522, 2018.
[91] Michael Shub and Steve Smale. Complexity of Bézout's theorem. I. Geometric aspects. J. Amer. Math. Soc., 6(2):459-501, 1993.
[92] B. W. Silverman. Density estimation for statistics and data analysis. Monographs on Statistics and Applied Probability. Chapman \& Hall, London, 1986.
[93] Rainer Sinn. Algebraic boundaries of convex semi-algebraic sets. Research in the Mathematical Sciences, 2:1-18, 2014.
[94] Steve Smale. Newton's method estimates from data at one point. In The merging of disciplines: new directions in pure, applied, and computational mathematics, pages 185-196. Springer, New York, 1986.
[95] Bernd Sturmfels. Gröbner bases and convex polytopes, volume 8 of University Lecture Series. American Mathematical Society, Providence, RI, 1996.
[96] Hiroaki Terao. Algebras generated by reciprocals of linear forms. J. Algebra, 250(2):549-558, 2002.
[97] Alexandre B. Tsybakov. Introduction to nonparametric estimation. Springer Series in Statistics. Springer, New York, 2009. Revised and extended from the 2004 French original, Translated by Vladimir Zaiats.
[98] A. Varchenko. Critical points of the product of powers of linear functions and families of bases of singular vectors. Compositio Math., 97(3):385-401, 1995.
[99] Cynthia Vinzant. What is . . . a spectrahedron? Notices Amer. Math. Soc., 61(5):492-494, 2014.
[100] G. Walter and J. Blum. Probability density estimation using delta sequences. Ann. Statist., 7(2):328-340, 1979.
[101] Guenther Walther. Detecting the presence of mixing with multiscale maximum likelihood. J. Amer. Statist. Assoc., 97(458):508-513, 2002.
[102] Larry Wasserman. All of nonparametric statistics. New York : Springer, 2006., New York, 2006.
[103] Ming Xu, Zhi-Bin Li, and Lu Yang. Quantifier elimination for a class of exponential polynomial formulas. J. Symbolic Comput., 68:146-168, 2015.
[104] Günter M. Ziegler. Lectures on polytopes. New York : Springer-Verlag, New York, 1995.

