

Nonlinear Algebra in Likelihood, Neurocomputing and Quantum Physics

Der Fakultät für Mathematik und Informatik
der Universität Leipzig
eingereichte

D I S S E R T A T I O N

zur Erlangung des akademischen Grades

DOCTOR RERUM NATURALIUM
(Dr. rer. nat.)

im Fachgebiet

Mathematik

vorgelegt

von MAST Maximilian Wiesmann
geboren am 5. September 1999 in Düsseldorf

Leipzig, den 27. Januar 2025

Authorship

The results presented in this thesis are based on six articles of mine which were written in collaboration with other researchers. Parts of this thesis are prepublished.

Chapter 1 is an introduction and written by myself.

Chapter 2 contains background material known from the literature. There are no original contributions in this chapter. Parts of this chapter are taken from the expositions in some of my articles: Section 2.2 partially appeared in [DHW23]; Section 2.3 partially follows [KKM⁺24a, §2]; Subsection 2.3.1 is carried over from [KLW24a, §3.1]; some material in Section 2.4 is taken from [KKM⁺24a, §2, §5]; Section 2.5 carries over the expository sections in [KLW24a, §2, §6]; Section 2.6 is carried over from [DPW23a, §2, App. A].

Chapter 3 is based on three articles. Section 3.1 is based on the paper [DHW23], jointly written with Eliana Duarte and Ben Hollering. All coauthors contributed equally to all aspects of the publication and mathematical results presented in the article. The article appeared in *Geometric Science of Information*, the proceedings to the homonymous conference. Passages from this article are carried over to all subsections of Section 3.1, except Subsection 3.1.2, which is written by myself and contains unpublished material.

Section 3.2 is based on the paper [TW24a], jointly written with Simon Telen. All coauthors contributed equally to all aspects of the publication and mathematical results presented in the article. The article was submitted to the *Journal of the European Mathematical Society* and is currently under review. Passages from this article are carried over to all subsections of Section 3.2.

Section 3.3 is based on the paper [KKM⁺24a], jointly written with Thomas Kahle, Lukas Kühne, Leonie Mühlherr and Bernd Sturmfels. All coauthors contributed equally to all aspects of the publication and mathematical results presented in the article. The article was submitted to the *Vietnam Journal of Mathematics* and is currently under review. Passages from this article are carried over to all subsections of Section 3.3, with the exception of Subsection 3.3.4. Subsection 3.3.4 is written by myself and contains unpublished material.

Chapter 4 is based on the paper [KLW24a], jointly written with Kaie Kubjas and Jiayi Li. All coauthors contributed equally to all aspects of the publication and mathematical results presented in the article. The article appeared in *Algebraic Statistics*. Passages from this article are carried over to all Sections of Chapter 4.

Chapter 5 is based on two articles. Section 5.1 is based on the paper [DPW23a], jointly written with Eliana Duarte and Dmitrii Pavlov. All coauthors contributed equally to all aspects of the publication and mathematical results presented in the article. The article was submitted to *Advances in Applied Mathematics* and is currently under review. Passages from this article are carried over to all subsections of Section 5.1.

Section 5.2 is based on the paper [BMW24], jointly written with Michael Borinsky and Chiara Meroni. All coauthors contributed equally to all aspects of the publication and mathematical results presented in the article. The article got accepted for publication in *Le Matematiche*. Passages from this article are carried over to all subsections of Section 5.2.

Acknowledgements

There are many people to whom I would like to express my gratitude. First and foremost, I would like to thank my two advisors, Eliana Duarte and Bernd Sturmfels. Eliana, thank you for guiding me into algebraic statistics and for always being there with your advice, even after you left the institute. Bernd, thank you so much for suggesting countless problems, connecting me with so many people, your constant encouragement and the energy you bring to the institute.

I would like to thank my friends and fellow PhD students, Barbara, Dmitrii, Laura and Leonie. You are a significant part of the reason why it was always a delight to come to the institute. You provided mental support and encouragement whenever I needed it. Thank you for every coffee break, fantastic party, kebab dinner or any other activity we shared.

This thesis is a result of my work at MPI MiS. The institute is a fantastic place because of the incredible people working there, and I would like to thank all nonlinear algebra members who have worked there during my time for making it such a welcoming place. Special thanks go to Simon, who, in the absence of my group leader, adopted me into his group; and to Ben who introduced me to algebraic statistics. Huge thanks also go to Mirke and Saskia who made the administrative aspects of my work life so much easier.

I am grateful for the opportunity to visit IMSI Chicago for a month to participate in their programme on algebraic statistics. Thank you to all the long-term visitors who made the institute a productive research environment and the stay an enjoyable experience.

To me, mathematics is most fun when it is done in collaboration with others. Thank you, Michael Borinsky, Eliana Duarte, Ben Hollering, Thomas Kahle, Kaie Kubjas, Lukas Kühne, Paul Larsen, Jiayi Li, Chiara Meroni, Leonie Mühlherr, Dmitrii Pavlov, Bernd Sturmfels and Simon Telen, for being fantastic coauthors.

I would like to thank my parents for their unwavering support of my mathematical endeavours. Last but by no means least, thank you, Evelyn, for your love and support.

Contents

List of Symbols	9
List of Figures	13
List of Tables	15
1 Introduction	17
2 Background	21
2.1 Toric varieties	22
2.2 Algebraic statistics	24
2.3 Commutative algebra	29
2.3.1 Symmetric tensor decomposition	30
2.4 Hypersurface arrangements	31
2.5 Neural networks	33
2.6 Quantum information theory	36
3 Likelihood geometry	41
3.1 Toric fibre products: Horn matrices and geometric modelling	42
3.1.1 Toric fibre products via point configurations	42
3.1.2 Fibre products in the category of toric varieties	43
3.1.3 The Horn matrix of ML degree one toric fibre products	46
3.1.4 Blending functions	48
3.1.5 Blending functions of toric fibre products	50
3.2 Euler stratifications of hypersurface families	54
3.2.1 Definitions and first examples	56
3.2.2 Points on the line	60
3.2.3 Projective hypersurfaces	64
3.2.4 Very affine hypersurfaces	67
3.2.5 Applications	72
3.2.6 Binary octics	72
3.2.7 Matroid stratification of bilinear forms	73
3.2.8 Feynman integrals	73
3.2.9 Maximum likelihood estimation for toric models	75
3.2.10 Hirzebruch surfaces	76
3.3 Arrangements and parametric likelihood	78
3.3.1 Arrangements and modules	78
3.3.2 Parametric likelihood in statistics and physics	83

3.3.3	Gentle, free and tame arrangements	86
3.3.4	Generic arrangements	88
3.3.5	Graphic arrangements	93
3.3.6	Software and computations	96
3.4	Conclusion	98
4	Polynomial neural networks	99
4.1	Introduction	100
4.2	Expressivity	102
4.2.1	Neuromanifolds and symmetric tensor decomposition	105
4.2.2	Neurovarieties	109
4.2.3	Dimension	111
4.2.4	The symmetries of an exceptional shallow network	114
4.3	Optimisation landscape	116
4.3.1	The learning degree of a PNN	116
4.3.2	Machine learning experiment	120
4.3.3	Case study: linear neural networks	122
4.4	Conclusion	123
5	Quantum physics	125
5.1	Quantum graphical models	126
5.1.1	Quantum conditional mutual information varieties	127
5.1.2	Petz varieties	131
5.1.3	Quantum graphical models from Gibbs manifolds	134
5.1.4	Gibbs varieties of linear systems of Hamiltonians	134
5.1.5	Gibbs varieties of unirational varieties	135
5.1.6	Quantum exponential families of commuting Hamiltonians	136
5.1.7	Quantum information projections	139
5.2	Exponential integrals and edge-bicoloured graphs	141
5.2.1	Laplace method and asymptotic expansions	141
5.2.2	Edge-bicoloured graphs	144
5.2.3	Efficient computation of the coefficients A_n	148
5.2.4	Asymptotics and critical points	149
5.3	Conclusion	156
Bibliography		156
Index		170

List of Symbols

$(2s - 1)!!$	double factorial $(2s - 1)(2s - 3) \cdots 3 \cdot 1$, 142
\dagger	complex conjugate transpose, 38
\prec	partition refinement, 60
\star	(entrywise) Hadamard product, 29
$ \psi\rangle$	quantum state, element of Hilbert space, 36
$[n]$	the set $\{1, 2, \dots, n\}$, 24
$\mathcal{A}(G)$	graphic arrangement for graph G , 33
$\text{Aut}(G)$	automorphism group of a labelled graph representing G , 145
$\text{Aut}(\Gamma)$	automorphism group of a labelled graph, 145
\mathcal{A}_X	localisation of hyperplane arrangement, 33
$A(X)$	Chow ring of X , 65
$B \times_{\mathcal{A}} C$	toric fibre product of point configurations, 43
$\chi(X)$	(topological) Euler characteristic, 28
cl_{Eucl}	closure w.r.t. Euclidean topology, 108
$c^M(X)$	total Chern–Mather class of X , 118
$c_i^M(X)$	i^{th} Chern–Mather class of X , 118
$\text{conv}(A)$	convex hull of the columns of A , 24
$D(\mathcal{A})$	log-derivation module of the arrangement \mathcal{A} , 79
$\deg(X)$	degree of projective variety X , 61
$\partial_{\text{alg}}(\mathcal{M})$	algebraic boundary of \mathcal{M} , 109
∂_{x_i}	short for $\frac{\partial}{\partial x_i}$, 69
Δ_{n-1}°	open probability simplex, 24
Δ_χ	Euler discriminant polynomial, 55
$\text{Der}_k(S)$	S -module of k -linear derivations, 32
$\text{Der}_S(\mathcal{A})$	module of logarithmic \mathcal{A} -derivations, 32
$D(F)$	hypersurface complement $\{x \in \mathbb{P}^d : F(x) \neq 0\}$, 57
$\mathcal{D}(\mathcal{H})$	set of all density matrices on Hilbert space \mathcal{H} , 37

$D(\rho \sigma)$	quantum relative entropy, 139
$\text{EDdeg}_{\text{gen}}(X)$	generic Euclidean distance degree of a variety X , 116
$\text{edim}(\mathcal{V}_{\mathbf{d},r})$	expected dimension of neurovariety, 103
F_θ	feedforward neural network with parameters θ , 33
f^s	likelihood function for data s , 78
\mathcal{G}	set of isomorphism classes of graphs, 145
\mathcal{G}_{-n}^*	set of isomorphism classes of edge-bicoloured graphs with vertex degrees at least three and Euler characteristic equal to $-n$, 147
$\text{GM}(\mathcal{L})$	Gibbs manifold of \mathcal{L} , 134
$\text{Gr}(d, V)$	Grassmannian of d -dimensional subspaces of V , 118
$\text{Gr}(k, n)$	Grassmannian of k -dimensional subspaces of \mathbb{R}^n , 111
$\text{GV}(\mathcal{L})$	Gibbs variety of \mathcal{L} , 134
$I(\mathcal{A})$	likelihood ideal of arrangement \mathcal{A} , 80
$I_0(\mathcal{A})$	pre-likelihood ideal of arrangement \mathcal{A} , 81
$I(A : C B)$	quantum conditional mutual information, 127
Id_n	$n \times n$ identity matrix, 73
$J(\mathcal{A})$	Jacobian syzygy module of arrangement \mathcal{A} , 79
$\mathcal{L}_{\mathcal{A}}$	(parametric) likelihood correspondence of an arrangement \mathcal{A} , 80
$L(\mathcal{A})$	intersection lattice of hyperplane arrangement \mathcal{A} , 33
$\lambda \vdash n$	λ is an integer partition of n , 72
$ \lambda $	length of a partition λ , 61
$\text{Ldeg}_\ell(p_{\mathbf{w}})$	learning degree of the PNN $p_{\mathbf{w}}$ w.r.t. loss ℓ , 116
$\ell_u(p)$	log-likelihood function w.r.t. data u , 26
\mathcal{L}_X	likelihood correspondence of a variety X , 27
\mathcal{M}	statistical model, 24
$\mathcal{M}_{0,n}$	moduli space of n labelled points in \mathbb{P}^1 , 86
\mathcal{M}_A	log-linear model, 24
$\mathcal{M}_{A,z}$	log-affine model, 24
$M(\mathcal{A})$	likelihood module of arrangement \mathcal{A} , 79
$\mathcal{M}_{\mathbf{d},r}$	neuromanifold, 101
$\text{MLdeg}(X)$	maximum likelihood degree of X , 26
$\text{MLdeg}(\mathcal{A})$	(parametric) ML degree of an arrangement \mathcal{A} , 83
$\mu(f, x)$	(local) Milnor number of f at x , 59
$\mu(X)$	total Milnor number of X , 59
$\text{mult}_x X$	multiplicity of $x \in X$, 63
∇_χ	Euler discriminant variety, 55

∇F	Gauss map, 57
∇_λ	Zariski closure of S_λ , 61
$\nabla_{m_0 \lambda m_\infty}$	Zariski closure of $S_{m_0 \lambda m_\infty}$, 62
NoToVar	category of (abstract) normal toric varieties, 44
$\mathcal{O}(h(z))$	big-O notation, 142
$\Omega_{T/S}^p$	T -module of S -valued Kähler p -forms, 32
$\Omega_{T/S}^p(\mathcal{A})$	module of logarithmic p -forms with poles along \mathcal{A} , 32
$\Omega_0^p(\mathcal{A})$	pruned module of logarithmic p -forms, 90
$\mathcal{O}_X(d)$	line bundle on projective variety X , 63
\mathbb{P}^n	n -dimensional complex projective space, 22
p_+	the sum $p_1 + p_2 + \dots + p_n$, 26
\mathcal{P}_G	Petz variety of tree G , 132
$\phi_{A,z}$	monomial map corresponding to $A \in \mathbb{Z}^{d \times n}$ and $z \in (\mathbb{C}^*)^n$, 22
$\varphi_{(H,\lambda)}$	Horn parametrisation, 29
$\mathcal{P}(n)$	number of integer partitions of n , 62
\mathcal{P}_N	Pauli group, 38
$\Psi_{\mathbf{d},\sigma}$	parameter map for neural networks, 34
$\mathcal{P}_{s,t}$	set of partitions of $\{1, \dots, s\} \sqcup \{1, \dots, t\}$, 146
$p_{\mathbf{w}}$	polynomial neural network with weights \mathbf{w} , 101
\mathcal{Q}_G	QCMI variety of tree G , 130
$\mathcal{Q}_{I(A:C B)}$	QCMI variety of 3-chain, 128
$\text{Relint}(P)$	relative interior of the polytope P , 49
\mathcal{R}_G	Petz recovery map for graph G , 131
ρ	density matrix, endomorphism on Hilbert space, 36
$\mathcal{R}(M)$	Rees algebra of a module M , 29
\mathcal{S}	Euler stratification, 56
(Σ, N)	fan Σ with ambient lattice N , representing an abstract normal toric variety, 44
σ_X	Pauli-X matrix, 37
σ_Y	Pauli-Y matrix, 37
σ_Z	Pauli-Z matrix, 37
S_λ	coincident root locus of partition λ , 60
$S_{m_0 \lambda m_\infty}$	coincident root locus in \mathbb{C}^* , 62
\mathbb{S}^n	space of symmetric $n \times n$ matrices, 127
\mathfrak{S}_n	symmetric group of degree n , 145
S^n	n -dimensional sphere, 149

$S_{d,k}^n(\mathbb{K})$	set of symmetric tensors $T \in \text{Sym}_d(\mathbb{K}^n)$ with symmetric rank k , 107
$S(\rho)$	von Neumann entropy of ρ , 127
$\text{Sym}(M)$	symmetric algebra of M , 29
θ_E	Euler derivation, 32
θ_k	Saito derivation, 95
θ_C^T	separator-based derivation, 95
Tr_A	partial trace over A , 37
u_+	the sum $u_1 + u_2 + \dots + u_n$, 26
u^v	for two vectors of the same dimension, $u^v = \prod_{i=1}^r u_i^{v_i}$, 29
$\mathcal{V}(I)$	vanishing locus of ideal I , 22
$\mathcal{V}_{\mathbf{d},r}$	neurovariety, 101
$V_{m,n,r}$	variety of $m \times n$ matrices with rank at most r , 118
$X^{>0}$	positive part of variety X , 26
X^\vee	dual variety of X , 63
X_A	projective toric variety (same as $X_{A,(1,1,\dots,1)}$), 22
$X_{A,z}$	scaled projective toric variety, 22
$X_{B \times_{\mathcal{A}} C, z_{B \times_{\mathcal{A}} C}}$	toric fibre product toric variety, 43
\mathcal{X}_F	family of hypersurface complements $\{(x, z) \in \mathbb{P}^d \times Z : F(x, z) \neq 0\}$, 57
\mathcal{X}_f^*	family of very affine hypersurface complements $\{(x, z) \in T \times Z : f(x, z) \neq 0\}$, 58
\mathcal{X}_F^c	family of hypersurfaces $\{(x, z) \in \mathbb{P}^d \times Z : F(x, z) = 0\}$, 57
$\mathcal{X}_{F,z}$	fibre of the map $\pi_F: \mathcal{X}_F \rightarrow Z$ over z , 57
$\mathcal{X}_{F,z}^c$	fibre of the map $\pi_F: \mathcal{X}_F^c \rightarrow Z$ over z , 57
X_{reg}	regular locus of X , 26
X_{sing}	singular locus of X , 59
$X \perp\!\!\!\perp Y Z$	conditional independence of X and Y given Z , 25

List of Figures

2.1	The polytope associated to the Segre embedding X_A from Example 2.1.2 (left) and its normal fan, the normal fan of $\mathbb{P}^1 \times \mathbb{P}^1$ (right).	24
2.2	An illustration of Birch's Theorem: the MLE (green) is the unique intersection point of a linear space (red) with the positive part (blue) of a toric variety (blue and yellow).	28
2.3	A neural network architecture with widths $\mathbf{d} = (2, 3, 3, 1)$, input $\mathbf{x} = (x_1, x_2)^T$ and output y_1	34
3.1	Toric fibre product of the point configurations B and C in Example 3.1.5. Each point configuration is displayed as a matrix with its corresponding convex hull below. The yellow vertices in each polytope have degree e_1 while the red vertices in each polytope have degree e_2 in the associated multigrading \mathcal{A} . The degree map is $\deg(\mathbf{b}_j^i) = \deg(\mathbf{c}_k^i) = \mathbf{a}^i$	44
3.2	Left: a toric morphism $\mathbb{P}^2 \rightarrow \mathbb{P}^1$ needs to be constant. Right: there is no nontrivial multigrading of the Veronese surface.	45
3.3	The Euler stratification of plane cubics.	60
3.4	Young diagrams with five boxes index the strata of $\pi: \mathcal{X}^c \subset \mathbb{P}^1 \times \mathbb{P}^5 \rightarrow \mathbb{P}^5$	61
3.5	The Euler stratification of five points in \mathbb{C}^*	62
3.6	Three Feynman diagrams: bubble, sunrise and parachute.	74
3.7	Euler stratifications for hyperplane sections of five toric Fano surfaces.	76
3.8	The octahedron and its edge graph.	94
4.1	The neuromanifold $\mathcal{M}_{(2,1,1),2}$ in $\text{Sym}_2(\mathbb{R}^2) \cong \mathbb{R}^3$	103
4.2	Coordinatewise comparison of the function most frequently learned by a PNN with $\mathbf{d} = (2, 2, 3)$ and $r = 2$ (red graph) with the ground truth (green graph).	122
5.1	Left: the graph G . Right: illustration of the Hamiltonian H_1	136
5.2	An edge-bicoloured graph with two connected components.	144
5.3	The system (5.2.9) for the function V from Example 5.2.16. Left: all values of $\lambda \in (0, 4)$ on the (reversed) vertical axis. At each level $\lambda = \text{const.}$ the black continuous curves are the maxima; the dashed curves are the minima. In blue, the curves for $\lambda = \frac{1}{3}$ and $\lambda = 3$ where the behaviour of the maxima changes. Right: the section $\lambda = 2$, with its maxima (squares) and its minima (crosses).	151
5.4	The behaviour of $\alpha(\lambda)$ and $c(\lambda)$ in the Ising model from Example 5.2.16. The phase transitions at $\lambda = \frac{1}{3}, 3$ can be detected in both quantities. At $\lambda = 3$, the function α is C^1 - but not C^2 -differentiable.	152

- 5.5 The system (5.2.14) for the function V from Example 5.2.19, for the values $\lambda = \frac{1}{2}, \frac{1}{4}, 4$, from left to right. The solutions are marked in black. The solutions that are (equally) closest to (but distinct from) the origin, are marked with squares. Notice the different scaling in the y -axis, for the sake of clarity. 154
- 5.6 The roots of $A_n(\lambda)$ under the assumptions of Example 5.2.2, for $n = 10, 25, 200$. The blue crosses are the phase transitions $\lambda = \frac{1}{3}, 3$ 155

List of Tables

3.1	Degrees of generators for the ideal of coincident root loci of a binary octic. . .	72
4.1	Properties of neurovarieties for shallow polynomial neural networks with widths $d_i \in [3]$ and activation degree $r = 2$	104
4.2	An overview of architectures for which we give results about their neuro-manifolds in Subsection 4.2.1.	105

Chapter 1

Introduction

Die Mathematik sollte stets verknüpft gehalten werden mit allem, was den Menschen bewegt.

— Felix Klein

Nonlinear algebra [DM07, MS21] concerns the study of solution sets to systems of polynomial equations, with a focus on computations and applications. This focus distinguishes nonlinear algebra from classical algebraic geometry. Wherever polynomials appear in areas like statistics, data science, optimisation or physics, the nonlinear algebraist enters the stage. Their toolbox comprises tools from algebraic geometry, but also other branches of mathematics like combinatorics, commutative algebra or topology. Importantly, the nonlinear algebraist is never missing their computer: often, the desired output to a problem is a (certified) computational result or an algorithm. Moreover, in the spirit of experimental mathematics [BBG04], computer-aided examples provide a useful guide for novel results. This has been made possible by relatively recent advances in computer algebra systems [GS02, OSC24] and numerical algebraic geometry software [BT18]. The focus on applications is a bidirectional interaction: the mathematical tools address the applied problem, while the application often inspires new and intrinsically valuable mathematical concepts.

The applications treated in this thesis are threefold. In Chapter 3 the problem of maximum likelihood estimation (MLE) is studied from an algebro-geometric perspective. MLE is ubiquitous in statistical applications as it connects the statistical model with real-world data, see e.g. [Fel81, HTF09, Jam06]. If the statistical model at hands is representable by an algebraic variety, which constitutes the framework of algebraic statistics [Sul18], MLE amounts to solving a polynomial system of equations. The second application from the title, neurocomputing, is addressed in Chapter 4 in the context of polynomial neural networks. The output of a network of this class is a tuple of polynomials in the input data. Although neural networks are a core concept of machine learning and widely used with overwhelming success [GBC16], the theoretical understanding of this success is often missing. For polynomial neural networks, we can use nonlinear algebra to study the expressivity and the optimisation process of these networks. The final Chapter 5 contains interactions between nonlinear algebra and quantum physics. In quantum information theory, quantum states are represented by density matrices [NC02]. In the first part of the chapter, we associate algebraic varieties to sets of states satisfying certain conditions on their mutual information depending on the structure of a graph. This generalises the use of nonlinear algebra in classical graphical models to the quantum world. The second

part of the chapter draws inspiration from perturbative quantum field theory [BIZ80] to enumerate edge-coloured graphs.

A common thread between the chapters is the appearance of a set of critical equations governing the problem. In Chapter 3 these are the likelihood equations whose solutions are the critical points of the log-likelihood function. In Chapter 4 these are the critical equations of the loss function appearing in the training process of the network. And in Chapter 5 these are on the one hand the equations computing the quantum information projection (the quantum analogue of MLE) and on the other hand the critical points governing the asymptotics of the number of edge-coloured graphs. Of particular interest is the number of solutions to these equations. This is often a degree, an invariant of the algebraic variety underlying the problem. In the context of MLE it is the ML degree [CHKS06], for training of polynomial neural networks we introduce the notion of learning degree. The computation of these numbers relates to such fields as intersection theory [Ful13], singularity theory [GLS07] and discriminants [GKZ08]. All of these topics appear in this thesis.

In the following, we give an overview of the results presented in this thesis. The process of maximum likelihood estimation associates to data a point in a statistical model that best describes the data. Questions relating the geometry of MLE [HS14] are studied in Chapter 3. A particular emphasis is laid on models representable by toric varieties. These correspond to log-affine models in statistics. The use of toric varieties [CLS11] opens up a wide range of geometric, algebraic and combinatorial tools to study MLE. The number of critical points of the log-likelihood function on the underlying variety is the ML degree. It captures the algebraic complexity of MLE for a particular model. In the special case that the ML degree equals one, the algebraic variety representing the model admits a beautiful rational parametrisation by associating to a data point its maximum likelihood estimate. This parametrisation has been described by June Huh in terms of the Horn matrix of the variety [Huh14]. In Section 3.1 we provide an explicit description of the Horn matrix for toric fibre products [Sul07]. This construction allows one to build higher dimensional models with ML degree one from lower dimensional ones. Moreover, we connect toric fibre products to geometric modelling. This field aims to find simple parametrisations of shapes for modelling in a computer via so-called blending functions. Here, models with ML degree one play a special role since they admit blending functions satisfying rational linear precision [GPS10]. We explicitly construct blending functions with this property for a toric fibre product. This section is based on the article [DHW23].

Log-affine statistical models include a scaling of the underlying toric variety. This scaling affects the ML degree of the model [ABB⁺19]. It is a difficult problem to determine which scaling leads to which ML degree. This problem is studied in Section 3.2, based on the article [TW24a]. A key fact employed is that the ML degree is the Euler characteristic of a very affine variety [Huh13]. Motivated by this, we introduce Euler stratifications. This is a stratification of the base locus of a parametric family of varieties such that the Euler characteristic of the fibre is constant over each stratum. Besides several structural results, we develop algorithms to compute Euler stratifications for projective and very affine hypersurface families. Since an Euler stratification of a very affine hypersurface family captures the dependence of the ML degree of a toric variety on the scaling parameters completely, this gives a computational answer to the problem described above.

A common perspective in algebraic statistics is to describe a statistical model implicitly via the ideal cutting out the underlying variety. In other branches of statistics it is more natural to work with a parametric description of the model. We take up this point of view in Section 3.3 by introducing the parametric likelihood correspondence. This connects data to critical points of the parametric likelihood equations. We study this correspondence by

relating it to the theory of hypersurface arrangements, and in particular to modules of logarithmic derivations [OT13]. The main result identifies the ideal defining the likelihood correspondence as the presentation ideal of the Rees algebra of the likelihood module, a module related to logarithmic derivations. This gives rise to a new way of computing this ideal which proves to be more efficient in practice than previous methods. The computation is especially simple if the hypersurface arrangement is gentle; we investigate this new arrangement property. This section is based on the paper [KKM⁺24a].

We then turn to neurocomputing, more precisely polynomial neural networks [KTB19], in Chapter 4. Two questions are of most concern to us: What is the expressivity [RPK⁺17] of the network, i.e. which functions can the network represent exactly? And, what does the optimisation process [GBC16, Part II, §8] look like? More concretely, how many functions does the network learn in an empirical risk minimisation procedure after different initialisations? The former question asks for a description of the neuromanifold, also called functional space, and is studied in Section 4.2. For polynomial neural networks, this space is a semialgebraic set that can be described by polynomial equations and inequalities. A good approximation to this space is its Zariski closure, the neurovariety. We study both neuromanifolds and neurovarieties for several network architectures. For shallow networks, these spaces relate to objects known from the theory of tensors [Lan11].

The latter question is addressed in Section 4.3. We introduce the learning degree of a network, an invariant of the neurovariety counting the number of nonsingular critical points of the loss function on the neurovariety; here, we enter the field of metric algebraic geometry [BKS24]. The learning degree provides an upper bound on the number of functions a polynomial neural network can learn via commonly applied training methods like gradient descent after different initialisations. It is bounded by the generic Euclidean distance degree [DHO⁺15], a common notion in metric algebraic geometry. We argue that our concept of learning degree provides meaningful and novel insights into the network's training process. Chapter 4 is based on the article [KLW24a].

The final Chapter 5 explores interactions between nonlinear algebra and quantum physics. In algebraic statistics, there has been a very successful approach to associate algebraic varieties to graphical models, where a graph represents the conditional independence structure between random variables. Section 5.1 aims to generalise this approach to quantum information theory [NC02] by associating algebraic varieties to quantum graphical models, based on the paper [DPW23a]. Such models encode sets of quantum states satisfying certain constraints on the mutual information between the states of subsystems [LP08]. Quantum states are represented by density matrices. If we restrict to diagonal matrices then our varieties associated to quantum graphical models reduce to the classical graphical models in algebraic statistics. One can therefore view quantum graphical models as a noncommutative generalisation of classical graphical models. The quantum analogue to MLE is the quantum information projection, minimising the quantum relative entropy between a set of quantum states (the model) and an additional state. We prove a generalisation of Birch's Theorem (a well-known result for MLE on log-affine models) for quantum information projections to quantum exponential families of certain Hamiltonians.

Lastly, in Section 5.2, we turn to the combinatorial problem of enumerating edge-coloured graphs, based on the paper [BMW24]. Here, we draw inspiration from perturbative quantum field theory, where a path integral is expanded in terms of Feynman diagrams, see e.g. [Wei22, Ch. 4]. Building on this idea, we relate the generating function of edge-bicoloured graphs to the asymptotic expansion of an exponential integral. This representation then allows us to find a formula for the asymptotic number of regular edge-bicoloured graphs with large Euler characteristic. The formula involves certain criti-

cal points of a polynomial governing the vertex incidence structure of the graphs. This is where nonlinear algebra enters the game. Interestingly, one observes “phase transitions” in the asymptotics when a coupling parameter between the colours is varied. We relate this phenomenon to phase transitions in the Ising model from statistical physics [DC23].

Gian-Carlo Rota, founder of modern combinatorics and philosopher, remarked the following: “Every field has its taboos. In algebraic geometry the taboos are (1) writing a draft that can be followed by anyone but two or three of one’s closest friends, (2) claiming a result has applications, (3) mentioning the word ‘combinatorial’ [...].” [Rot97, p. 231]. In nonlinear algebra, we are allowed to break all of these taboos. We hope the reader might enjoy this and in particular will find taboo (1) to be violated.

Chapter 2

Background

In this chapter we introduce preliminary material providing the background for the following chapters. The choice of topics explained here is very subjective and reflects the author's (un)knowledge. The reader might wish to skip some sections, revisit the chapter at a later point or consult the literature that is referenced throughout the chapter for more details. An exposition of toric varieties is provided in Section 2.1. Toric varieties play a crucial role throughout Chapter 3 as they represent log-affine models in algebraic statistics. They reappear in the context of quantum exponential families in Section 5.1. Algebraic statistics is introduced in Section 2.2, including undirected graphical models and the fundamentals of likelihood geometry. Parts of this section appeared in [DHW23]. In Section 2.3, partially following [KKM⁺24a, §2], we collect some results from commutative algebra; these will mostly be used in Section 3.3. Moreover, we introduce symmetric tensor decomposition which appears in Section 4.2 in the context of polynomial neural networks; this is taken from [KLW24a, §3.1]. An exposition on hypersurface arrangements is found in Section 2.4. Some of this is material taken from [KKM⁺24a, §2, §5]. The notions discussed therein are used in Section 3.3. In Section 2.5 we introduce basic notions of neural networks, following the expository sections in [KLW24a, §2, §6]. Finally, in Section 2.6, we present the fundamentals of quantum information theory necessary for Section 5.1 on quantum graphical models. This exposition is taken from [DPW23a, §2, App. A].

2.1. Toric varieties

Toric varieties are a class of algebraic varieties that exhibit, besides their algebraic and geometric structure, a rich combinatorial structure. In pure mathematics, this makes them a common testing ground for conjectures, for example in the context of mirror symmetry, see e.g. [CK99]. Despite being a very specific class of varieties, one might argue that most varieties a human comes up with are in fact toric. Moreover, nature seems to favour toric varieties as well, as has been argued with the slogan “the world is toric” in the book [MS21]. Therein, the authors argue that toric varieties naturally occur in many applications of nonlinear algebra to the sciences. In this thesis, we encounter toric varieties in the contexts of algebraic statistics (see Section 2.2) and geometric modelling (see Section 3.1), and as certain quantum exponential families in Subsections 5.1.6 and 5.1.7.

Let $A \in \mathbb{Z}^{d \times n}$ be an integer matrix, let $a_i \in \mathbb{Z}^d$ be the i^{th} column of A and let $t \in (\mathbb{C}^*)^d$ be a point in the algebraic torus. We write t^{a_i} shorthand for $\prod_{j=1}^d t_j^{a_{ji}}$. Moreover, let $z \in (\mathbb{C}^*)^n$ be a vector of nonzero complex scalings. This data gives rise to a monomial map

$$\phi'_{A,z}: (\mathbb{C}^*)^d \rightarrow (\mathbb{C}^*)^n, \quad t \mapsto (z_1 t^{a_1}, \dots, z_n t^{a_n}).$$

The projective space \mathbb{P}^{n-1} contains an $(n-1)$ -dimensional dense torus

$$T = \mathbb{P}^{n-1} \setminus \mathcal{V}(x_1 \dots x_n) = \{(1 : t_1 : \dots : t_{n-1}) : t_1, \dots, t_{n-1} \in \mathbb{C}^*\} \cong (\mathbb{C}^*)^{n-1}.$$

The scaling invariance of homogeneous coordinates gives rise to a short exact sequence

$$1 \rightarrow \mathbb{C}^* \rightarrow (\mathbb{C}^*)^n \xrightarrow{\pi} T \rightarrow 1.$$

We define the map $\phi_{A,z}$ to be the composition of the monomial map $\phi'_{A,z}$ with π , followed by the inclusion of T into \mathbb{P}^{n-1} , i.e.

$$\phi_{A,z}: (\mathbb{C}^*)^d \xrightarrow{\phi'_{A,z}} (\mathbb{C}^*)^n \xrightarrow{\pi} T \hookrightarrow \mathbb{P}^{n-1}.$$

Definition 2.1.1 ([CLS11, Def. 2.1.1]). The *scaled projective toric variety* $X_{A,z}$ is the Zariski closure of the image of $\phi_{A,z}$, in symbols $X_{A,z} = \overline{\text{im}(\phi_{A,z})} \subseteq \mathbb{P}^{n-1}$.

If $z = (1, 1, \dots, 1)$, we write $X_A = X_{A,(1,1,\dots,1)}$ for short. Moreover, we might just drop the adjectives “scaled projective” from now on and simply call $X_{A,z}$ a toric variety. Sometimes it will be convenient to assume that the first row of A is the all-ones vector. Then we can directly define $X_{A,z}$ to be the image closure of $(\mathbb{C}^*)^d \rightarrow \mathbb{P}^{n-1}$, $t \mapsto (z_1 t^{a_1} : \dots : z_n t^{a_n})$. Whenever this is the case, we will simply denote this map by $\phi_{A,z}$. One should however be careful that $X_{A,z}$ is then $(d-1)$ -dimensional, and not d -dimensional as before. It will be apparent from the context which convention we are using.

Example 2.1.2. The scaling $z = (1, 1, 1, 1)$ together with the integer matrix

$$A = \begin{pmatrix} 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{pmatrix}$$

give rise to the map $\phi_{A,z}: (\mathbb{C}^*)^2 \rightarrow \mathbb{P}^3$, $(t_1, t_2) \mapsto (1 : t_1 : t_2 : t_1 t_2)$. The Zariski closure of this map is the variety $X_A = X_{A,z} = \mathcal{V}(x_1 x_4 - x_2 x_3) \subset \mathbb{P}^3$, which can be seen to be the closed image of the Segre embedding $\mathbb{P}^1 \times \mathbb{P}^1 \hookrightarrow \mathbb{P}^3$. \diamond

In general, the map $\phi_{A,z}$ might be many-to-one or even have positive dimensional fibres. However, it is always possible to find a different matrix A' such that $X_{A,z} = X_{A',z}$ and $\phi_{A',z}$ is one-to-one, which is the content of the following statement.

Proposition 2.1.3. *There exists an $A' \in \mathbb{Z}^{d' \times n}$ such that $X_{A,z} = X_{A',z}$ and $\phi_{A',z}$ is one-to-one.*

This result is fairly standard; however, we did not find a proof in the literature, so we provide a short argument here, making use of the Smith normal form.

Proof. Let $S = PAQ$ be the Smith normal form of A . Then S is of the form

$$S = \begin{pmatrix} s_1 & & 0 & \dots & 0 \\ & \ddots & \vdots & \ddots & \vdots \\ & s_{d'} & 0 & \dots & 0 \\ 0 & & 0 & \dots & 0 \\ & \ddots & \vdots & \ddots & \vdots \\ & & 0 & 0 & \dots & 0 \end{pmatrix} \in \mathbb{Z}^{d \times n},$$

where $s_i \neq 0$ for $i = 1, \dots, d'$. For any $x \in \text{im}(\phi_{A,z})$, the fibre $\phi_{A,z}^{-1}(x)$ is isomorphic to

$$\phi_{A,z}^{-1}(x) \cong \left\{ t \in (\mathbb{C}^*)^d : \phi_{A,z}(t) = (1 : 1 : \dots : 1) \right\} \cong \mu_{s_1} \times \dots \times \mu_{s_{d'}} \times (\mathbb{C}^*)^{d-d'}$$

Here, μ_{s_i} denotes the group scheme of s_i^{th} roots of unity. Hence, setting A' to be the matrix consisting of the first d' rows of Q^{-1} , we find that A' has the desired properties. \square

From now on, we assume the map $\phi_{A,z}$ to be one-to-one. If $X_{A,z}$ has dimension d , then a d -dimensional algebraic torus $(\mathbb{C}^*)^d$ is densely embedded into $X_{A,z}$ via the map $\phi_{A,z}$. The torus comes with an algebraic group action by itself which can be extended to $X_{A,z}$.

Proposition 2.1.4 ([Tel22, Prop. 3.4]). *The natural torus action of the dense torus $(\mathbb{C}^*)^d \hookrightarrow X_{A,z}$ given by multiplication on itself extends to $X_{A,z}$.*

Quite often, the statement above is taken as a definition for toric varieties, see e.g. [CLS11, Def. 3.1.1]. This gives rise to a notion of *abstract* toric varieties. The toric varieties from Definition 2.1.1 are embedded projectively. The two notions relate as follows. Let X be an abstract projective normal toric variety. The data of an embedding of X into projective space is equivalent to a very ample divisor D on X . This divisor can be assumed to be torus-invariant. Associated to it is a polytope P_D . Collect the lattice points of P_D in the columns of an integer matrix A . Then $X \cong X_A$. See [CLS11, Ch. 6] for more details. Abstract toric varieties will appear in Subsection 3.1.2 in the context of toric fibre products.

Example 2.1.5. Consider the matrix A from Example 2.1.2. Regarding the columns of A as lattice points in \mathbb{Z}^2 , they form the lattice points of the square shown in Figure 2.1 (left). The normal fan of this polytope is the fan depicted in Figure 2.1 (right), which is the normal fan of $\mathbb{P}^1 \times \mathbb{P}^1$. One can equip each ray with the normalised volume of the corresponding facet of the polytope; in this case, this is one for all rays. This gives a very ample (Weil) divisor on $\mathbb{P}^1 \times \mathbb{P}^1$; its line bundle defines the Segre embedding of $\mathbb{P}^1 \times \mathbb{P}^1 \hookrightarrow \mathbb{P}^3$. \diamond

The orbits of the torus action admit a nice combinatorial description. This is commonly known as the *orbit-cone-correspondence*. Here, this is given by an orbit-face-correspondence.

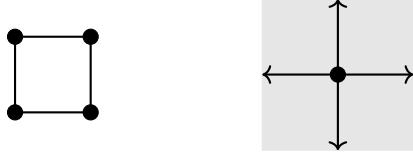


Figure 2.1: The polytope associated to the Segre embedding X_A from Example 2.1.2 (left) and its normal fan, the normal fan of $\mathbb{P}^1 \times \mathbb{P}^1$ (right).

Theorem 2.1.6 (Orbit-face-correspondence, [MS21, Cor. 8.21]). *The torus orbits of $X_{A,z}$ are in inclusion-preserving bijection with the faces of the polytope $P = \text{conv}(A)$. A d -dimensional face $F \subseteq P$ corresponds to the d -dimensional torus orbit $\{x \in X_{A,z} : x_i \neq 0 \Leftrightarrow a_i \in F\}$.*

Example 2.1.7. For X_A as in Examples 2.1.2 and 2.1.5, the four lattice points correspond to the four points $(1 : 0 : 0 : 0), \dots, (0 : 0 : 0 : 1)$. The facets correspond to the one-dimensional orbits $\{(s : t : 0 : 0) \in \mathbb{P}^1 \times \mathbb{P}^1\}, \dots, \{(0 : 0 : s : t) \in \mathbb{P}^1 \times \mathbb{P}^1\}$. The whole square corresponds to the two-dimensional dense torus $\phi_A : (\mathbb{C}^*)^2 \hookrightarrow \mathbb{P}^1 \times \mathbb{P}^1$. \diamond

Also, the degree of $X_{A,z}$ is a combinatorial property of $P = \text{conv}(A)$.

Theorem 2.1.8 (Kouchnirenko's Theorem, [Tel22, Thm. 3.16]). *Let $P = \text{conv}(A) \subset \mathbb{R}^n$, then the degree of $X_{A,z}$ equals the normalised volume $\deg(X_{A,z}) = n! \text{Vol}(P)$.*

Example 2.1.9. The normalised volume of the unit square is two which is the degree of the Segre embedding $\mathbb{P}^1 \times \mathbb{P}^1 \subset \mathbb{P}^3$. \diamond

2.2. Algebraic statistics

In algebraic statistics one uses tools from algebra, combinatorics and geometry to study problems in statistics. In this section we provide basic notions of (algebraic) statistics, with a particular emphasis on maximum likelihood estimation, which is the main problem of interest in Chapter 3. This exposition mostly follows the book [Sul18] and [DHW23].

A discrete statistical model with n outcomes is a subset \mathcal{M} of the open probability simplex

$$\Delta_{n-1}^\circ = \left\{ (p_1, \dots, p_n) : \sum_{i=1}^n p_i = 1, p_i > 0 \forall i = 1, \dots, n \right\}.$$

Each point in Δ_{n-1}° specifies a probability distribution for a random variable X with outcome space $[n] := \{1, \dots, n\}$ by setting $p_i = P(X = i)$. A broad class of statistical models are *log-affine models*, also called *discrete regular exponential families*.

Definition 2.2.1 ([Sul18, Def. 6.2.1]). Let $A \in \mathbb{Z}^{d \times n}$ be an integer matrix with the all-ones vector in its rowspan, and let $z \in \mathbb{R}_{>0}^n$. The *log-affine model* $\mathcal{M}_{A,z}$ is given by

$$\mathcal{M}_{A,z} = \{p \in \Delta_{n-1}^\circ : \log(p) \in \log(z) + \text{rowspan}(A)\}.$$

If $z = (1, 1, \dots, 1)$, the model is called *log-linear* and simply denoted \mathcal{M}_A .

A key observation of algebraic statistics is that many statistical models are *semialgebraic sets*, meaning they can be described by polynomial equations and inequalities. In particular, log-affine models can be represented by toric varieties.

Proposition 2.2.2 ([Sul18, §6.2]). A log-affine model $\mathcal{M}_{A,z}$ can be represented as the intersection of a scaled toric variety with the open probability simplex. More precisely, let $\varphi: \mathbb{P}^{n-1} \rightarrow \mathbb{C}^n$ be the map defined by $[p_1 : \cdots : p_n] \mapsto \frac{1}{p_1 + \cdots + p_n}(p_1, \dots, p_n)$. Then

$$\mathcal{M}_{A,z} = \varphi(X_{A,z}) \cap \Delta_{n-1}^\circ. \quad (2.2.1)$$

Example 2.2.3. The Segre embedding X_A from Example 2.1.2 represents the *independence model* on two binary random variables. Indeed, let X_1 and X_2 be two such random variables and let p_{ij} denote the joint probability $p_{ij} = P(X_1 = i, X_2 = j)$ for $i, j \in \{0, 1\}$. Independence of X_1 and X_2 is equivalent to $p_{00}p_{11} = p_{01}p_{10}$. This is precisely the implicit definition of X_A . Hence, the statistical model \mathcal{M}_A consisting of all joint probability distributions of two binary independent random variables is represented by X_A . \diamond

A key example of log-affine models in statistics are *undirected graphical models* [Lau96], with applications to, among others, phylogenetics, causal inference and medical diagnosis [KF09, MDLW18]. Such models arise from graphs imposing certain *conditional independence* statements on random variables represented by nodes in the graphs.

Let $G = (V, E)$ be a simple undirected graph with labelled vertices $V = \{1, \dots, n\}$ and cliques $\mathcal{C}(G)$. Consider n discrete random variables associated to the vertices of G , where the random variable X_i has state space S_{X_i} . We define a matrix $A(G)$ as follows. The columns of A are indexed by elements of the product state space $\prod_{i=1}^n S_{X_i}$, the rows of A are indexed by pairs (C, s) , where $C \in \mathcal{C}(G)$ is a clique of the graph G and $s \in \prod_{i \in C} S_{X_i}$. Now the entry $(A(G))_{(C,s), s'}$ is set to be 1 if the projection of s' to C equals s and 0 otherwise.

Definition 2.2.4 ([GMS06, §2]). The toric variety associated to the discrete undirected graphical model of G and $\{X_1, \dots, X_n\}$ is the variety $X_{A(G)}$.

Example 2.2.5. Let G be the graph consisting of two nodes and no edges, and let $\{X_1, X_2\}$ be two binary random variables. Then this undirected graphical model is the independence model and its associated toric variety is X_A from Examples 2.1.2 and 2.2.3. \diamond

We stated above that graphical models encode conditional independence (CI) statements and explain this now. Two random variables X and Y are conditionally independent given a third random variable Z if the conditional probability distributions factor as

$$P(X = x, Y = y | Z = z) = P(X = x | Z = z)P(Y = y | Z = z)$$

for all $x \in S_X, y \in S_Y$ and $z \in S_Z$. This property is denoted $X \perp\!\!\!\perp Y | Z$. Conditional independence gives binomial constraints in the indeterminates representing the joint probabilities, see [Sul18, Prop. 4.1.6]. Hence, to each collection of CI statements \mathcal{S} one can associate an ideal $I_{\mathcal{S}}$ generated by all the binomial constraints given by the CI statements. We describe two particularly important collections of CI statements.

Definition 2.2.6 ([Sul18]). Let $G = (V, E)$ be a graph and $\{X_1, \dots, X_n\}$ be a collection of discrete random variables as above.

1. The set of *global Markov statements* $\text{glob}(G)$ consists of all CI statements

$$X_A \perp\!\!\!\perp X_C | X_B,$$

for all disjoint triples of vertices $A, B, C \subset V$ such that B separates A from C , i.e. any path from a vertex in A to a vertex in C must pass through B . Here, X_A is the joint random variable of all X_i with $i \in A$ and similarly for X_B and X_C .

2. The set of *pairwise Markov statements* $\text{pw}(G)$ consists of all CI statements

$$X_i \perp\!\!\!\perp X_j \mid X_{V \setminus \{i,j\}},$$

for any pair of vertices $i, j \in V$ such that $(i, j) \notin E$.

These Markov statements give rise to ideals $I_{\text{glob}(G)}$ and $I_{\text{pw}(G)}$. On the level of varieties, we obtain the inclusions [GMS06, Eq. 4.6]

$$X_{A(G)} \subseteq \mathcal{V}(I_{\text{glob}(G)}) \subseteq \mathcal{V}(I_{\text{pw}(G)}). \quad (2.2.2)$$

In general, these inclusions can be strict. However, the Hammersley–Clifford Theorem asserts that, restricted to positive parts, all the varieties above agree.

Theorem 2.2.7 (Hammersley–Clifford, [GMS06, Thm. 4.1]). *For any graph G , the positive parts in (2.2.2) agree, i.e. we have*

$$X_{A(G)}^{>0} = \mathcal{V}^{>0}(I_{\text{glob}(G)}) = \mathcal{V}^{>0}(I_{\text{pw}(G)}).$$

See the beginning of Section 5.1 for another example of an undirected graphical model.

For an i.i.d. sample $\mathcal{D} = \{X_1, \dots, X_N\}$ of a random variable X , let u_i be the number of appearances of outcome i in \mathcal{D} and set $u = (u_1, \dots, u_n)$. Often we refer to u as *data*. Given data u and a statistical model \mathcal{M} , a common task in statistics is to find the probability distribution $\hat{p} \in \mathcal{M}$ that best explains the data u . To this end, one employs maximum likelihood estimation. The *maximum likelihood estimator* of the model \mathcal{M} is the function $\Phi : \mathbb{N}^n \rightarrow \mathcal{M}$ that assigns to u the point in \mathcal{M} maximising the *log-likelihood function*

$$\ell_u(p) = \left(\sum_{i=1}^n u_i \log(p_i) \right) - u_+ \log(p_+). \quad (2.2.3)$$

Here, we make use of the notation $u_+ = u_1 + \dots + u_n$ and $p_+ = p_1 + \dots + p_n$. For discrete regular exponential families, the log-likelihood function is concave, and under certain genericity conditions on $u \in \mathbb{N}^n$, existence and uniqueness of the *maximum likelihood estimate* (MLE) $\hat{p} = \Phi(u)$ are guaranteed [Hab74]. The maximum likelihood estimator is an algebraic function. However, this does not necessarily imply that the MLE \hat{p} is given in closed form. In practice, one uses e.g. the iterative proportional scaling algorithm [DR72].

In the process of maximising the log-likelihood function one is interested in the number of critical points of (2.2.3). This number is called the *maximum likelihood degree* and can be thought of as an algebraic complexity measure of maximum likelihood estimation.

Definition 2.2.8 ([HS14, Def. 1.5]). Let $X \subseteq \mathbb{P}^{n-1}$ be a projective variety (representing a statistical model) and let \mathcal{H} be the hyperplane arrangement

$$\mathcal{H} = \left\{ (p_1 : p_2 : \dots : p_n) \in \mathbb{P}^{n-1} : p_1 p_2 \dots p_n p_+ = 0 \right\}. \quad (2.2.4)$$

The *maximum likelihood degree* $\text{MLdeg}(X)$ of X is the number of critical points of the log-likelihood function ℓ_u (see (2.2.3)) on $X_{\text{reg}} \setminus \mathcal{H}$ for generic data u .

Here, X_{reg} is the regular locus of X , i.e. the set of nonsingular points. The exclusion of the distinguished hyperplane arrangement is necessary to ensure that (2.2.3) is well-defined. Definition 2.2.8 entails that the number of critical points on $X_{\text{reg}} \setminus \mathcal{H}$ is constant for generic data u . This can be shown as follows, using the likelihood correspondence.

Definition 2.2.9 ([HS14, Def. 1.5]). The *likelihood correspondence* \mathcal{L}_X of a projective variety $X \subseteq \mathbb{P}^{n-1}$ is the Zariski closure in $\mathbb{P}^{n-1} \times \mathbb{P}^{n-1}$ of

$$\{(p, u) : p \in X_{\text{reg}} \setminus \mathcal{H} \text{ and } p \text{ is a critical point of } \ell_u\}.$$

The likelihood correspondence comes with a projection $\pi: \mathcal{L}_X \rightarrow \mathbb{P}_u^{n-1}$ to data space. It can be shown that this map is generically finite-to-one [HS14, Thm. 1.6]. The cardinality of a generic fibre of π is $\text{MLdeg}(X)$, thus establishing well-definedness of Definition 2.2.8.

It should be emphasised that in algebraic statistics models are mostly studied *implicitly*, in contrast to a parametric representation. This has the effect that the maximum likelihood estimate is a probability distribution and not a set of parameters giving this point in the model. In particular, there can be many more solutions to the likelihood equations when they are solved parametrically. This parametric view is taken in Section 3.3. For example, a *parametric* likelihood correspondence is introduced in Definition 3.3.4. See also Remark 3.3.15 for a comparison between the implicit and parametric ML degrees.

Example 2.2.10 (confer [HS14, Ex. 1.14]). We compute the MLE for the independence model X_A from Examples 2.1.2 and 2.2.3 using Lagrange multipliers. Let $u \in \mathbb{N}^4$ be a data point and let p_1, \dots, p_4 be coordinates of \mathbb{P}^3 . The critical points of ℓ_u on X_A satisfy

$$p_1 p_4 - p_2 p_3 = 0 \quad \text{and} \quad \left(\frac{u_i}{p_i} - \frac{u_+}{p_+} \right)_{i=1,\dots,4} = \lambda(p_4, -p_3, -p_2, p_1).$$

The latter condition can equivalently be expressed as

$$\text{rk} \begin{pmatrix} u_1 & u_2 & u_3 & u_4 \\ p_1 & p_2 & p_3 & p_4 \\ p_1 p_4 & -p_2 p_3 & -p_2 p_3 & p_1 p_4 \end{pmatrix} = 2.$$

Let M be the matrix above and let I be the ideal $I = \langle p_1 p_4 - p_2 p_3 \rangle + \langle 3 \times 3 \text{ minors of } M \rangle$. The variety $\mathcal{V}(I)$ encodes all pairs (p, u) such that p is a critical point of ℓ_u ; however, this is not quite the likelihood correspondence. Since X_A is smooth, there are no singular points to exclude. But we need to remove points in the distinguished arrangement \mathcal{H} . To this end we compute the saturation $I : (p_1 \dots p_4 p_+)^{\infty}$. The resulting ideal defines the likelihood correspondence and is minimally generated by $p_1 p_4 - p_2 p_3$ and the entries of

$$U \cdot (p_1, p_2, p_3, p_4)^T, \quad \text{where} \quad U = \begin{pmatrix} 0 & 0 & u_2 + u_4 & -u_1 - u_3 \\ 0 & u_3 + u_4 & 0 & -u_1 - u_2 \\ u_3 + u_4 & 0 & -u_1 + u_4 & -u_1 - u_3 \\ u_2 + u_4 & -u_1 + u_4 & 0 & -u_1 - u_2 \end{pmatrix}.$$

For generic data u , the matrix U has rank three and there is a unique vector \hat{p} in the kernel satisfying $\hat{p}_1 \hat{p}_4 - \hat{p}_2 \hat{p}_3 = 0$. This solution is given by the rational function

$$\hat{p} = u_+^{-2} ((u_1 + u_3)(u_1 + u_2), (u_2 + u_4)(u_1 + u_2), (u_1 + u_3)(u_3 + u_4), (u_2 + u_4)(u_3 + u_4)).$$

Hence, the independence model X_A has ML degree one. \diamond

For log-affine models, the MLE can be computed by intersecting a linear space with the associated toric variety. This is the content of *Birch's Theorem*, illustrated in Figure 2.2.

Theorem 2.2.11 (Birch's Theorem, [DSS08, Prop. 2.1.5]). Let $u \in \mathbb{N}_{>0}^n$ be a positive data vector and let $\mathcal{M}_{A,z}$ be a log-affine model as in (2.2.1). Then the MLE $\hat{p} \in \mathcal{M}_{A,z}$ for data u is the unique nonnegative solution to the system of equations

$$A\hat{p} = Au \quad \text{and} \quad \hat{p} \in X_{A,z}.$$

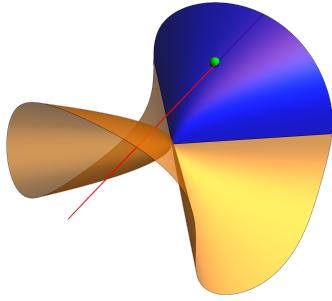


Figure 2.2: An illustration of Birch’s Theorem: the MLE (green) is the unique intersection point of a linear space (red) with the positive part (blue) of a toric variety (blue and yellow).

As a consequence, for *generic* scaling parameters z , the scaled toric variety $X_{A,z}$ has $\text{MLdeg}(X_{A,z}) = \deg(X_{A,z})$, which, by Theorem 2.1.8, is just the normalised volume of the associated polytope $\text{conv}(A)$. However, for non-generic z , the ML degree can drop. We have already seen an instance of this phenomenon: the variety X_A representing the independence model of two binary random variables has degree two (Example 2.1.9), however, the ML degree (for the scaling $z = (1, 1, \dots, 1)$) is equal to one (Example 2.2.10). Determining when and how this happens is precisely the content of Problem 3.0.1.

We now turn to a more geometric formulation of the ML degree in terms of an Euler characteristic of a *very affine variety*. A very affine variety is a variety admitting a closed embedding into an algebraic torus. Let $\varphi: \mathbb{P}^{n-1} \dashrightarrow (\mathbb{C}^*)^n$ be the rational map defined by $[p_1 : \dots : p_n] \mapsto \frac{1}{p_1 + \dots + p_n} (p_1, \dots, p_n)$. Restricted to $X \setminus \mathcal{H}$, where \mathcal{H} is the distinguished hyperplane arrangement (2.2.4), the map φ gives a closed embedding of $X \setminus \mathcal{H}$ into $(\mathbb{C}^*)^n$, turning $X \setminus \mathcal{H}$ into a very affine variety. Then there is the following result.

Theorem 2.2.12 ([Huh13, Thm. 1]). *If $X \setminus \mathcal{H}$ is a smooth very affine variety of dimension d , then*

$$\text{MLdeg}(X) = (-1)^d \chi(X \setminus \mathcal{H}),$$

where χ denotes the topological Euler characteristic.

Example 2.2.13 ([HS14, p. 72]). Again, consider the independence model $X_A \cong \mathbb{P}^1 \times \mathbb{P}^1$ from Examples 2.1.2 and 2.2.3. Let us choose coordinates $((x_1 : x_2), (y_1 : y_2))$ on $\mathbb{P}^1 \times \mathbb{P}^1$. Then the distinguished hyperplane arrangement \mathcal{H} can be written as

$$\mathcal{H} = \mathcal{V}(x_1 x_2 y_1 y_2 (x_1 + x_2)(y_1 + y_2)).$$

Therefore, $X_A \setminus \mathcal{H} = (\mathbb{P}^1 \setminus \mathcal{V}(x_1 x_2 (x_1 + x_2))) \times (\mathbb{P}^1 \setminus \mathcal{V}(y_1 y_2 (y_1 + y_2)))$. Each factor is a \mathbb{P}^1 with three points removed. By the excision property of the Euler characteristic, each factor has Euler characteristic (-1) , therefore $\chi(X_A \setminus \mathcal{H}) = 1$. This confirms that $\text{MLdeg}(X_A) = 1$, as already seen in Example 2.2.10. \diamond

The case where a model has ML degree one is of particular interest: in this case the coordinate functions of Φ are rational functions in the data u , thus the MLE has a closed form expression. This is in fact completely determined in terms of a *Horn matrix* as we explain below, following [Huh14, DMS21]. It is an open problem in algebraic statistics to characterise the class of toric varieties with ML degree one and their respective Horn matrices. Only for toric surfaces such a classification is known [DDPS23].

Definition 2.2.14. A *Horn matrix* is an $r \times n$ integer matrix with all column sums being zero. Given a Horn matrix H with columns h_1, h_2, \dots, h_n and a vector $\lambda \in \mathbb{R}^n$, the *Horn parametrisation* $\varphi_{(H,\lambda)}: \mathbb{R}^r \rightarrow \mathbb{R}^n$ is the rational map defined by

$$u \mapsto \lambda \star (Hu)^H = \left(\lambda_1(Hu)^{h_1}, \lambda_2(Hu)^{h_2}, \dots, \lambda_n(Hu)^{h_n} \right).$$

Here, \star denotes the (entrywise) Hadamard product, and for two vectors $u, v \in \mathbb{R}^r$ we use the shorthand notation $u^v = \prod_{i=1}^r u_i^{v_i}$.

Definition 2.2.15. The pair (H, λ) is called a *Horn pair* if

1. the coordinates of $\varphi_{(H,\lambda)}$ sum up to one, i.e.

$$\lambda_1(Hu)^{h_1} + \lambda_2(Hu)^{h_2} + \dots + \lambda_n(Hu)^{h_n} = 1, \quad \text{and}$$

2. $\varphi_{(H,\lambda)}$ is defined for all positive vectors and maps these to positive vectors.

An example for a Horn pair is given in Example 3.1.7. The relation to maximum likelihood estimation becomes apparent in the following statement.

Theorem 2.2.16 ([DMS21, Thm. 1]). *If \mathcal{M} is a statistical model with ML degree one, there exists a Horn pair (H, λ) such that the maximum likelihood estimator Φ for \mathcal{M} satisfies $\Phi = \varphi_{(H,\lambda)}$.*

2.3. Commutative algebra

In this section we introduce some concepts from commutative algebra that are beyond the standard curriculum as for example in [AM18]. The first such concept is the Rees algebra of a module. We are following [EHU03, SUV03, KKM⁺24a] in this exposition.

Let R be a Noetherian ring. The Rees algebra of an ideal $I \subseteq R$ is a well-known object as it appears as the blowup algebra. More precisely, the Rees algebra of I is the algebra

$$\mathcal{R}(I) = \bigoplus_{n=0}^{\infty} I^n = R[Is] \subseteq R[s].$$

If $Z = \mathcal{V}(I)$ is the variety defined by I , the blowup of $\text{Spec}(R)$ along Z is described by $\mathcal{R}(I)$, i.e. $\text{Bl}_Z \text{Spec}(R) = \text{Proj}(\mathcal{R}(I))$. To define the Rees algebra of a module, one wishes to embed the module into a free module and define the Rees algebra as the image of the symmetric algebra of that module. However, this construction should not depend on the chosen embedding. One way to circumvent this problem is to find a versal embedding into a free module. This approach is equivalent to the following. Let M be a finitely generated R -module with m generators. The *symmetric algebra* of M is a commutative R -algebra with m generators that satisfy the same relations as the generators of M . More precisely, if $M = \text{coker}(A)$ for some matrix $A \in R^{m \times l}$, then

$$\text{Sym}(M) = R[s_1, \dots, s_m] / \langle (s_1, \dots, s_m) A \rangle. \tag{2.3.1}$$

Definition 2.3.1. Let R be a Noetherian domain and let M be a finitely generated R -module. The *Rees algebra* $\mathcal{R}(M)$ of M is the quotient of the symmetric algebra $\text{Sym}(M)$ by its R -torsion submodule, $\mathcal{R}(M) = \text{Sym}(M)/\mathcal{T}$.

The Rees algebra can also be defined over rings that are not domains (see [EHU03]), but we won't need this definition here. If the symmetric algebra $\text{Sym}(M)$ has no R -torsion, the module M is of *linear type* and $\mathcal{R}(M) = \text{Sym}(M)$. The following is derived from [EHU03].

Proposition 2.3.2. *Let R and M be as in Definition 2.3.1. Then $\mathcal{R}(M)$ is a domain.*

Lastly, we will need the following fact from [Eis18, § 2].

Proposition 2.3.3. *Taking the Rees algebra commutes with localisation.*

We now turn to more general results from commutative algebra that we will use in later sections. The following statement offers a useful criterion for minimality of a graded resolution and can be found in [SV86, §0, Lem. 4.4].

Lemma 2.3.4. *Let k be a field and let R be a Noetherian graded k -algebra with maximal ideal $\mathfrak{m} = \bigoplus_{i \geq 1} R_i$. Moreover, let M be a finitely generated graded R -module with a graded free resolution,*

$$\cdots \rightarrow F_i \xrightarrow{\varphi_i} F_{i-1} \xrightarrow{\varphi_{i-1}} \cdots \xrightarrow{\varphi_1} F_0 \xrightarrow{\varphi_0} M \rightarrow 0, \quad (2.3.2)$$

where each F_i is finitely generated. Assume we are given bases of F_i for $i \geq 0$, and matrices representing φ_i for $i \geq 1$. Then (2.3.2) is minimal if and only if all matrix entries lie in \mathfrak{m} .

The following is a classical result and can be found e.g. in [Eis13, Thm. 20.15].

Theorem 2.3.5 (Hilbert–Burch). *Let R be a local ring with an ideal $I \subset R$ such that*

$$R^n \xrightarrow{f} R^m \rightarrow R \rightarrow R/I \rightarrow 0$$

is a free resolution of R/I , then $m = n - 1$ and $I = aJ$ where a is a regular element of R and J is a depth-two ideal generated by all $m \times m$ minors of the matrix representing f .

Finally, we will need the first hypertor spectral sequence, see [Wei94, Appl. 5.7.8].

Theorem 2.3.6 (First hypertor spectral sequence). *Let R be a ring, A_\bullet a complex of R -modules and B an R -module. Then the following spectral sequence converges:*

$$E_{pq}^1 = \text{Tor}_q^R(A_p, B) \Rightarrow \text{Tor}_{p+q}^R(A_\bullet, B).$$

2.3.1. Symmetric tensor decomposition

This exposition on symmetric tensor decomposition appeared as Section 3.1 in [KLW24a].

An order- r tensor is a multidimensional array in $\mathbb{K}^{n_1 \times \dots \times n_r}$, where \mathbb{K} is a field. To us, the case of primary interest is $\mathbb{K} = \mathbb{R}$. A tensor $T = (T_{j_1 \dots j_r}) \in \mathbb{R}^{d_0 \times \dots \times d_0}$ is *symmetric* if $T_{j_1 \dots j_r} = T_{\sigma(j_1) \dots \sigma(j_r)}$ for all permutations $\sigma \in \mathfrak{S}_r$. Given an order- r symmetric tensor $T = (T_{j_1 \dots j_r})$ of format $d_0 \times d_0 \times \dots \times d_0$, one can associate the following homogeneous polynomial of degree r in d_0 variables to the tensor T :

$$F(\mathbf{x}) = \sum_{1 \leq j_1, j_2, \dots, j_r \leq d_0} T_{j_1 j_2 \dots j_r} x_{j_1} x_{j_2} \cdots x_{j_r}. \quad (2.3.3)$$

Monomials generally appear more than once in the sum above.

For $\mathbf{i} = (i_1, \dots, i_{d_0}) \in \mathbb{N}^{d_0}$ with $i_1 + \dots + i_{d_0} = r$, we define the multinomial coefficient

$$\binom{r}{i_1, \dots, i_{d_0}} = \frac{r!}{i_1! \cdots i_{d_0}!}.$$

Using multinomial coefficients, the polynomial (2.3.3) can be rewritten as

$$F(\mathbf{x}) = \sum_{i_1+i_2+\dots+i_{d_0}=r} \binom{r}{i_1, \dots, i_{d_0}} a_{i_1 i_2 \dots i_{d_0}} x_1^{i_1} x_2^{i_2} \cdots x_{d_0}^{i_{d_0}} \quad (2.3.4)$$

such that each monomial appears precisely once in the sum.

To obtain the converse map from homogeneous polynomials to symmetric tensors, we note that there is a bijection between the monomials of degree r in d_0 variables and unique entries of a general order- r symmetric tensor of format $d_0 \times d_0 \times \cdots \times d_0$. The bijection is given by the map $f : \mathbb{N}^{d_0} \rightarrow \mathbb{N}^r, \mathbf{i} \mapsto \mathbf{j}$, where i_k denotes the number of appearances of k in \mathbf{j} and the entries of \mathbf{j} are in ascending order. The homogeneous polynomial (2.3.4) maps to the symmetric tensor T with the entries

$$T_{\mathbf{j}} = a_{f^{-1}(\mathbf{j})},$$

where the entries of \mathbf{j} are in ascending order. The remaining entries of T are obtained via symmetry. For more details on the bijection, we refer the reader to [CM96, CGLM08].

The outer product of r vectors $v_i \in \mathbb{R}^{n_i}$ is an order- r tensor defined as

$$v_1 \otimes \cdots \otimes v_r = (v_{1,i_1} \cdots v_{r,i_r})_{i_1, \dots, i_r=1}^{n_1, \dots, n_r}.$$

Definition 2.3.7. Let $T \in \mathbb{K}^{d_0 \times \cdots \times d_0}$ be a symmetric tensor. The *symmetric rank* of T over a field \mathbb{K} is the smallest $k \in \mathbb{N}$ such that

$$T = \sum_{i=1}^k \lambda_i v_i \otimes \cdots \otimes v_i,$$

where $\lambda_i \in \mathbb{K}$ and $v_i \in \mathbb{K}^{d_0}$ for $i \in [k]$. If $\mathbb{K} = \mathbb{R}$ (respectively $\mathbb{K} = \mathbb{C}$), then we call it the *real* (respectively *complex*) *symmetric tensor rank*.

For symmetric matrices (i.e. order-two tensors), the symmetric rank is equal to the rank. It is currently unknown whether this is also true for higher order tensors, see [Dra24].

Let $I_1 \cup I_2 = [r]$ be a partition of the set $[r]$. Let $D_j = \prod_{i \in I_j} n_i$ for $j = 1, 2$. Every partition $I_1 \cup I_2 = [r]$ induces a *flattening* of a tensor $T \in \mathbb{K}^{n_1 \times \cdots \times n_r}$ to a matrix in $\mathbb{K}^{D_1 \times D_2}$. The minors of flattenings yield relations that hold for low rank tensors. For example, a nonzero tensor has rank one if and only if all 2×2 minors of all its flattenings vanish [Lan11, §3.4].

Example 2.3.8. Fix $d_0 = 2$ and $r = 3$. Then we consider binary cubics or, equivalently, order-three $2 \times 2 \times 2$ symmetric tensors. The polynomial $x_1^3 + 3x_1x_2^2 + 3x_2^3$ corresponds to the symmetric tensor whose flattening induced by the partition $\{1, 2\} \cup \{3\}$ is

$$\begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 3 \end{pmatrix}^T. \quad \diamond$$

2.4. Hypersurface arrangements

In this section we recall some basic notions of hypersurface arrangements, in particular modules of logarithmic derivations and differential forms. Most of these are very well-known for hyperplane arrangements, see e.g. [OT13]. Hypersurface arrangements are less commonly studied, but many concepts generalise straightforwardly. We follow [DS09, DSS⁺13] in this exposition. Some of this is material appeared in [KKM⁺24a, §2, §5].

Let $\{f_1, f_2, \dots, f_m\} \in R = \mathbb{C}[x_1, \dots, x_n]$ a finite set of homogeneous polynomials. The union \mathcal{A} of their vanishing loci, $\mathcal{A} = \bigcup_{i=1}^m \mathcal{V}(f_i) \subseteq \mathbb{P}^{n-1}$ is a *hypersurface arrangement*. Sometimes, by abuse of terminology, we also call the collection of polynomials itself an arrangement of hypersurfaces and write $\mathcal{A} = \{f_1, f_2, \dots, f_m\}$. If all f_i are linear, \mathcal{A} is a hyperplane arrangement. Let $f = \prod_{i=1}^m f_i$ be the defining polynomial of \mathcal{A} . Unlike e.g. [Dup15], we do not require $\mathcal{V}(f_i)$ to be smooth or irreducible. However, we do want that the divisor associated to \mathcal{A} is reduced, so we assume that the polynomial f is squarefree.

For an S -algebra T , let $\text{Der}_S(T)$ be the T -module of S -linear derivations on T , i.e. S -linear maps $d: T \rightarrow T$ satisfying the Leibniz rule $d(ab) = ad(b) + bd(a)$. If $S = \mathbb{C}$ we may omit the subscript and simply write $\text{Der}(T)$.

Definition 2.4.1. The module of *logarithmic \mathcal{A} -derivations* $\text{Der}_S(\mathcal{A})$ is defined by

$$\text{Der}_S(\mathcal{A}) = \{\theta \in \text{Der}_S(T) : \theta(f) \in \langle f \rangle\}. \quad (2.4.1)$$

The notion depends on the S -algebra T which is, however, not apparent in the notation since it will be clear from the context. We simply write $\text{Der}(\mathcal{A})$ if $S = \mathbb{C}$ and $T = R$.

Dual to the module of derivations, we have the T -module of S -valued Kähler differentials $\Omega_{T/S}^1 = \text{Hom}(\text{Der}_S(T), T)$, and the modules of Kähler p -forms $\Omega_{T/S}^p = \Lambda^p \Omega_{T/S}^1$. Let $\Omega_{T/S}^p(\star \mathcal{A}) = \Omega_{T_f/S}^p$ be the T_f -module of S -valued rational differential p -forms with poles along the hypersurface arrangement \mathcal{A} . Here, T_f denotes localisation of T at f .

Definition 2.4.2. The module $\Omega_{T/S}^p(\mathcal{A})$ of *logarithmic p -forms with poles along \mathcal{A}* is given by

$$\Omega_{T/S}^p(\mathcal{A}) = \left\{ \omega \in \Omega_{T/S}^p(\star \mathcal{A}) : f\omega \in \Omega_{T/S}^p \text{ and } f d\omega \in \Omega_{T/S}^{p+1} \right\}.$$

If $S = \mathbb{C}$ and $T = R$ we simply write $\Omega^p(\mathcal{A})$.

Modules of logarithmic differential forms are naturally graded as follows. Let $\omega \in \Omega^p(\mathcal{A})$. By definition, $f\omega \in \Omega^p$ and we can write $f\omega = \sum_I g_I dx_I$. Then we define the degree of ω to be $\deg(\omega) = p + \deg(g_I) - \deg(f)$. We write $\Omega^\bullet(\mathcal{A})_k$ for the k^{th} graded part of $\Omega^\bullet(\mathcal{A})$. In the following we collect some key properties, see [DSS⁺13, §1].

Proposition 2.4.3. 1. *The module $\text{Der}(\mathcal{A})$ always contains the Euler derivation*

$$\theta_E = \sum_{i=1}^n x_i \partial_{x_i}.$$

It induces a splitting $\text{Der}(\mathcal{A}) \cong \text{Der}_0(\mathcal{A}) \oplus R\theta_E$ by writing $\theta = \theta' + \frac{1}{\deg(f)} \frac{\theta(f)}{f} \theta_E$.

2. *There is a perfect pairing*

$$\text{Der}(\mathcal{A}) \times \Omega^1(\mathcal{A}) \rightarrow R, \quad \left(\sum_{i=1}^n \theta_i \partial_{x_i}, \sum_{i=1}^n \eta_i dx_i \right) \mapsto \sum_{i=1}^n \theta_i \eta_i.$$

3. *There is an identification*

$$\text{Der}(\mathcal{A}) \xrightarrow{\sim} \Omega^{n-1}(\mathcal{A}), \quad \theta \mapsto \frac{1}{f} \sum_{i=1}^n (-1)^{i-1} \theta(x_i) dx_1 \wedge \cdots \wedge \widehat{dx_i} \wedge \cdots \wedge dx_n.$$

4. *The natural map $j_p: \Lambda^p \Omega^1(\mathcal{A}) \rightarrow \Omega^p(\mathcal{A})$ is injective.*

5. $\Omega^p(\mathcal{A})$ is reflexive for all p , i.e. $\Omega^p(\mathcal{A})^{\vee\vee} \cong \Omega^p(\mathcal{A})$.

In the hyperplane arrangement literature, there has been a large focus on *free* arrangements. The famous Terao conjecture states that freeness is a combinatorial property, i.e. it only depends on the underlying matroid [Ter80].

Definition 2.4.4. A hypersurface arrangement \mathcal{A} is *free* if $\text{Der}(\mathcal{A})$ is a free R -module.

Another well-studied property of arrangements is tameness which is a weakening of the freeness property. Analogously to the hyperplane case (see e.g. [CDFV11, Def. 2.2]), we define it as follows.

Definition 2.4.5. A hypersurface arrangement \mathcal{A} is *tame* if

$$\text{pd}_R(\Omega^p(\mathcal{A})) \leq p \quad \text{for all } 0 \leq p \leq n.$$

Clearly, every free arrangement is tame. The converse is not true, as can be seen e.g. in Example 3.3.17, for the arrangement of four random ternary quadrics and a line.

Hyperplane arrangements. We now turn to the specific case where all polynomials f_i are linear, so \mathcal{A} is a hyperplane arrangement. There is a useful criterion to check whether the arrangement is free, known as Saito's criterion.

Theorem 2.4.6 (Saito's criterion, [Sai80]). *Let $\theta_1, \dots, \theta_n \in \text{Der}(\mathcal{A})$ be logarithmic derivations and let Θ be the $n \times n$ matrix defined via $\Theta_{ij} = \theta_j(x_i)$. The following are equivalent:*

1. $\det(\Theta) = uf$, where $u \in \mathbb{C}^*$,
2. $\text{Der}(\mathcal{A})$ is free with basis $\{\theta_1, \dots, \theta_n\}$.

Moreover, for hyperplane arrangements there exists the notion of localisation. Let \mathcal{A} be a hyperplane arrangement and let $L(\mathcal{A})$ denote the intersection lattice of the hyperplanes $H_i = \{f_i = 0\}$ for $f_i \in \mathcal{A}$. If $X \in L(\mathcal{A})$ then the *localisation* of \mathcal{A} at X is

$$\mathcal{A}_X = \{f_i \in \mathcal{A} : X \subseteq H_i\}.$$

A specific class of hyperplane arrangements are *graphic arrangements*, which constitute a prominent topic in combinatorics. Let $G = (V, E)$ be a simple undirected graph with vertex set $V = \{1, \dots, n\}$. The graphic arrangement $\mathcal{A}(G)$ consists of the hyperplanes $\{x_i - x_j : \{i, j\} \in E\}$. This arrangement lives in \mathbb{P}^{n-1} , but we can also view it in the space \mathbb{P}^{n-2} obtained by projecting from the point $(1 : 1 : \dots : 1)$ which lies in all hyperplanes.

2.5. Neural networks

In this section we provide the very basic setup for the study of neural networks, following the expository sections in [KLW24a, §2, §6].

A general L -layer (*feedforward*) *neural network* is a composition of L affine-linear maps with coordinatewise nonlinearity in between [Hay98]. More precisely, let F_θ be a feedforward neural network with parameters θ , then it can be written as

$$F_\theta(\mathbf{x}) = f_L \circ \sigma_{L-1} \circ f_{L-1} \circ \dots \circ f_2 \circ \sigma_1 \circ f_1(\mathbf{x}),$$

where

$$f_l(\mathbf{x}) : \mathbb{R}^{d_{l-1}} \rightarrow \mathbb{R}^{d_l}, \quad \mathbf{x} \mapsto W_l \mathbf{x} + \mathbf{b}_l$$

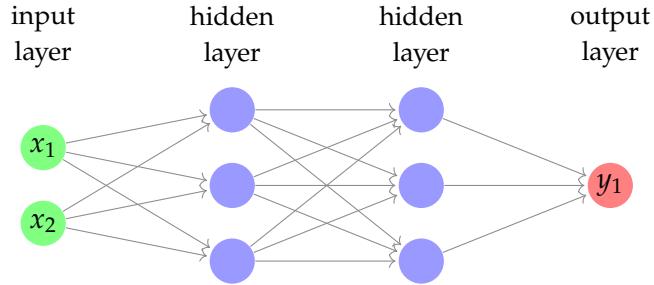


Figure 2.3: A neural network architecture with widths $\mathbf{d} = (2, 3, 3, 1)$, input $\mathbf{x} = (x_1, x_2)^T$ and output y_1 .

and the function $\sigma_l: \mathbb{R}^{d_l} \rightarrow \mathbb{R}^{d_l}$ is the *activation function*, typically a real function that is applied coordinatewise, i.e. $\sigma_l = (\sigma_{l,1}, \dots, \sigma_{l,d_l})$ with $\sigma_{l,j}: \mathbb{R} \rightarrow \mathbb{R}$. The entries of the matrices W_1, \dots, W_L are the *weights* \mathbf{w} of the neural network, and b_1, \dots, b_L are referred to as the *biases*. Together, the weights and biases constitute the parameter set θ .

Pictorially, a neural network architecture can be represented as in Figure 2.3. Each node is called a *neuron* and each column of neurons forms a *layer* of the network. The first layer is called *input layer*, the last layer is called *output layer* and the remaining layers are referred to as *hidden layers*. A neural network with exactly one hidden layer is called *shallow*. The number d_i of neurons in the i^{th} layer is the i^{th} *width*. The vector of widths $\mathbf{d} = (d_0, d_1, \dots, d_L)$ and a choice of activation functions $\sigma = (\sigma_1, \sigma_2, \dots, \sigma_{L-1})$ constitute the *architecture* of the network. Arrows between layers represent the maps f_l . All architectures in this thesis are considered to be fully-connected feedforward neural networks.

The goal of deep learning is to approximate a target function $f: \mathbb{R}^{d_0} \rightarrow \mathbb{R}^{d_L}$ with a neural network F_θ of a chosen architecture $(\mathbf{d} = (d_0, d_1, \dots, d_L), \sigma)$. This amounts to an optimisation task over the space of parameters, the “learning” or “training” process.

The *parameter map*

$$\Psi_{\mathbf{d}, \sigma}: \mathbb{R}^N \rightarrow \text{Fun}(\mathbb{R}^{d_0}, \mathbb{R}^{d_L}), \quad \theta \mapsto F_\theta$$

associates a tuple of parameters θ with the corresponding neural network F_θ . Its image is called the *neuromanifold*¹ $\mathcal{M}_{\mathbf{d}, \sigma}$ and consists of all functions a network with architecture (\mathbf{d}, σ) can learn. The neuromanifold is also referred to as *functional space* in the literature.

Typical choices of activation functions are *ReLU* or *sigmoid* functions

$$\sigma_{\text{ReLU}}(x) = \max\{0, x\}, \quad \sigma_{\text{sigmoid}}(x) = \frac{e^x}{1 + e^x}.$$

In Chapter 4 we study networks where the activation function is given by coordinatewise exponentiation $\sigma(x) = x^r$, leading to *polynomial neural networks*.

In supervised machine learning tasks, a neural network is trained through *empirical risk minimisation (ERM)*. Consider a training dataset $\mathcal{D} = \{(\mathbf{x}_1, \mathbf{y}_1), \dots, (\mathbf{x}_N, \mathbf{y}_N)\}$, where each \mathbf{x}_i is an input vector and \mathbf{y}_i is the corresponding output or label. The objective in ERM is to find a function f^* from a hypothesis space \mathcal{H} that minimises the empirical risk \hat{R} , which is defined as the average loss over the training data

$$\hat{R}(f) = \frac{1}{N} \sum_{i=1}^N L(\mathbf{y}_i, f(\mathbf{x}_i)).$$

¹The term *neuromanifold* is used by an abuse of terminology as this space is rarely a smooth manifold.

Here, L is a loss function measuring the discrepancy between the predicted value $f(\mathbf{x}_i)$ and the actual label \mathbf{y}_i . The function sought after is then

$$f^* = \arg \min_{f \in \mathcal{H}} \hat{R}(f). \quad (2.5.1)$$

To minimise \hat{R} , gradient-based optimisation methods such as (stochastic) gradient descent are commonly used. For convex objective functions, gradient-based algorithms can converge to the unique global minimum with appropriate hyperparameter choices. However, when optimising neural networks, the loss landscape is typically non-convex. As a result, there is generally no theoretical guarantee that gradient-based methods will reach a global optimum f^* for feedforward neural networks.

In the neural network setting, the hypothesis class \mathcal{H} is the corresponding neuromani-fold, and the empirical minimisation problem becomes

$$\boldsymbol{\theta}^* = \arg \min_{\boldsymbol{\theta}} \ell_{F_{\boldsymbol{\theta}}} \quad \text{where } \ell_{F_{\boldsymbol{\theta}}} := \frac{1}{N} \sum_{i=1}^N \ell(\mathbf{y}_i, F_{\boldsymbol{\theta}}(\mathbf{x}_i)) \quad (2.5.2)$$

For example, choosing the ℓ_2 loss, we minimise the average distance between $F_{\boldsymbol{\theta}}(\mathbf{x})$ and \mathbf{y} .

To solve for (2.5.1), one mostly uses iterative optimisation algorithms, including gradient descent and its variants, to find a local minimum of the objective function. In each step, the parameter vector $\boldsymbol{\theta}$ is adjusted in the direction opposite to the gradient of the function at the current point, as the gradient points in the direction of the steepest ascent. The update rule at step t can be expressed as

$$\boldsymbol{\theta}^{(t+1)} = \boldsymbol{\theta}^{(t)} - \eta \nabla_{\boldsymbol{\theta}} \ell_{F_{\boldsymbol{\theta}^{(t)}}}. \quad (2.5.3)$$

The tuning parameter η is referred to as the *learning rate* of the optimisation process. In practice, the value of η is crucial as it determines how big a step is taken on each iteration. If η is too small, the algorithm may take too long to converge; if it is too large, the algorithm may overshoot the minimum or fail to converge.

To compute the gradient of the loss function in (2.5.3) efficiently one uses the *backpropagation algorithm*. Since Werbos introduced this algorithm in his 1974 PhD thesis [Wer74], backpropagation lies at the heart of modern successes in deep learning. We give a brief overview of the backpropagation algorithm following [Nie15, Ch. 2].

Let ℓ be a loss function (e.g. as in (4.3.1)) satisfying the following assumptions:

1. ℓ is smooth;
2. ℓ can be written as $\ell = \sum_{j=1}^N \ell_{\hat{\mathbf{x}}_j}$, where each $\ell_{\hat{\mathbf{x}}_j}$ depends only on the sample $\hat{\mathbf{x}}_j$;
3. ℓ depends only on the output of the network $F_{\boldsymbol{\theta}}(\mathbf{x})$, not on the state of intermediate layers; the training data $f(\hat{\mathbf{x}}_j)$ are considered as parameters.

Backpropagation provides a way to compute the gradient $\nabla_{\mathbf{w}} \ell(\hat{\mathbf{x}}_1, \dots, \hat{\mathbf{x}}_N)$ efficiently. In particular, if $\ell_{\hat{\mathbf{x}}}$ is chosen to be simply the network $F_{\boldsymbol{\theta}}(\hat{\mathbf{x}})$ evaluated at the samples, this gives a way for computing the gradient $\nabla_{\mathbf{w}} F_{\boldsymbol{\theta}}(\hat{\mathbf{x}})$ of the neural network itself. This observation is used in Subsection 4.2.3 to compute the dimension of a neuromani-fold efficiently.

Let z_j^l denote the input into the j^{th} neuron of the l^{th} layer. The *error* δ_j^l of this neuron is

$$\delta_j^l := \frac{\partial \ell}{\partial z_j^l}.$$

Theorem 2.5.1 ([Nie15]). Let a_j^l be the output of the j^{th} neuron in the l^{th} layer, and let L denote the last layer of the network. Then the following equations hold:

$$\delta_j^L = \frac{\partial \ell}{\partial a_j^L} \sigma'_{L,j}(z_j^L), \quad \delta_j^l = \sigma'_{l,j}(z_j^l) \sum_{k=1}^{d_l} w_{l+1,j,k} \delta_k^{l+1}, \quad \frac{\partial \ell}{\partial w_{l,j,k}} = a_k^{l-1} \delta_j^l.$$

Here, $\sigma'_{l,j}$ is the derivative of the single-variable activation function of the j^{th} neuron in the l^{th} layer.

Proof. These are simple consequences of applying the chain rule. \square

These equations give rise to Algorithm 1 below. The idea is to first compute the outputs of each neuron a_j^l via a “forward” run through the network and then compute the errors δ_j^l using the first two equations of Theorem 2.5.1 running “backwards” through the network. Finally, the gradient can be computed using the third equation from Theorem 2.5.1.

Algorithm 1 Backpropagation

Input: A neural network F_θ and a loss function ℓ with single input sample \hat{x} satisfying the assumptions above

Output: The gradient $\nabla_w \ell(\hat{x})$, where w are the weights of F_θ

- ```

1: $a^0 \leftarrow \hat{x}$
2: for l from 1 to L : do ▷ forward loop
3: $z^l \leftarrow W_l a^{l-1}$
4: $a^l \leftarrow \sigma_l(z^l)$
5: end for
6: $\delta^L \leftarrow \nabla_a \ell \star \sigma'_L(z^L)$ ▷ \star denotes the Hadamard (elementwise) product
7: for l from $L - 1$ to 1: do ▷ backward loop
8: $\delta^l \leftarrow (W_l^T \delta^{l+1}) \star \sigma'_l(z^l)$
9: end for
10: return $\nabla_w \ell(x) = (a_k^{l-1} \delta_j^l)_{j,k,l}$

```
- 

## 2.6. Quantum information theory

In this section we collect some fundamental notions of quantum information theory. This exposition is taken from [DPW23a, §2, App. A]. The reader is referred to [Lan19], [NC02, Part III] for more detailed introductions to the subject.

A quantum state on  $N$  qudits is represented by a vector  $|\psi\rangle \in \mathcal{H} = \mathcal{H}_1 \otimes \cdots \otimes \mathcal{H}_N$  of unit length, where  $\mathcal{H}_i$  is the Hilbert space  $\mathcal{H}_i \cong \mathbb{C}^d$ ,  $i = 1, \dots, N$ . Here, we make use of the Dirac notation, i.e.  $|\psi\rangle$  denotes a column vector and  $\langle\psi|$  its complex conjugate transpose. In the case  $N = 1$  and  $d = 2$ ,  $|\psi\rangle$  is called a *qubit* and this will be our primary focus.

An ensemble of quantum states is a collection  $\{p_i, |\psi_i\rangle\}_i$  where  $\{p_i\}_i$  is a discrete probability distribution. Besides the uncertainty inherent to quantum systems through superpositions, ensembles of quantum states take a further statistical uncertainty into account. Such an ensemble is described by its *density matrix*

$$\rho = \sum_i p_i |\psi_i\rangle \langle\psi_i|.$$

Equivalently, we can characterise density matrices as positive semidefinite endomorphisms on  $\mathcal{H}$  with unit trace. From now on, we will use the terms “quantum state” and “density

matrix" interchangeably. If  $\rho$  has rank one, i.e.  $\rho$  is of the form  $\rho = |\psi\rangle\langle\psi|$ , we call  $\rho$  a *pure state*; otherwise, it is *mixed*. The set of all density matrices on  $\mathcal{H}$  is denoted by  $\mathcal{D}(\mathcal{H})$ .

Let  $\rho_{AB}$  be a bipartite quantum state, i.e. a density operator on  $\mathcal{H}_A \otimes \mathcal{H}_B$ . We define the *partial trace* over the  $B$ -system on elementary tensors via

$$\mathrm{Tr}_B(|a_i\rangle\langle a_j| \otimes |b_k\rangle\langle b_l|) := |a_i\rangle\langle a_j| \cdot \mathrm{Tr}(|b_k\rangle\langle b_l|) = |a_i\rangle\langle a_j| \cdot \langle b_l|b_k\rangle,$$

where  $|a_i\rangle, |a_j\rangle \in \mathcal{H}_A$  and  $|b_k\rangle, |b_l\rangle \in \mathcal{H}_B$ , and extend this operation linearly to  $\rho_{AB}$ . Note that  $\mathrm{Tr}_B \rho_{AB}$  is a density operator on  $\mathcal{H}_A$ , and therefore we use the notation  $\rho_A := \mathrm{Tr}_B \rho_{AB}$ . One can think of the partial trace operation as the quantum analogue of marginalisation in statistics. Physically speaking,  $\rho_A$  describes the state of the subsystem  $A$  of the composite system  $AB$ . It can be shown that the partial trace is the unique sensible way to obtain the state of a subsystem, see [NC02, p. 107].

**Example 2.6.1.** Set  $|0\rangle := (1, 0)^T, |1\rangle := (0, 1)^T$  as a basis of  $\mathbb{C}^2$  (this basis is called *computational basis* in the context of quantum information theory); we also adapt the notation to write  $|ij\rangle$  for  $|i\rangle \otimes |j\rangle$ ,  $i, j \in \{0, 1\}$ . Consider the *Bell state*

$$\rho_{AB} = \left( \frac{1}{\sqrt{2}} (|00\rangle + |11\rangle) \right) \left( \frac{1}{\sqrt{2}} (\langle 00| + \langle 11|) \right) = \frac{1}{2} (|00\rangle\langle 00| + |00\rangle\langle 11| + |11\rangle\langle 00| + |11\rangle\langle 11|)$$

on  $\mathbb{C}^2 \otimes \mathbb{C}^2$ . Then the partial trace over  $B$  is computed as

$$\mathrm{Tr}_B \rho_{AB} = \frac{1}{2} (|0\rangle\langle 0| |0\rangle\langle 0| + |0\rangle\langle 1| |1\rangle\langle 0| + |1\rangle\langle 0| |0\rangle\langle 1| + |1\rangle\langle 1| |1\rangle\langle 1|) = \frac{1}{2} (|0\rangle\langle 0| + |1\rangle\langle 1|) = \frac{1}{2} \mathrm{Id}_2.$$

This is the *maximally mixed state*, meaning  $\rho_{AB}$  is *maximally entangled* [NC02, §2]. Notice that we started with a pure state  $\rho_{AB}$  with no statistical uncertainty about the state, and ended up with a mixed state with maximal statistical uncertainty. In other words, without knowing what state system  $B$  is in, we are uncertain whether system  $A$  is in state  $|0\rangle\langle 0|$  or  $|1\rangle\langle 1|$ . The concept of entropy (see Subsection 5.1.1) captures this uncertainty.  $\diamond$

When we speak of a *Hamiltonian*  $H$ , we simply mean a real symmetric matrix. The reason we restrict to this class (and do not consider Hermitian matrices as would be natural in many contexts in quantum physics) is that symmetric matrices form a linear space, while the set of Hermitian matrices is not an algebraic variety. Likewise, in Section 5.1 we consider density matrices to be real positive semidefinite (PSD). Note that  $\exp(H)$  is positive definite and thus can (up to normalisation) be regarded as a quantum state.

We now briefly describe the *stabiliser formalism*. This framework is commonly used in quantum error correction for a very convenient description of quantum code spaces. In this exposition we follow [NC02, §10.5.1].

The three *Pauli matrices* are the Hermitian matrices

$$\sigma_X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_Y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix},$$

called *Pauli-X*, *Pauli-Y* and *Pauli-Z* matrix, respectively. In quantum mechanics, these matrices play a fundamental role as observables. They satisfy the commutation relation

$$[\sigma_j, \sigma_k] = 2i\epsilon_{jkl}\sigma_l, \tag{2.6.1}$$

where  $\epsilon_{jkl}$  is the Levi-Civita symbol, and we denote  $\sigma_1 = \sigma_X, \sigma_2 = \sigma_Y$  and  $\sigma_3 = \sigma_Z$ . The *Pauli group*  $\mathcal{P}_1$  is the group

$$\mathcal{P}_1 := \{\pm \mathrm{Id}_2, \pm i\mathrm{Id}_2, \pm \sigma_X, \pm i\sigma_X, \pm \sigma_Y, \pm i\sigma_Y, \pm \sigma_Z, \pm i\sigma_Z\}.$$

The set of all  $N$ -fold tensor products of elements of  $\mathcal{P}_1$  equipped with multiplication as the group operation forms the Pauli group  $\mathcal{P}_N$  of order  $4^{N+1}$ . Any subgroup  $S \leq \mathcal{P}_N$  of the Pauli group acts on the vector space of  $N$  qubit states by multiplication. The *vector space stabilised by  $S$*  is denoted  $V_S$ , and we call  $S$  the *stabiliser* of  $V_S$ . In quantum error correction,  $V_S$  is the code space. The generators of  $S$  provide an efficient description of  $V_S$ .

Let  $S$  be generated by  $S = \langle p_1, \dots, p_l \rangle$ ; the generators  $p_1, \dots, p_l$  are called *independent* if for all  $i = 1, \dots, l$ :  $\langle p_1, \dots, \hat{p}_i, \dots, p_l \rangle \not\subseteq S$ , where the hat means that the element is omitted. The following standard fact is used in Subsection 5.1.6, and we provide a proof to make the reader more familiar with the concepts.

**Lemma 2.6.2** ([NC02, Prop. 10.5]). *Let  $S = \langle p_1, \dots, p_{N-k} \rangle \leq \mathcal{P}_N$  be a subgroup of the Pauli group generated by  $N - k$  independent and commuting Pauli product matrices such that  $-\text{Id}_{2^N} \notin S$ . Then  $V_S$  has dimension  $2^k$ .*

*Proof.* First note that any Pauli matrix  $\sigma \in \{\sigma_X, \sigma_Y, \sigma_Z\}$  has eigenvalues  $\pm 1$ , and the projector on the  $\pm 1$ -eigenspace of  $\sigma$  is  $\frac{\text{Id}_2 \pm \sigma}{2}$ . For any  $\mathbf{x} = (x_1, \dots, x_{N-k}) \in (\mathbb{Z}/2\mathbb{Z})^{N-k}$ , define

$$P_S^\mathbf{x} := \frac{1}{2^{N-k}} \prod_{j=1}^{N-k} (\text{Id}_{2^N} + (-1)^{x_j} p_j);$$

clearly,  $P_S^0$  is the projector onto  $V_S$ .

**Claim 2.6.3.** For any  $\mathbf{x} \in (\mathbb{Z}/2\mathbb{Z})^{N-k}$ ,  $\dim(\text{im}(P_S^\mathbf{x})) = \dim(\text{im}(P_S^0))$ .

We represent a Pauli product matrix  $p \in \mathcal{P}_N$  as a (row) vector  $v_p \in (\mathbb{Z}/2\mathbb{Z})^{2N}$  via

$$(v_p)_i = \begin{cases} 1 & \text{if } i \leq N \text{ and the } i\text{th tensor factor of } v_p \text{ is either } \sigma_X \text{ or } \sigma_Y, \\ 1 & \text{if } i > N \text{ and the } (i-N)\text{th tensor factor of } v_p \text{ is either } \sigma_Z \text{ or } \sigma_Y, \\ 0 & \text{else.} \end{cases}$$

Then two Pauli product matrices  $p, p' \in \mathcal{P}_N$  commute if and only if  $v_p \Lambda v_{p'}^T = 0$ , where  $\Lambda$  is the  $2N \times 2N$  matrix

$$\Lambda = \begin{pmatrix} 0 & \text{Id}_N \\ \text{Id}_N & 0 \end{pmatrix}.$$

Let  $p_1, \dots, p_{N-k}$  be the independent generators of  $S$ . For any  $i = 1, \dots, N - k$  there exists a  $p \in \mathcal{P}_N$  such that  $pp_i p^\dagger = -p_i$  and  $pp_j p^\dagger = p_j$  for all  $j \neq i$ . Indeed, consider the  $(N - k) \times 2N$  matrix  $P$  with rows  $v_{p_1}, \dots, v_{p_{N-k}}$ ; as the generators are independent, one can check that the rows of  $P$  are linearly independent. Therefore, the linear system  $P \Lambda x = e_i$ , where  $e_i$  is the  $i$ th standard basis vector, has a solution, say  $s \in (\mathbb{Z}/2\mathbb{Z})^{2N}$ . Then we define  $p \in \mathcal{P}_N$  by  $v_p = s^T$ . Thus, for any  $j \neq i$  we have  $v_{p_j} \Lambda v_p = 0$ , so  $p$  and  $p_j$  commute and  $pp_j p^\dagger = p_j$ . Moreover,  $v_{p_i} \Lambda v_p = 1$ , hence  $pp_i p^\dagger = -p_i$ . This shows that for any  $\mathbf{x} \in (\mathbb{Z}/2\mathbb{Z})^{N-k}$ , there exists  $p_\mathbf{x} \in \mathcal{P}_N$  such that  $P_S^\mathbf{x} = p_\mathbf{x} P_S^0 p_\mathbf{x}^\dagger$ , proving the claim.

Let  $\mathbf{x}, \mathbf{x}' \in (\mathbb{Z}/2\mathbb{Z})^{N-k}$  be two distinct vectors, i.e. there exists an  $i \in \{1, \dots, N - k\}$  such that  $\mathbf{x}_i \neq \mathbf{x}'_i$ . Then  $\text{im}(P_S^\mathbf{x})$  and  $\text{im}(P_S^{\mathbf{x}'})$  are orthogonal. Indeed, the Hilbert–Schmidt inner product between  $P_S^\mathbf{x}$  and  $P_S^{\mathbf{x}'}$  evaluates to

$$\langle P_S^\mathbf{x}, P_S^{\mathbf{x}'} \rangle = \frac{1}{2^{2(N-k)}} \text{Tr} \left( (\text{Id} + p_i)(\text{Id} - p_i) \prod_{j \neq i} (\text{Id} + (-1)^{x_j} p_j)(\text{Id} + (-1)^{x'_j} p_j) \right) = 0$$

as  $(\text{Id} + p_i)/2$  and  $(\text{Id} - p_i)/2$  are projectors on complementary eigenspaces of  $p_i$ . Finally,

$$\sum_{\mathbf{x} \in (\mathbb{Z}/2\mathbb{Z})^{N-k}} P_S^{\mathbf{x}} = \text{Id}_{2^N},$$

so the  $2^{N-k}$  many vector spaces  $\text{im}(P_S^{\mathbf{x}})$  form an equidimensional partition of  $\mathbb{C}^{2^N}$ , hence  $\dim(V_S) = \dim(\text{im}(P_S^0)) = 2^k$ .  $\square$

In this chapter we have introduced basic notions from various branches of mathematics and its applications. We hope this equips the reader with sufficient background material to follow the presentation in subsequent chapters. The need for preliminary material is, however, highly subjective, and throughout the text we provide references to more detailed expositions the reader might wish to consult.



## Chapter 3

# Likelihood geometry

In this chapter we study the problem of maximum likelihood estimation from a geometric viewpoint. Throughout this chapter, the guiding problem is the following.

**Problem 3.0.1.** Given a scaled projective toric variety  $X_{A,z}$ , what is  $\text{MLdeg}(X_{A,z})$ ? How does it depend on the scaling  $z$ , and what is the parameter locus  $Z \subseteq (\mathbb{C}^*)^n$  such that for  $z \in Z$ ,  $X_{A,z}$  attains a certain ML degree?

Firstly, in Section 3.1, we focus on the case of ML degree one. The toric fibre product is a combinatorial way to construct higher dimensional toric varieties from lower dimensional ones. We study the relation between toric fibre products and Horn matrices, determining the parametrisation of the MLE. Moreover, this gives rise to applications in geometric modelling, constructing novel blending functions. This section is based on [DHW23].

In Section 3.2, taken from [TW24a], we take a more holistic view on Problem 3.0.1. We introduce the notion of Euler stratifications and develop algorithms to compute these for hypersurface families. In particular, this gives a computational answer to Problem 3.0.1, as the stratification of  $Z$  into parameter loci where  $\text{MLdeg}(X_{A,z})$  is constant can be seen as an Euler stratification. Furthermore, we apply our algorithms to particle physics, where the Euler characteristic of a very affine variety measures the number of master integrals.

The last part, Section 3.3, is based on [KKM<sup>+</sup>24a] and considers statistical models in a parametric description. This is done by connecting likelihood geometry with the theory of hypersurface arrangements. In particular, we describe a new way to compute the ideal of the (parametric) likelihood correspondence via the Rees algebra of the likelihood module. This module is closely connected to the module of logarithmic derivations which is well-studied in the arrangement literature.

## 3.1. Toric fibre products: Horn matrices and geometric modelling

The toric fibre product (TFP), introduced by Sullivant [Sul07], is an operation that takes two toric varieties  $X_1, X_2$  and, using compatibility criteria determined by a multigrading  $\mathcal{A}$ , creates a (usually) higher dimensional toric variety  $X_1 \times_{\mathcal{A}} X_2$ . This operation is used to construct a Markov basis for  $X_1 \times_{\mathcal{A}} X_2$  by using Markov bases of  $X_1$  and  $X_2$ . Interestingly, the ML degree of a TFP is the product of the ML degrees of its factors, therefore the TFP of two models with ML degree one yields a model with ML degree one [AKK20]. The Cartesian product of two statistical models is an instance of a TFP. Another example is the class of decomposable graphical models, each of these models has ML degree one and can be constructed iteratively from lower dimensional ones using TFPs [Sul07, Lau96]. We investigate the relation between the combinatorially constructed TFPs and categorical fibre products of abstract normal toric varieties in Subsection 3.1.2.

For a variety with ML degree one, the Horn matrix describes the rational maximum likelihood estimator, see Definition 2.2.15 and the paragraph thereafter. In the toric case, sometimes geometric information about the associated polytope determines a Horn matrix for the model. Instances of these phenomena are present in the characterisation of polytopes with the more restrictive property of strict linear precision [CC20], and in the classification of two-dimensional toric models with ML degree one [DDPS23]. With this perspective in mind, in Subsection 3.1.3, we present an explicit construction of a Horn matrix for the toric fibre product of two toric varieties with ML degree one. This construction reformulates [AKK20, Thm. 5.5] in terms of Horn matrices.

Geometric modelling is, broadly speaking, the study of how geometric objects can best be represented by a computer. *Blending functions* provide parametrisations employed in CAD software used by architects to model buildings. These blending functions are associated to polytopes and have desirable characteristics if the polytope has the property of *rational linear precision*. It is an open problem to classify polytopes in dimension  $d \geq 3$  having rational linear precision [CC20]. There is a remarkable connection between geometric modelling and algebraic statistics: a polytope has rational linear precision if and only if its corresponding toric variety has ML degree one [GPS10]. Inspired by algebraic statistics, we introduce the toric fibre product construction to geometric modelling. In statistics, the interest is in the closed form expression for the MLE; in geometric modelling, the interest is in finding blending functions defined on the polytope satisfying the property of linear precision. Our main Theorem 3.1.14 gives an explicit formula for the blending functions defined on the toric fibre product of two polytopes that have rational linear precision.

### 3.1.1. Toric fibre products via point configurations

In the paper [Sul07], toric fibre products are introduced algebraically on the level of ideals. Here, we take a different approach and define toric fibre products combinatorially via point configurations. These two approaches yield the same object, see Remark 3.1.4.

Let  $r \in \mathbb{N}$  and  $s_i, t_i \in \mathbb{N}$  for  $1 \leq i \leq r$ . Fix integral point configurations  $\mathcal{A} = \{\mathbf{a}^i : i \in [r]\} \subseteq \mathbb{Z}^d$ ,  $B = \{\mathbf{b}_j^i : i \in [r], j \in [s_i]\} \subseteq \mathbb{Z}^{d_1}$  and  $C = \{\mathbf{c}_k^i : i \in [r], k \in [t_i]\} \subseteq \mathbb{Z}^{d_2}$ . For any point configuration  $P$ , we use  $P$  interchangeably to denote a set of points or the matrix whose columns are the points in  $P$ ; the symbol  $|P|$  will be used to denote the indexing set of  $P$ . For each  $i \in |\mathcal{A}|$ , set  $B^i = \{\mathbf{b}_j^i : j \in [s_i]\}$  and  $C^i = \{\mathbf{c}_k^i : k \in [t_i]\}$ . The indices  $i, j, k$  are reserved for elements in  $|\mathcal{A}|, |B^i|$  and  $|C^i|$ , respectively.

**Definition 3.1.1.** A point configuration  $B \subset \mathbb{Z}^{d_1}$  is *multigraded* by  $\mathcal{A} \subset \mathbb{Z}^d$  if there exists a linear map  $\deg: \mathbb{Z}^{d_1} \rightarrow \mathbb{Z}^d$  such that  $\deg(\mathbf{b}_j^i) = \mathbf{a}^i$  for all  $i \in |\mathcal{A}|$  and  $j \in |B^i|$ .

Throughout, we assume linear independence of  $\mathcal{A}$  and the existence of an  $\bar{\omega} \in \mathbb{Q}^d$  such that  $\bar{\omega}\mathbf{a}^i = 1$  for all  $i$ ; the latter condition ensures that if a toric ideal is homogeneous with respect to a multigrading  $\mathcal{A}$  then it is also homogeneous in the usual sense. Linear independence of  $\mathcal{A}$  implies that only the cardinality of  $\mathcal{A}$  is relevant. Moreover, we always assume that a row of all-ones is the first row of the matrices  $B$  and  $C$  so to meet the assumptions of Section 2.1. This row of all-ones is omitted from notation.

**Definition 3.1.2.** The *toric fibre product* (TFP) of two integral point configurations  $B$  and  $C$  multigraded by  $\mathcal{A}$  is the point configuration  $B \times_{\mathcal{A}} C$  given by

$$B \times_{\mathcal{A}} C = \{(\mathbf{b}_j^i, \mathbf{c}_k^i) : i \in |\mathcal{A}|, j \in |B^i|, k \in |C^i|\}.$$

The definition above gives rise to a definition of toric fibre products of toric varieties as follows. Let  $z$  and  $\tilde{z}$  be vectors of weights for  $B$  and  $C$ , respectively, i.e.  $z \in (\mathbb{C}^*)^{|B|}$  and  $\tilde{z} \in (\mathbb{C}^*)^{|C|}$ . We set the vector of *weights* for  $B \times_{\mathcal{A}} C$  to be

$$z_{B \times_{\mathcal{A}} C} := (z_j^i \tilde{z}_k^i)_{(j,k) \in |B^i \times C^i|}^{i \in |\mathcal{A}|}.$$

**Definition 3.1.3.** Let  $B$  and  $C$  be point configurations multigraded by  $\mathcal{A}$ , and let  $X_{B,z}$  and  $X_{C,\tilde{z}}$  be scaled projective toric varieties (see Definition 2.1.1). The *toric fibre product* between  $X_{B,z}$  and  $X_{C,\tilde{z}}$  is the toric variety associated to  $B \times_{\mathcal{A}} C$  and  $z_{B \times_{\mathcal{A}} C}$ , denoted by  $X_{B \times_{\mathcal{A}} C, z_{B \times_{\mathcal{A}} C}}$ .

**Remark 3.1.4.** Sullivant [Sul07] introduces the TFP as an operation on toric ideals which are multigraded by  $\mathcal{A}$ ; such condition, as explained in [EKS14], is equivalent to the existence of linear maps  $\pi_1: \mathbb{Z}^{d_1} \rightarrow \mathbb{Z}^r$  and  $\pi_2: \mathbb{Z}^{d_2} \rightarrow \mathbb{Z}^r$  with  $\pi_1(\mathbf{b}_j^i) = \mathbf{a}^i$  for all  $i$  and  $j$ , and  $\pi_2(\mathbf{c}_k^i) = \mathbf{a}^i$  for all  $i$  and  $k$ . Therefore, our Definition 3.1.3 agrees with the original one.

We use  $\deg$  to denote the projections  $\pi_1, \pi_2$ . The following example illustrates the toric fibre product of two point configurations.

**Example 3.1.5.** Consider the two point configurations

$$B = \{(0,0), (1,0), (0,1), (1,1)\}, \quad C = \{(0,0), (1,0), (2,0), (1,1), (0,1)\}.$$

Recall that a 1 in the first coordinate of each point is omitted. To check the linearity of the degree map it is important to take the leading 1 into account. Let  $\mathcal{A} = \{e_1, e_2\} \subset \mathbb{Z}^2$  consist of two basis vectors. We define the multigrading via the two linear maps

$$\pi_1 = \pi_2: \mathbb{Z}^3 \rightarrow \mathbb{Z}^2, \quad (1,0,0) \mapsto \mathbf{a}^1 = e_1, \quad (1,1,0) \mapsto \mathbf{a}^1 = e_1, \quad (1,0,1) \mapsto \mathbf{a}^2 = e_2.$$

The corresponding toric fibre product  $B \times_{\mathcal{A}} C$  is displayed in Figure 3.1. ◊

### 3.1.2. Fibre products in the category of toric varieties

The naming of the construction described above as toric fibre product is derived from the idea that a  $(\dim(\mathcal{A}) - 1)$  dimensional torus is quotiented out from the product of the two varieties, see the representation of a toric fibre product as a GIT quotient in [EKS14, Prop. 2.2]. However, it is natural to wonder whether the construction is also a fibre product in the categorical sense. It turns out that this is not necessarily the case.

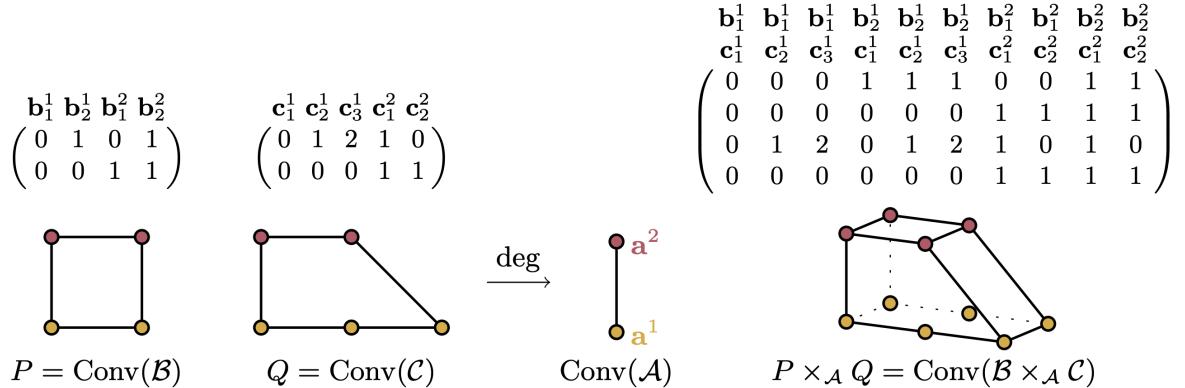


Figure 3.1: Toric fibre product of the point configurations  $B$  and  $C$  in Example 3.1.5. Each point configuration is displayed as a matrix with its corresponding convex hull below. The yellow vertices in each polytope have degree  $e_1$  while the red vertices in each polytope have degree  $e_2$  in the associated multigrading  $\mathcal{A}$ . The degree map is  $\deg(\mathbf{b}_j^i) = \deg(\mathbf{c}_k^i) = \mathbf{a}^i$ .

Let **NoToVar** denote the category of (abstract) normal toric varieties. Its objects are represented by tuples  $(\Sigma, N)$ , where  $\Sigma$  is a fan in the ambient lattice  $N$ . The reader is referred to [CLS11, Ch. 3] for details. A toric morphism  $\phi$  between  $(\Sigma_1, N_1)$  and  $(\Sigma_2, N_2)$  is a  $\mathbb{Z}$ -linear map  $\phi: N_1 \rightarrow N_2$  with the property that for every cone  $\sigma_1 \in \Sigma_1$  there exists a cone  $\sigma_2 \in \Sigma_2$  such that  $\phi_{|\mathbb{R}}(\sigma_1) \subseteq \sigma_2$ , where  $\phi_{|\mathbb{R}}$  is the induced linear map on  $N_1 \otimes \mathbb{R}$ .

The category **NoToVar** possesses fibre products via the following combinatorial construction from [Mol21, §2.2]. Consider three toric varieties with toric morphisms (omitting the  $\mathbb{R}$  from the notation of the maps) in the diagram

$$\begin{array}{ccc} & & (\Sigma_2, N_2) \\ & & \downarrow \phi_2 \\ (\Sigma_1, N_1) & \xrightarrow{\phi_1} & (\Sigma_3, N_3) \end{array}$$

Then their fibre product is the toric variety with fan

$$\Sigma_1 \times_{\Sigma_3} \Sigma_2 = \{\sigma_1 \times_{\sigma_3} \sigma_2 : \sigma_i \in \Sigma_i, \phi_1(\sigma_1) \subseteq \sigma_3, \phi_2(\sigma_2) \subseteq \sigma_3\}$$

inside the ambient lattice  $N_1 \times_{N_3} N_2$ . Here,  $\sigma_1 \times_{\sigma_3} \sigma_2$  is the set-theoretic fibre product, i.e.

$$\sigma_1 \times_{\sigma_3} \sigma_2 = \{(s_1, s_2) \in \sigma_1 \times \sigma_2 : \phi_1(s_1) = \phi_2(s_2)\},$$

and similarly for the fibre product of lattices. It should be noted that this fibre product in **NoToVar** does not necessarily agree with the scheme-theoretic fibre product of the same varieties, see e.g. [Mol21, Ex. 2.2.2 & Ex. 2.2.3].

The toric varieties we have seen earlier (and everywhere else in this thesis) are projectively embedded normal toric varieties represented by their associated polytopes. However, we can forget about the concrete embedding and consider them as abstract normal toric varieties by taking the normal fans to their polytopes. The question is whether the toric fibre product is then also a fibre product in **NoToVar** of the corresponding abstract varieties. This is not necessarily the case as is shown by the following statement.

**Theorem 3.1.6.** Let  $B$  and  $C$  be point configurations representing normal toric varieties  $X_B$  and  $X_C$  and abstract normal toric varieties  $(\Sigma_1, N_1)$  and  $(\Sigma_2, N_2)$ . Let  $\mathcal{A}$  be a multigrading of  $B$  and

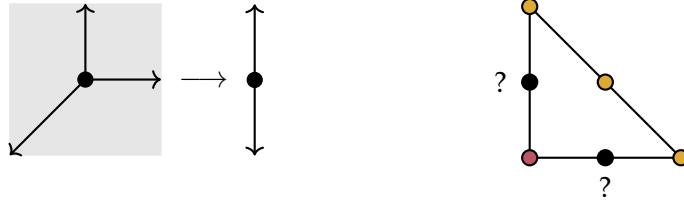


Figure 3.2: Left: a toric morphism  $\mathbb{P}^2 \rightarrow \mathbb{P}^1$  needs to be constant. Right: there is no nontrivial multigrading of the Veronese surface.

C. Then the abstract toric variety underlying the toric fibre product  $X_{B \times_A C}$  is not necessarily a fibre product of  $(\Sigma_1, N_1)$  and  $(\Sigma_2, N_2)$  in **NoToVar**.

*Proof.* Let  $B = C = \{(0,0), (1,0), (0,1)\}$  be two point configurations representing the projective plane  $\mathbb{P}^2$ . We equip both with the same multigrading

$$\deg((0,0)) = \mathbf{a}^1, \deg((1,0)) = \deg((0,1)) = \mathbf{a}^2,$$

where  $\mathcal{A} = \{\mathbf{a}^1, \mathbf{a}^2\}$  is linearly independent (recall that we omit ones in the first coordinate of the points). The toric fibre product  $X_{B \times_A C}$  is of dimension three and represented by a pyramid, the tip corresponding to the product of the origins. Let  $\Sigma$  be the normal fan of  $\text{conv}(B)$  and let  $\Sigma_{B \times_A C}$  be the normal fan of  $X_{B \times_A C}$ . We show the impossibility of

$$\begin{array}{ccc} (\Sigma_{B \times_A C}, \mathbb{Z}^4) & \longrightarrow & (\Sigma, \mathbb{Z}^2) \\ \downarrow & & \downarrow \phi_2 \\ (\Sigma, \mathbb{Z}^2) & \xrightarrow{\phi_1} & (\Xi, N) \end{array}$$

being a Cartesian diagram, for any toric variety  $(\Xi, N)$  and toric morphisms  $\phi_1$  and  $\phi_2$ .

If  $\Xi$  were zero- or one-dimensional, the only toric map from  $\Sigma$  to  $\Xi$  would be constant (the reader can easily convince themselves of this by looking at the left part of Figure 3.2) and  $\Sigma \times_{\Xi} \Sigma$  would be four-dimensional. Hence, assume that  $\Xi$  is at least two-dimensional and  $\phi_1$  is injective, mapping  $\Sigma$  into a two-dimensional slice  $S_1$  of  $\Xi$ . First assume that  $\phi_2$  is constant. Then  $\Sigma \times_{\Xi} \Sigma \cong \Sigma$  is two-dimensional, again a contradiction. Therefore, we may assume that  $\phi_2$  is injective as well, mapping  $\Sigma$  into a two-dimensional slice  $S_2$  of  $\Xi$ .

First consider the case that  $S_1$  and  $S_2$  intersect in a one-dimensional space  $L$ . This  $L$  intersects both  $\phi_1(\Sigma)$  and  $\phi_2(\Sigma)$  in two of their cones. Therefore,  $\Sigma \times_{\Xi} \Sigma$  has two cones and not the five cones of  $\Sigma_{B \times_A C}$ . Finally, if  $S_1 = S_2$ , the fibre product  $\Sigma \times_{\Xi} \Sigma$  would be two-dimensional. In no case we can recover  $\Sigma_{B \times_A C}$ .  $\square$

The example above bears another curiosity. If the toric fibre product were compatible with the category-theoretic fibre product, after passing to the abstract underlying variety it should be agnostic to the specific embedding chosen before. However, consider a dilated two-dimensional simplex  $\Delta$ , still representing  $\mathbb{P}^2$ , but in a Veronese embedding. Corollary 3.1.18 below asserts that the points in  $\Delta$  which are assigned the same multidegree form a face of  $\Delta$ . Therefore, we see that  $\Delta$  can only be trivially multigraded with all points being graded by the same multidegree (see Figure 3.2 (right)). Then the toric fibre product  $X_{\Delta} \times_A X_{\Delta}$  is isomorphic to  $\mathbb{P}^2 \times \mathbb{P}^2$  which is not the variety  $X_{B \times_A C}$  from the proof above.

### 3.1.3. The Horn matrix of ML degree one toric fibre products

In this subsection we give an explicit description of a Horn pair for the toric fibre product of two toric varieties with ML degree one. This construction uses a Horn pair for each factor and for the  $(\mathcal{A} - 1)$ -dimensional probability simplex. Throughout this section we use notation and setup for the toric fibre product introduced in Subsection 3.1.1. Recall the definitions of a Horn matrix (Definition 2.2.14) and a Horn pair (Definition 2.2.15).

If  $X_B$  and  $X_C$  have ML degree one, by Theorem 2.2.16, there exist Horn pairs  $(H_B, \lambda_B)$  and  $(H_C, \lambda_C)$  such that the maximum likelihood estimates  $\hat{p}_B$  and  $\hat{p}_C$  can be expressed in terms of Horn parametrisations, i.e.

$$\hat{p}_B = \lambda_B \star (H_B u_B)^{H_B} \quad \text{and} \quad \hat{p}_C = \lambda_C \star (H_C u_C)^{H_C}$$

for data vectors  $u_B$  and  $u_C$ . It follows from [AKK20, Thm. 5.5] that the toric fibre product of the two models  $X_{B \times_{\mathcal{A}} C}$  has again ML degree one and must therefore admit a Horn pair  $(H_{B \times_{\mathcal{A}} C}, \lambda_{B \times_{\mathcal{A}} C})$ . We give an explicit description of  $(H_{B \times_{\mathcal{A}} C}, \lambda_{B \times_{\mathcal{A}} C})$  in Theorem 3.1.8.

To set up the notation, let

$$u = \left( u_{j,k}^i \right)_{(j,k) \in |B^i \times C^i|}^{i \in |\mathcal{A}|}$$

denote a data vector. As before, we reserve  $i, j$  and  $k$  for indices of  $\mathcal{A}, B^i$  and  $C^i$ , respectively. We use “+” to denote summation over all possible values of the respective index, e.g.  $u_{j,+}^i = \sum_{k \in |C^i|} u_{j,k}^i = (u_B)_j^i$ . In a similar vein, we denote by

$$p = \left( p_{j,k}^i \right)_{(j,k) \in |B^i \times C^i|}^{i \in |\mathcal{A}|}$$

a joint probability distribution for the model  $X_{B \times_{\mathcal{A}} C}$ .

In general, if a statistical model possesses a Horn pair, i.e. the Horn parametrisation yields a parametrisation of the model, the Horn pair is not unique. However, there exists a minimal Horn matrix to a model with ML degree one, see [DMS21].

**Example 3.1.7.** A Horn pair corresponding to the simplex  $\Delta_n$  is given by letting the Horn matrix be the  $(n + 1)$ -dimensional identity matrix with an additional row of  $(-1)$ s at the bottom and with  $\lambda$  being the vector of all  $(-1)$ s. For example, for the simplex  $\Delta_1$  we have

$$H = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ -1 & -1 \end{pmatrix}, \quad \lambda = (-1, -1), \quad \Phi(u_1, u_2) = \lambda \star (Hu)^H = \left( \frac{u_1}{u_1 + u_2}, \frac{u_2}{u_1 + u_2} \right).$$

For another illustration, consider the two models  $X_B$  and  $X_C$  defined by the polytopes  $P = \text{conv}(B)$  and  $Q = \text{conv}(C)$  from Example 3.1.13. We remark that  $X_B$  represents the independence model of two binary random variables (see Example 2.2.3), and  $X_C$  represents a multinomial staged tree (see [DMS21, §3]). For toric surfaces with ML degree one, the Horn pair can be directly read off from the lattice distance functions (see Example 3.1.13) and the normal fan of the polytope, see [DDPS23, Prop. 3.1]. Concretely, we have

$$H_B = \begin{pmatrix} \mathbf{b}_1^1 & \mathbf{b}_2^1 & \mathbf{b}_1^2 & \mathbf{b}_2^2 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ -1 & -1 & -1 & -1 \\ -1 & -1 & -1 & -1 \end{pmatrix}, \quad H_C = \begin{pmatrix} \mathbf{c}_1^1 & \mathbf{c}_2^1 & \mathbf{c}_3^1 & \mathbf{c}_1^2 & \mathbf{c}_2^2 \\ 0 & 1 & 2 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 2 & 1 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 & 0 \\ -1 & -1 & -1 & -1 & -1 \\ -2 & -2 & -2 & -1 & -1 \end{pmatrix}$$

and  $\lambda_B = (1, 1, 1, 1)$ ,  $\lambda_C = (-1, -2, -1, 1, 1)$ ; the columns of the Horn matrices are labelled by the vectors of  $B$  and  $C$ , respectively.  $\diamond$

**Theorem 3.1.8.** *Let  $X_B$  and  $X_C$  be toric varieties with ML degree one and corresponding Horn pairs  $(H_B, \lambda_B)$  and  $(H_C, \lambda_C)$ , respectively, where  $H_B \in \mathbb{Z}^{r_1 \times |B|}$ ,  $H_C \in \mathbb{Z}^{r_2 \times |C|}$ . Fix  $H_{\mathcal{A}}$  to be the minimal Horn matrix associated to the  $(|\mathcal{A}| - 1)$ -dimensional probability simplex, so  $H_{\mathcal{A}} \in \mathbb{Z}^{(|\mathcal{A}| + 1) \times |\mathcal{A}|}$ . Denote the columns of  $H_B$ ,  $H_C$ , and  $H_{\mathcal{A}}$  by  $h_j^i$ ,  $h_k^i$ , and  $h^i$ , respectively. Then  $(H_{B \times \mathcal{A} C}, \lambda_{B \times \mathcal{A} C})$  is a Horn pair for the toric fibre product  $X_{B \times \mathcal{A} C}$ . Here, the vector  $\lambda_{B \times \mathcal{A} C}$  of coefficients is given by*

$$\lambda_{B \times \mathcal{A} C} = \left( \lambda_{j,k}^i \right)_{(j,k) \in |B^i \times C^i|}^{i \in |\mathcal{A}|} \text{ with } \lambda_{j,k}^i = -\lambda_j^i \lambda_k^i \text{ and } \lambda_B = (\lambda_j^i)_{j \in |B^i|}^{i \in |\mathcal{A}|}, \lambda_C = (\lambda_k^i)_{k \in |C^i|}^{i \in |\mathcal{A}|},$$

and the Horn matrix  $H_{B \times \mathcal{A} C}$  is given in block form by

$$H_{B \times \mathcal{A} C} = (H_{B^1 \times C^1} \mid H_{B^2 \times C^2} \mid \cdots \mid H_{B^{|\mathcal{A}|} \times C^{|\mathcal{A}|}}). \quad (3.1.1)$$

For each  $i \in |\mathcal{A}|$ , the column  $h_{j,k}^i$  of block  $H_{B^i \times C^i}$  is the vertical concatenation of  $h_j^i$ ,  $h_k^i$  and  $-h^i$ . Explicitly, if  $\rho = r_1 + r_2 + |\mathcal{A}| + 1$  and  $\alpha \in [\rho]$ , then the row  $\alpha$  of  $h_{j,k}^i$ , denoted by  $h_{j,k}^{\alpha,i}$ , is given by

$$h_{j,k}^{\alpha,i} = \begin{cases} h_j^{\alpha,i} & \text{for } 1 \leq \alpha \leq r_1 \\ h_k^{(\alpha-r_1),i} & \text{for } r_1 + 1 \leq \alpha \leq r_1 + r_2 \\ -h^{(\alpha-r_1-r_2),i} & \text{for } r_1 + r_2 + 1 \leq \alpha \leq \rho. \end{cases}$$

Where,  $h_j^{\alpha,i}$ ,  $h_k^{(\alpha-r_1),i}$ , and  $h^{(\alpha-r_1-r_2),i}$ , are the entries  $\alpha$ ,  $\alpha - r_1$ , and  $\alpha - r_1 - r_2$  of the columns  $h_j^i$ ,  $h_k^i$  and  $h^i$ , respectively.

*Proof.* It suffices to check that the pair  $(H_{B \times \mathcal{A} C}, \lambda_{B \times \mathcal{A} C})$  gives rise to a Horn parametrisation yielding the correct expression for the maximum likelihood estimate of  $X_{B \times \mathcal{A} C}$ ; then the pair will automatically be friendly and positive, thus a Horn pair for  $X_{B \times \mathcal{A} C}$ , see [DMS21].

By [AKK20, Thm. 5.5], the MLE of  $X_{B \times \mathcal{A} C}$  is given by

$$\hat{p} = \left( \hat{p}_{j,k}^i \right)_{(j,k) \in |B^i \times C^i|}^{i \in |\mathcal{A}|} \text{ with } \hat{p}_{j,k}^i = \frac{\hat{p}_j^i \hat{p}_k^i}{\hat{p}_{+,+}^i}.$$

The  $(i, j, k)$ th entry of the Horn parametrisation computes as

$$\left( \lambda_{B \times \mathcal{A} C} \star (H_{B \times \mathcal{A} C} u)^{H_{B \times \mathcal{A} C}} \right)_{j,k}^i = \lambda_{j,k}^i \prod_{\alpha=1}^{\rho} \left( \sum_{(\tilde{i}, \tilde{j}, \tilde{k}) \in |B \times \mathcal{A} C|} h_{\tilde{j}, \tilde{k}}^{\alpha, \tilde{i}} u_{\tilde{j}, \tilde{k}}^{\tilde{i}} \right)^{h_{j,k}^{\alpha,i}} \quad (3.1.2)$$

Let us split the product above into three products  $P_1$ ,  $P_2$  and  $P_3$  where  $\alpha$  ranges over  $\{1, \dots, r_1\}$ ,  $\{r_1 + 1, \dots, r_1 + r_2\}$  and  $\{r_1 + r_2 + 1, \dots, \rho\}$ , respectively. Then we obtain

$$\begin{aligned} P_1 &= \prod_{\alpha=1}^{r_1} \left( \sum_{(\tilde{i}, \tilde{j}, \tilde{k}) \in |B \times \mathcal{A} C|} h_{\tilde{j}, \tilde{k}}^{\alpha, \tilde{i}} u_{\tilde{j}, \tilde{k}}^{\tilde{i}} \right)^{h_{B \times \mathcal{A} C, j,k}^{\alpha,i}} = \prod_{\alpha=1}^{r_1} \left( \sum_{(\tilde{i}, \tilde{j}, \tilde{k}) \in |B \times \mathcal{A} C|} h_{\tilde{j}}^{\alpha, \tilde{i}} u_{\tilde{j}, \tilde{k}}^{\tilde{i}} \right)^{h_j^{\alpha,i}} \\ &= \prod_{\alpha=1}^{r_1} \left( \sum_{(\tilde{i}, \tilde{j}) \in |B|} h_{\tilde{j}}^{\alpha, \tilde{i}} u_{\tilde{j}, +}^{\tilde{i}} \right)^{h_{B,j}^{\alpha,i}} = \frac{\hat{p}_{j,+}^i}{\lambda_j^i}, \end{aligned}$$

and similarly  $P_2 = \frac{\hat{p}_{+,k}^i}{\lambda_k^i}$ . Finally, we have

$$\begin{aligned} P_3 &= \prod_{\alpha=r_1+r_2+1}^{\rho} \left( \sum_{(\tilde{i}, \tilde{j}, \tilde{k}) \in |B \times_{\mathcal{A}} C|} h_{\tilde{j}, \tilde{k}}^{\alpha, i} u_{\tilde{j}, \tilde{k}}^{\tilde{i}} \right)^{h_{j,k}^{\alpha,i}} = \prod_{\alpha=r_1+r_2+1}^{\rho} \left( \sum_{\tilde{i} \in |\mathcal{A}|} -h^{(\alpha-r_1-r_2), \tilde{j}} u_{+,+}^{\tilde{i}} \right)^{(-h^{\alpha,i})} \\ &= \left( \prod_{\alpha=1}^{|\mathcal{A}|} (-u_{+,+}^{\alpha})^{\delta_{\alpha,i}} \right)^{-1} \cdot u_{+,+}^+ = -\frac{u_{+,+}^+}{u_{+,+}^i}. \end{aligned}$$

As  $\mathcal{A}$  is linearly independent,  $\hat{p}_{+,+}^i = \frac{u_{+,+}^i}{u_{+,+}^+}$ . Combining this with the above, we obtain

$$(3.1.2) = -\lambda_j^i \lambda_k^i P_1 P_2 P_3 = \frac{\hat{p}_{j,+}^i \hat{p}_{+,k}^i}{\hat{p}_{+,+}^i}. \quad \square$$

**Example 3.1.9.** The Horn pair for the toric fibre product  $X_B \times_{\mathcal{A}} X_C$  from Theorem 3.1.8, where  $X_B$ ,  $X_C$  and the multigrading are taken from Example 3.1.5 and the Horn matrices are computed in Example 3.1.7, is given by

$$H_{B \times_{\mathcal{A}} C} = \begin{pmatrix} \mathbf{b}_1^1 & \mathbf{b}_1^1 & \mathbf{b}_1^1 & \mathbf{b}_2^1 & \mathbf{b}_2^1 & \mathbf{b}_2^1 & \mathbf{b}_1^2 & \mathbf{b}_1^2 & \mathbf{b}_2^2 & \mathbf{b}_2^2 \\ \mathbf{c}_1^1 & \mathbf{c}_2^1 & \mathbf{c}_3^1 & \mathbf{c}_1^1 & \mathbf{c}_2^1 & \mathbf{c}_3^1 & \mathbf{c}_1^2 & \mathbf{c}_2^2 & \mathbf{c}_1^2 & \mathbf{c}_2^2 \\ 1 & 1 & 1 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 \\ -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 \\ 0 & 1 & 2 & 0 & 1 & 2 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 2 & 1 & 0 & 2 & 1 & 0 & 0 & 1 & 0 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 \\ -2 & -2 & -2 & -2 & -2 & -2 & -1 & -1 & -1 & -1 \\ -1 & -1 & -1 & -1 & -1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & -1 & -1 & -1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix}$$

and  $\lambda_{B \times_{\mathcal{A}} C} = (1, 2, 1, 1, 2, 1, -1, -1, -1, -1)$ . In almost all instances the Horn matrix as constructed in Theorem 3.1.8 is not minimal, as is also the case in this example. However, it can be transformed into a minimal one via an efficient algorithm [DMS21, Lem. 3].  $\diamond$

### 3.1.4. Blending functions

We now turn our attention to geometric modelling. The central object are blending functions, giving rise to convenient parametrisations of shapes to be modelled. In this subsection we introduce the notion of blending functions, and in Subsection 3.1.5 we give explicit descriptions of blending functions of toric fibre product patches, thereby creating a large class of new blending functions satisfying the desirable property of *rational linear precision*.

Let  $P \subset \mathbb{R}^d$  be a lattice polytope with facet representation  $P = \{\mathbf{p} \in \mathbb{R}^d : \langle \mathbf{p}, n_i \rangle \geq a_i, \forall i \in [R]\}$ , where  $n_i$  is a primitive inward facing normal vector to the facet  $F_i$ . Without

loss of generality, we will always assume that  $P$  is full-dimensional inside  $\mathbb{R}^d$ . The lattice distance of a point  $\mathbf{p} \in \mathbb{R}^d$  to  $F_i$  is  $h_i(\mathbf{p}) := \langle \mathbf{p}, n_i \rangle + a_i$ ,  $i \in [R]$ . Set  $B := P \cap \mathbb{Z}^d$ , so  $B$  is the set of lattice points in  $P$  and let  $z = (z_{\mathbf{b}})_{\mathbf{b} \in B}$  be a vector of positive weights. To each  $\mathbf{b} \in B$  we associate the rational functions  $\beta_{\mathbf{b}}, \beta_z, \beta_{z,\mathbf{b}} : P \rightarrow \mathbb{R}$  defined by

$$\beta_{\mathbf{b}}(\mathbf{p}) := \prod_{i=1}^R h_i(\mathbf{p})^{h_i(\mathbf{b})}, \quad \beta_z(\mathbf{p}) := \sum_{\mathbf{b} \in B} z_{\mathbf{b}} \beta_{\mathbf{b}}(\mathbf{p}), \text{ and} \quad \beta_{z,\mathbf{b}} := z_{\mathbf{b}} \beta_{\mathbf{b}} / \beta_z. \quad (3.1.3)$$

The functions  $\beta_{z,\mathbf{b}}$  for  $\mathbf{b} \in B$  are the *toric blending functions* of the pair  $(P, z)$ , introduced by Krasauskas [Kra02] as generalisations of Bézier curves and surfaces to more general polytopes. Blending functions usually satisfy additional properties that make them amenable for computation, see for instance [Kra02]. Given a set of control points  $\{Q_{\mathbf{b}}\}_{\mathbf{b} \in B}$ , a *toric patch* is defined by the rule  $F(\mathbf{p}) := \sum_{\mathbf{b} \in B} \beta_{\mathbf{b}}(\mathbf{p}) Q_{\mathbf{b}}$ .

**Definition 3.1.10.** The pair  $(P, z)$  has *rational linear precision* if there is a set of rational functions  $\{\hat{\beta}_{\mathbf{b}}\}_{\mathbf{b} \in B}$  on  $\mathbb{C}^d$  satisfying:

1.  $\sum_{\mathbf{b} \in B} \hat{\beta}_{\mathbf{b}} = 1$ .
2. The functions  $\{\hat{\beta}_{\mathbf{b}}\}_{\mathbf{b} \in B}$  define a rational parametrisation of the toric variety  $X_{B,z}$ ,

$$\hat{\beta} : \mathbb{C}^d \dashrightarrow X_{B,z} \subset \mathbb{P}^{|B|-1}, \quad \hat{\beta}(\mathbf{t}) = (\hat{\beta}_{\mathbf{b}}(\mathbf{t}))_{\mathbf{b} \in B}.$$

3. For every  $\mathbf{p} \in \text{Relint}(P) \subset \mathbb{C}^d$ ,  $\hat{\beta}_{\mathbf{b}}(\mathbf{p})$  is defined and is a nonnegative real number.
4. Linear precision:  $\sum_{\mathbf{b} \in B} \hat{\beta}_{\mathbf{b}}(\mathbf{p}) \mathbf{b} = \mathbf{p}$  for all  $\mathbf{p} \in P$ .

The property of rational linear precision does not hold for arbitrary toric patches, but it is desirable because the blending functions “provide barycentric coordinates for general control point schemes” [GPS10]. A deep relation to algebraic statistics is provided by the following statement.

**Theorem 3.1.11** ([GPS10]). *Let  $B = P \cap \mathbb{Z}^d$  be the set of lattice points of  $P$ . The pair  $(P, z)$  has rational linear precision if and only if  $X_{B,z}$  has ML degree one.*

**Remark 3.1.12.** Henceforth, to ease notation, we drop the usage of a vector of weights  $z$  for the blending functions  $\beta_{z,\mathbf{b}}$  and the scaled projective toric variety  $X_{B,z}$ . Although we will not in general write them explicitly in the proofs, the weights play an important role whether the toric variety has ML degree one or, equivalently, if the polytope has rational linear precision. How the ML degree depends on the weights is studied in Section 3.2.

**Example 3.1.13.** Consider the point configurations as in Example 3.1.5 (see also Figure 3.1)

$$B = \{(0,0), (1,0), (0,1), (1,1)\}, \quad C = \{(0,0), (1,0), (2,0), (1,1), (0,1)\}$$

and set  $P = \text{conv}(B)$  and  $Q = \text{conv}(C)$ . The facet presentation of  $P$  is

$$P = \{(x_1, x_2) \in \mathbb{R}^2 : x_1 \geq 0, x_2 \geq 0, 1 - x_1 \geq 0, 1 - x_2 \geq 0\}.$$

The lattice distance functions of a point  $(x_1, x_2) \in \mathbb{R}^2$  to the facets of  $P$  are

$$h_1 = x_1, \quad h_2 = x_2, \quad h_3 = 1 - x_1, \quad h_4 = 1 - x_2.$$

Therefore, the toric blending functions of  $P$  with weights  $z = (1, 1, 1, 1)$  are:

$$\beta_{(0)} = (1-x_1)(1-x_2), \quad \beta_{(1)} = x_2(1-x_1), \quad \beta_{(0)} = x_1(1-x_2), \quad \beta_{(1)} = x_1x_2. \quad (3.1.4)$$

These toric blending functions satisfy the conditions in Definition 3.1.10; when this is the case,  $P$  is said to have *strict linear precision*. The polytope  $Q$  has rational linear precision for the vector of weights  $\tilde{z} = (1, 2, 1, 1, 1)$ . In this case, the toric blending functions do not satisfy condition 4 in Definition 3.1.10, however, as explained in [CC20], the following functions do:

$$\begin{aligned}\tilde{\beta}_{\binom{0}{0}} &= \frac{(1-y_2)(2-y_1-y_2)^2}{(2-y_2)^2}, & \tilde{\beta}_{\binom{1}{0}} &= \frac{2y_1(1-y_2)(2-y_1-y_2)}{(2-y_2)^2}, & \tilde{\beta}_{\binom{2}{0}} &= \frac{y_1^2(1-y_2)}{(2-y_2)^2}, \\ \tilde{\beta}_{\binom{0}{1}} &= \frac{y_2(2-y_1-y_2)}{2-y_2}, & \tilde{\beta}_{\binom{1}{1}} &= \frac{y_1y_2}{2-y_2}.\end{aligned}$$

◊

### 3.1.5. Blending functions of toric fibre products

In this subsection we show that the blending functions of the toric fibre product of two polytopes with rational linear precision can be constructed from the blending functions of the original polytopes and give an explicit formula for them. Throughout this section we use the setup for the toric fibre product introduced in Subsection 3.1.1. We let  $P = \text{conv}(B)$  and  $Q = \text{conv}(C)$  be polytopes with rational linear precision and denote their blending functions satisfying Definition 3.1.10 by  $\{\beta_j^i\}_{j \in |B_i|}^{i \in |\mathcal{A}|}$  and  $\{\beta_k^i\}_{k \in |C_i|}^{i \in |\mathcal{A}|}$ , respectively.

**Theorem 3.1.14.** *If  $P$  and  $Q$  are polytopes with rational linear precision for weights  $z, \tilde{z}$ , respectively, then the toric fibre product  $P \times_{\mathcal{A}} Q$  has rational linear precision with vector of weights  $z_{P \times_{\mathcal{A}} Q}$ . Moreover, blending functions with rational linear precision for  $P \times_{\mathcal{A}} Q$  are given by*

$$\beta_{j,k}^i(\mathbf{p}, \mathbf{q}) = \frac{\beta_j^i(\mathbf{p})\beta_k^i(\mathbf{q})}{\sum_{j' \in |B^i|} \beta_{j'}^i(\mathbf{p})} = \frac{\beta_j^i(\mathbf{p})\beta_k^i(\mathbf{q})}{\sum_{k' \in |C^i|} \beta_{k'}^i(\mathbf{q})} \quad (3.1.5)$$

where  $(\mathbf{p}, \mathbf{q}) \in P \times_{\mathcal{A}} Q$ .

**Remark 3.1.15.** The two expressions on the right-hand side of (3.1.5) are well-defined on  $\text{Relint}(P \times_{\mathcal{A}} Q)$ . The morphism  $\beta_{j,k}^i$  extends to a rational function  $\beta_{j,k}^i: \mathbb{C}^d \dashrightarrow \mathbb{C}$  where  $d = \dim(P \times_{\mathcal{A}} Q)$ . By abuse of notation, we will sometimes write

$$\beta_{j,k}^i(\mathbf{t}) = \frac{1}{N^i(\mathbf{t})} \beta_j^i(\mathbf{t}) \beta_k^i(\mathbf{t}),$$

where  $\mathbf{t} \in \mathbb{C}^d$  and  $N^i(\mathbf{t})$  denotes the denominator as in (3.1.5). The most difficult part in the proof of Theorem 3.1.14 is to show that the two expressions on the right-hand side of (3.1.5) agree on  $\text{Relint}(P \times_{\mathcal{A}} Q)$ , which might seem surprising at first.

The following example illustrates the construction in Theorem 3.1.14.

**Example 3.1.16.** Consider the polytopes  $P$  and  $Q$  from Example 3.1.13, with their vectors of weights  $z$  and  $\tilde{z}$ . By Theorem 3.1.14, the blending functions for  $P \times_{\mathcal{A}} Q$  are

$$\frac{\beta_j^i \tilde{\beta}_k^i}{\sum_j \beta_j^i} = \frac{\beta_j^i \tilde{\beta}_k^i}{\sum_k \tilde{\beta}_k^i}.$$

For example, the blending function corresponding to  $(\mathbf{b}_2^1 \ \mathbf{c}_3^1)^T$  is

$$\beta_{2,3}^1 = \frac{\beta_2^1 \tilde{\beta}_3^1}{\beta_1^1 + \beta_1^2} = \frac{x_1(1-x_2)y_1^2(1-y_2)}{(1-x_2)(2-y_2)^2} = \frac{\beta_2^1 \tilde{\beta}_3^1}{\tilde{\beta}_1^1 + \tilde{\beta}_2^1 + \tilde{\beta}_3^1} = \frac{x_1(1-x_2)y_1^2(1-y_2)}{(1-y_2)(2-y_2)^2}.$$

While the denominators are different, the two expressions agree on  $\text{Relint}(P \times_{\mathcal{A}} Q)$ . ◊

Before proving Theorem 3.1.14 we will prove two lemmas which will be used in the final proof. Our first lemma demonstrates how the blending functions behave on certain faces of  $P$  and  $Q$ . The second lemma shows that the two parametrisations in (3.1.5) yield the same MLE for a generic data point  $u$ .

**Lemma 3.1.17.** *Let  $P^i$  be the subpolytope defined by  $P^i = \text{conv}\{\mathbf{b}_j^i : j \in |B^i|\}$ . Then, for  $\mathbf{p} \in P^i$ ,*

$$\sum_{j \in |B^i|} \beta_j^i(\mathbf{p}) = 1.$$

*Proof.* By assumption,  $\beta: \mathbb{C}^{d_1} \dashrightarrow X_B$ ,  $\beta(\mathbf{t}) = (\beta_j^i(\mathbf{t}))_{j \in |B^i|}^{i \in |\mathcal{A}|}$  is a rational parametrisation of  $X_B$ . Let  $X_B^i$  be the toric variety associated to  $P^i$ ; we claim that  $X_B^i$  is parametrised by  $(\beta_j^i(\mathbf{t}))_{j \in |B^i|}$  and setting all other coordinates of  $\beta$  to zero. Indeed, consider the linear map

$$\deg: P \rightarrow \text{conv}(\mathcal{A}), \mathbf{b}_j^i \mapsto \mathbf{a}^i.$$

As  $\mathcal{A}$  is linearly independent,  $\mathbf{a}^i$  is a vertex of  $\text{conv}(\mathcal{A})$ . Note that  $P^i = \deg^{-1}(\mathbf{a}^i)$ ; as preimages of faces under linear maps are again faces,  $P^i$  is a face of  $P$ . The claim then follows from the Orbit-face-correspondence (Theorem 2.1.6): we know that  $\sum_{(i,j) \in |B|} \beta_j^i = 1$ . On  $P^i$ , all  $\beta_j^{i'}$  for  $i' \neq i$  vanish, so we must have  $\sum_{j \in |B^i|} \beta_j^i(\mathbf{p}) = 1$  for  $\mathbf{p} \in P^i$ .  $\square$

We record the following fact as a consequence of the proof above.

**Corollary 3.1.18.** *Let  $P$  be a polytope equipped with a linearly independent multigrading  $\mathcal{A}$ . Then  $P^i = \text{conv}(\mathbf{b}_j^i \mid j \in |B^i|)$  is a face of  $P$ .*

**Example 3.1.19.** For the polytope  $Q$  and the grading  $\deg$  as shown in Figure 3.1, we have  $Q^1 = \text{conv}(\mathbf{c}_1^1, \mathbf{c}_2^1, \mathbf{c}_3^1)$  and  $Q^2 = \text{conv}(\mathbf{c}_1^2, \mathbf{c}_2^2)$ . To illustrate the result of Lemma 3.1.17, note that the sum of the blending functions associated to  $Q^1$  is equal to  $1 - y_2$ .  $\diamond$

**Lemma 3.1.20.** *Let  $P$  and  $Q$  be polytopes with rational linear precision and  $\beta_1, \beta_2$  be two rational functions defined by*

$$\beta_1(\mathbf{t}) = \left( \frac{\beta_j^i(\mathbf{t})\beta_k^i(\mathbf{t})}{\sum_{j' \in |B^i|} \beta_{j'}^i(\mathbf{t})} \right)_{(j,k) \in |B^i \times C^i|}^{i \in |\mathcal{A}|} \quad \text{and} \quad \beta_2(\mathbf{t}) = \left( \frac{\beta_j^i(\mathbf{t})\beta_k^i(\mathbf{t})}{\sum_{k' \in |C^i|} \beta_{k'}^i(\mathbf{t})} \right)_{(j,k) \in |B^i \times C^i|}^{i \in |\mathcal{A}|}.$$

For a data point  $u = (u_{j,k}^i)_{(j,k) \in |B^i \times C^i|}^{i \in |\mathcal{A}|}$ , set

$$\mathbf{p} = \sum_{(i,j,k) \in |B \times \mathcal{A} C|} \frac{u_{j,k}^i}{u_{+,+}^i} \mathbf{m}_{j,k}^i \in \mathbb{C}^d.$$

Then the maximum likelihood estimate for  $X_{B \times \mathcal{A} C}$  is

$$\beta_1(\mathbf{p}) = \beta_2(\mathbf{p}) = (\hat{p}_{j,k}^i)_{(j,k) \in |B^i \times C^i|}^{i \in |\mathcal{A}|}.$$

*Proof.* As  $P$  and  $Q$  have rational linear precision, by [CC20, Prop. 8.4] we have  $\beta_j^i(\mathbf{p}) = (\hat{p}_B)_j^i$  and  $\beta_k^i(\mathbf{p}) = (\hat{p}_C)_k^i$ . Furthermore, by [AKK20, Thm. 5.5], the MLE of the toric fibre product is given by

$$\hat{p}_{j,k}^i = \frac{(\hat{p}_B)_j^i (\hat{p}_C)_k^i}{(\hat{p}_{\mathcal{A}})^i}.$$

As a consequence of Birch's Theorem (Theorem 2.2.11), it follows from the proof of [AKK20, Lem. 5.10] that  $(\hat{p}_B)_+^i = \frac{u_{+,+}^i}{u_{+,+}^{i+}} = (\hat{p}_{\mathcal{A}})^i$ , and analogously  $(\hat{p}_C)_+^i = (\hat{p}_{\mathcal{A}})^i$ . Therefore,

$$\sum_{j' \in |B^i|} \beta_{j'}^i(\mathbf{p}) = (\hat{p}_B)_+^i = \sum_{k' \in |C^i|} \beta_{k'}^i(\mathbf{p}) = (\hat{p}_C)_+^i = (\hat{p}_{\mathcal{A}})^i$$

and the desired statement follows.  $\square$

We are now ready to prove Theorem 3.1.14.

*Proof.* Having rational linear precision is equivalent to having ML degree one by Theorem 3.1.11. Then the first statement is a direct consequence of the multiplicativity of the ML degree under toric fibre products [AKK20, Thm. 5.5].

We first show that both expressions in (3.1.5) define rational parametrisations

$$\beta_1(\mathbf{t}) = \left( \frac{\beta_j^i(\mathbf{t}) \beta_k^i(\mathbf{t})}{\sum_{j' \in |B^i|} \beta_{j'}^i(\mathbf{t})} \right)_{(j,k) \in |B^i \times C^i|}^{i \in |\mathcal{A}|}, \quad \beta_2(\mathbf{t}) = \left( \frac{\beta_j^i(\mathbf{t}) \beta_k^i(\mathbf{t})}{\sum_{k' \in |C^i|} \beta_{k'}^i(\mathbf{t})} \right)_{(j,k) \in |B^i \times C^i|}^{i \in |\mathcal{A}|}$$

of  $X_{B \times_{\mathcal{A}} C}$ . To do this, we first show that the products  $\beta_j^i \beta_k^i$  parametrise  $X_{B \times_{\mathcal{A}} C}$  and the result then follows since  $\beta_1$  and  $\beta_2$  are equivalent to  $\beta_j^i \beta_k^i$  under the torus action associated to the multigrading  $\mathcal{A}$ . Let  $\phi : \mathbb{C}^{|B|} \times \mathbb{C}^{|C|} \rightarrow \mathbb{C}^{|B \times_{\mathcal{A}} C|}$  be the map given by

$$\phi(\mathbf{x}, \mathbf{y}) = (x_{j,k}^i y_k^i)_{(j,k) \in |B^i \times C^i|}^{i \in |\mathcal{A}|}.$$

Then the toric fibre product  $X_{B \times_{\mathcal{A}} C}$  is precisely given by  $\phi(X_B \times X_C)$ . Since the blending functions  $\beta_j^i$  and  $\beta_k^i$  parametrise  $X_B$  and  $X_C$ , respectively, and  $\beta_j^i \beta_k^i = \phi \circ (\beta_j^i, \beta_k^i)$ , we immediately get that the  $\beta_j^i \beta_k^i$  parametrise  $X_{B \times_{\mathcal{A}} C}$ . Now observe that the multigrading  $\mathcal{A}$  induces an action of the torus  $T_{\mathcal{A}} = (\mathbb{C}^*)^{|\mathcal{A}|}$  via

$$T_{\mathcal{A}} \times X_{B \times_{\mathcal{A}} C} \rightarrow X_{B \times_{\mathcal{A}} C}, \quad (t^1, \dots, t^{|\mathcal{A}|}) \cdot (x_{j,k}^i)_{(j,k) \in |B^i \times C^i|}^{i \in |\mathcal{A}|} = (t^i x_{j,k}^i)_{(j,k) \in |B^i \times C^i|}^{i \in |\mathcal{A}|}.$$

Define  $\tau : \mathbb{C}^{d_1} \rightarrow T_{\mathcal{A}}$  by

$$\tau = (\tau^1, \dots, \tau^{|\mathcal{A}|}), \quad \tau^i(\mathbf{t}) = \begin{cases} \left( \sum_{j \in |B^i|} \beta_j^i(\mathbf{t}) \right)^{-1} & \text{if } \sum_{j \in |B^i|} \beta_j^i(\mathbf{t}) \neq 0 \\ 1 & \text{else.} \end{cases}$$

Note that  $\tau(\mathbf{t}) \in T_{\mathcal{A}}$  and

$$\tau(\mathbf{x}) \cdot (\beta_j^i(\mathbf{x}) \beta_k^i(\mathbf{x}))_{j \in |B^i|, k \in |C^i|}^{i \in |\mathcal{A}|} = \beta_1(\mathbf{x}) \quad \text{for all } \mathbf{x} \in P \times_{\mathcal{A}} Q,$$

showing that  $\beta_j^i(\mathbf{x}) \beta_k^i(\mathbf{x})$  and  $\beta_1(\mathbf{x})$  lie in the same  $T_{\mathcal{A}}$ -orbit. A similar argument shows the same for  $\beta_2(\mathbf{x})$ , thus both  $\beta_1$  and  $\beta_2$  parametrise  $X_{B \times_{\mathcal{A}} C}$ .

We will now show that the two expressions in 3.1.5 are equal. Let us define a new  $\tau : \mathbb{C}^{d_1+d_2} \rightarrow T_{\mathcal{A}}$  by

$$\tau = (\tau^1, \dots, \tau^{|\mathcal{A}|}), \quad \tau^i(\mathbf{t}) = \begin{cases} \frac{\sum_{j \in |B^i|} \beta_j^i(\mathbf{t})}{\sum_{k \in |C^i|} \beta_k^i(\mathbf{t})} & \text{if } \sum_{j \in |B^i|} \beta_j^i(\mathbf{t}) \neq 0 \neq \sum_{k \in |C^i|} \beta_k^i(\mathbf{t}) \\ 1 & \text{else.} \end{cases}$$

Clearly,  $\tau(\mathbf{t}) \in T_{\mathcal{A}}$ ; we claim that  $\tau(\mathbf{x}).\beta_1(\mathbf{x}) = \beta_2(\mathbf{x})$  for  $\mathbf{x} \in P \times_{\mathcal{A}} Q$ . First consider the case  $\mathbf{x} \in P^i \times Q^i$ , with  $P^i$  and  $Q^i$  defined as in Lemma 3.1.17. By the Orbit-face-correspondence (Theorem 2.1.6) applied to the  $T_{\mathcal{A}}$ -action, all coordinates in  $\beta_1(\mathbf{x})$  and  $\beta_2(\mathbf{x})$  vanish except for those graded by  $\mathbf{a}^i$ . By Lemma 3.1.17,  $\sum_{j \in |B^i|} \beta_j^i(\mathbf{x}) = \sum_{k \in |C^i|} \beta_k^i(\mathbf{x}) = 1$ , so in particular the claim holds. Now consider the case where  $\mathbf{x} \notin \bigcup_{i \in |\mathcal{A}|} P^i \times Q^i$ . Then, again by the Orbit-face-correspondence applied to the  $T_{\mathcal{A}}$ -action, for each  $i \in |\mathcal{A}|$  there exist  $j \in |B^i|$  and  $k \in |C^i|$  such that  $\beta_j^i(\mathbf{x}), \beta_k^i(\mathbf{x}) \neq 0$ . Thus, by definition,  $\tau(\mathbf{x}).\beta_1(\mathbf{x}) = \beta_2(\mathbf{x})$ . We conclude that for all  $\mathbf{x} \in P \times_{\mathcal{A}} Q$ ,  $\beta_1(\mathbf{x})$  and  $\beta_2(\mathbf{x})$  lie in the same  $T_{\mathcal{A}}$ -orbit. Equality of  $\beta_1$  and  $\beta_2$  then follows once there exists at least one point in each orbit where the two parametrisations agree. This is indeed the case: for the maximal orbit this is the point given in Lemma 3.1.20, for smaller orbits corresponding to faces of  $P^i \times Q^i$  we can pick a point as in Lemma 3.1.17.

It now remains to show that the  $\beta_{j,k}^i$  sum to one and that they satisfy linear precision. This follows from direct computation. Firstly, we have

$$\sum_{(i,j,k) \in |B \times_{\mathcal{A}} C|} \beta_{j,k}^i = \sum_{i \in |\mathcal{A}|, k \in |C^i|} \beta_k^i \sum_{j \in |B^i|} \frac{\beta_j^i}{\sum_{j' \in |B^i|} \beta_{j'}^i} = \sum_{i \in |\mathcal{A}|, k \in |C^i|} \beta_k^i = 1.$$

Finally, we compute

$$\begin{aligned} \sum_{(i,j,k) \in |B \times_{\mathcal{A}} C|} \beta_{j,k}^i(\mathbf{p}) \mathbf{m}_{j,k}^i &= \sum_{i \in |\mathcal{A}|, j \in |B^i|} \beta_j^i(\mathbf{p}) \sum_{k \in |C^i|} \frac{\beta_k^i(\mathbf{p})}{\sum_{k' \in |C^i|} \beta_{k'}^i(\mathbf{p})} (\mathbf{b}_j^i, 0) \\ &\quad + \sum_{i \in |\mathcal{A}|, k \in |C^i|} \beta_k^i(\mathbf{p}) \sum_{j \in |B^i|} \frac{\beta_j^i(\mathbf{p})}{\sum_{j' \in |B^i|} \beta_{j'}^i(\mathbf{p})} (0, \mathbf{c}_k^i) \\ &= \left( \sum_{i \in |\mathcal{A}|, j \in |B^i|} \beta_j^i(\mathbf{p}) \mathbf{b}_j^i, \sum_{i \in |\mathcal{A}|, k \in |C^i|} \beta_k^i(\mathbf{p}) \mathbf{c}_k^i \right) = \mathbf{p}. \end{aligned}$$

Therefore, the  $\beta_{j,k}^i$  constitute blending functions with rational linear precision.  $\square$

## 3.2. Euler stratifications of hypersurface families

In this section we take a more holistic view on Problem 3.0.1. Instead of focussing on the case of ML degree one, we develop algorithms to give a complete computational answer to Problem 3.0.1. To this end, we introduce *Euler stratifications*. We study them in the general setup of hypersurface families. Euler stratifications stratify a parameter locus into strata over which each fibre has constant Euler characteristic. In particular, for a certain choice of hypersurface family, an Euler stratification divides a parameter space  $Z$  into Euler strata  $S_i$  such that for all  $z \in S_i$ , the ML degree of  $X_{A,z}$  is constant. The relationship between the ML degree and the Euler characteristic is established by Theorem 2.2.12.

A homogeneous polynomial  $0 \neq F \in \mathbb{C}[x_0, \dots, x_d]$  of degree  $n$  defines a hypersurface  $\mathcal{V}(F) \subset \mathbb{P}^d$ . We study the dependence of the (topological) Euler characteristic of  $\mathcal{V}(F)$  on the coefficients of  $F$ . More precisely, we consider a family  $\mathcal{V}(F(x,z)) \subset \mathbb{P}^d \times Z$  of hypersurfaces over an irreducible variety  $Z$ . We seek to compute an explicit description of the loci in  $Z$  on which the hypersurface  $\mathcal{V}(F(x,z))$  has a fixed Euler characteristic. An Euler stratification decomposes  $Z$  into such loci, see Subsection 3.2.1 for a precise definition.

We must clarify what we mean by an “explicit description” of these loci, or of the strata in an Euler stratification. It turns out that Euler strata are constructible, meaning essentially that they can be described by polynomials. We illustrate this for plane conics.

**Example 3.2.1** ( $d = n = 2$ ). Let  $Z = \mathbb{P}^5$  and consider a generic ternary quadric

$$F(x_0, x_1, x_2, z) = z_0 x_0^2 + z_1 x_0 x_1 + z_2 x_0 x_2 + z_3 x_1^2 + z_4 x_1 x_2 + z_5 x_2^2. \quad (3.2.1)$$

The Euler characteristic of the curve  $\mathcal{V}(F(x,z))$  is either two or three. Indeed, if  $\mathcal{V}(F(x,z))$  is smooth and  $F$  is irreducible then the Euler characteristic is  $\chi(\mathcal{V}(F(x,z))) = 2$ . The same is true when  $F = L^2$  is the square of a linear form  $L$ . If  $\mathcal{V}(F(x,z))$  is singular, then it is the union of two distinct lines, and it has Euler characteristic three. Algebraically, we have

$$\chi(\mathcal{V}(F(x,z))) = 3 \iff \text{rank } M(z) = 2, \quad \text{where } M(z) = \begin{pmatrix} 2z_0 & z_1 & z_2 \\ z_1 & 2z_3 & z_4 \\ z_2 & z_4 & 2z_5 \end{pmatrix}. \quad (3.2.2)$$

This defines a constructible subset of  $\mathbb{P}^5$ , the regular locus of the hypersurface

$$\nabla = \{z \in Z : \Delta = 0\}, \text{ where } \Delta = 8z_0z_3z_5 - 2z_0z_4^2 - 2z_1z_5^2 + 2z_1z_2z_4 - 2z_2^2z_3. \quad (3.2.3)$$

The singular locus  $\nabla_{\text{sing}}$  is cut out by the  $2 \times 2$ -minors of the matrix  $M(z)$  in (3.2.2).  $\diamond$

Euler strata in the parameter space  $Z$  are closely related to the equisingular loci of our hypersurface families, i.e. the sets of parameter values for which the singularities of  $\mathcal{V}(F(x,z))$  are of the same type, in some appropriate sense. These loci are not well understood in general. For points in  $\mathbb{P}^1$  ( $d = 1$ ), the problem comes down to computing coincident root loci [Chi03]. Computations up to degree  $n = 7$  are reported in [LS16]. For plane curves, the most well studied equisingular loci are Severi varieties. These are loci of nodal curves of fixed degree with a fixed number of nodes [Ful82]. For tropical approaches to Severi varieties, see for instance [DHT17, Yan13]. Severi varieties for surfaces in  $\mathbb{P}^3$  were studied in [CC99]. In 1988, Diaz and Harris wrote that the equisingular stratification of plane curves of a fixed degree was out of reach [DH88, p. 1]. This section addresses the problem using modern tools from computational algebraic geometry. We elaborate more on equisingular loci in a paragraph at the end of Section 3.2.1.

A special case which deserves extra attention is when  $F$  factors as

$$F(x, z) = x_0 \cdots x_d f(x, z). \quad (3.2.4)$$

Let  $T \subset \mathbb{P}^d$  be the dense torus of  $\mathbb{P}^d$ , i.e.  $T \cong (\mathbb{C}^*)^d \subset \mathbb{P}^d$ . In this case, we have

$$\chi(\mathcal{V}(F)) = \chi(\mathbb{P}^d) - \chi(T) + \chi(\mathcal{V}(F) \cap T) = (d+1) + \chi(\mathcal{V}(f) \cap T).$$

Hence, an Euler stratification for the projective hypersurface  $\mathcal{V}(F)$  is an Euler stratification for the very affine hypersurface  $\mathcal{V}(f) \cap T$ . Very affine varieties are central in tropical geometry [MS21]. They appear as statistical models in algebraic statistics [CHKO24, HS14] and as integration spaces in particle physics [MHMT23]. In these applications, the Euler characteristic of the very affine variety at hand plays a crucial role; see for example Theorem 2.2.12 and [MHMT23, Thm. 3.14]. In statistics, the absolute value of the Euler characteristic coincides with the maximum likelihood degree of the corresponding model, which measures the algebraic complexity of maximum likelihood estimation [CHKS06]. In physics,  $|\chi(\mathcal{V}(f)) \cap T|$  is the dimension of a vector space of integrals, and it measures the complexity of integration by parts reduction [AFST24, BBKP19, MHMT23].

With regard to Equation (3.2.4), we may consider equisingular loci for the zero set of  $f(x, z)$  in  $T$ . If, in addition, the parameters  $z$  appear as coefficients of  $f$  (as in (3.2.1)), then this leads us to study  $A$ -discriminants and principal  $A$ -determinants as defined by Gel'fand, Kapranov and Zelevinsky (GKZ) [GKZ08, Ch. 9–10]. We illustrate this for our conics example from before.

**Example 3.2.2.** Let  $f(x, z)$  be the polynomial in (3.2.1). Its exponents are the columns of

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

The cubic defining equation  $\Delta$  of the hypersurface  $\nabla$  in Example 3.2.1 is the  $A$ -discriminant associated to this matrix. By [ABB<sup>+</sup>19, Thm. 13], the Euler characteristic of  $\mathcal{V}(f(x, z)) \cap T$  equals  $-4$ , unless the principal  $A$ -determinant vanishes:

$$E_A = \Delta \cdot (z_1^2 - 4z_0z_3) \cdot (z_2^2 - 4z_0z_5) \cdot (z_4^2 - 4z_3z_5) \cdot z_0 \cdot z_3 \cdot z_5 = 0. \quad (3.2.5)$$

For a generic point in the 4-dimensional variety  $\nabla_\chi = \{z \in \mathbb{P}^5 : E_A(z) = 0\}$ , the Euler characteristic is  $\chi(\mathcal{V}(f(x, z)) \cap T) = -3$ . Overall, the possible Euler characteristics are  $0, -1, -2, -3, -4$ . In Section 3.2.4, we will decompose  $\nabla_\chi$  into 70 Euler strata.  $\diamond$

Esterov studies Euler characteristics and multisingularity strata in the GKZ setting in [Est13, Est17]. In particular, in [Est13, §1.2] he coins the term “Euler discriminant”, which was picked up in the context of Feynman integrals in [FMT24]. In this section, the Euler discriminant variety  $\nabla_\chi$  is the subvariety of  $Z$  obtained as the closure of all  $z$  for which the Euler characteristic  $\chi(\mathcal{V}(F(x, z)))$  takes non-generic values. If  $\nabla_\chi$  is defined by a single equation  $\Delta_\chi = 0$ , then we call  $\Delta_\chi$  the Euler discriminant polynomial. Since computing an Euler stratification is an iterated computation of Euler discriminants, understanding these discriminants is crucial to our story. The formulae for  $\chi(\mathcal{V}(F))$  stated by Dimca and Papadima [DP03] and Huh [Huh12, Huh13] are a key to success for computing Euler discriminants using computer algebra software.

Euler stratifications are coarsenings of Whitney stratifications [Whi65]. More precisely, by Thom’s first isotopy lemma [Mat12, Prop. 11.1], the strata of a Whitney stratification

parametrise varieties of constant topological type, and thus in particular of constant Euler characteristic. Brown's definition of the Landau variety in [Bro09, Def. 54] is based on Thom's isotopy lemma; it is the union of all Whitney strata of codimension one. This variety captures the singular locus of Feynman integrals, viewed as multivalued functions of kinematic parameters. The Landau variety contains the Euler discriminant, but the two do not always coincide, see Examples 3.2.9 and 3.2.10. However, in physics, they often do.

Recent efforts by Helmer and Nanda have lead to symbolic algorithms for computing Whitney stratifications [HN23]. This was applied to Feynman integrals and Landau varieties in [HPT24]. That paper shows that the algorithms work well in small examples, but they run out of steam for more challenging integrals, such as those tackled in [FMT24]. However, the efficient symbolic-numerical methods in [FMT24] often lead to incomplete results, resulting in a variety that is strictly contained in the Euler discriminant, called the principal Landau determinant.

The section is organised as follows. Subsection 3.2.1 gives a definition of Euler stratifications (Definition 3.2.3), proves existence for a general class of hypersurface families (Lemma 3.2.6) and presents some first examples. We discuss the relation to Whitney stratifications and multisingularity strata. Subsection 3.2.2 revisits coincident root loci for binary forms. We summarise known results and deduce an explicit characterisation of the Euler stratification (Theorems 3.2.13 and 3.2.14). We also present a new algorithm for the Euler stratification of hyperplane sections of smooth projective curves (Algorithm 2). In Subsection 3.2.3, we develop general algorithms (Algorithms 3, 4 and 5) for computing the Euler stratification of projective hypersurface families. For families of plane curves with isolated singularities, we prove that the Euler discriminant equals the polar discriminant (Proposition 3.2.24). Subsection 3.2.4 is dedicated to the very affine setting. We prove that the degree of the Gauss map counts the critical points of a logarithmic potential (Proposition 3.2.28). This is then related to Huh's result (Theorem 3.2.27) using likelihood degenerations [ABF<sup>+</sup>23]. We prove two structural results in Theorem 3.2.31 and Proposition 3.2.32, and present the specialised Algorithm 6 for computing Euler discriminants in this setting. Following Subsection 3.2.5, we present a gallery of examples, including coincident root loci of binary octics (Subsection 3.2.6), matroid stratifications of bilinear forms (Subsection 3.2.7), Landau varieties of Feynman integrals (Subsection 3.2.8), maximum likelihood stratifications for toric Fano varieties (Subsection 3.2.9) and Hirzebruch surfaces (Subsection 3.2.10). Our code and computational results are available online at MathRepo [TW24b].

### 3.2.1. Definitions and first examples

Our first task is to define Euler stratifications, and to show that they exist for sufficiently general families of varieties. A *quasi-projective variety* is a subset  $X \subset \mathbb{P}^d$  of projective space which can be written as  $X = V \setminus W$ , where  $V, W \subset \mathbb{P}^d$  are closed subvarieties.

**Definition 3.2.3.** Let  $\mathcal{X}, Z$  be quasi-projective varieties, with  $Z$  irreducible. Consider a surjective morphism  $\pi : \mathcal{X} \rightarrow Z$  whose fibres have constant dimension. An *Euler stratification* of  $\pi$  is a finite set  $\mathcal{S}$  of quasi-projective subvarieties of  $Z$  such that

1. when  $S \neq S' \in \mathcal{S}$ , then  $S \cap S' = \emptyset$ , and  $\bigsqcup_{S \in \mathcal{S}} S = Z$ ,
2. for each stratum  $S \in \mathcal{S}$ , the closure  $\overline{S}$  is a union of strata:  $\overline{S} = \bigsqcup_{S' \subseteq \overline{S}} S'$ ,
3. the Euler characteristic of the fibre  $\chi(\pi^{-1}(z))$  is constant for  $z \in S$ .

The set  $\mathcal{S}$  is partially ordered as follows:  $S \preceq S'$  if and only if  $\overline{S} \subseteq \overline{S}'$ .

The quasi-projective subvarieties  $S$  in an Euler stratification  $\mathcal{S}$  are called *(Euler) strata*, and their closures  $\bar{S}$  are the *closed (Euler) strata*. Euler strata need not be irreducible.

We make the following concrete choices for  $\mathcal{X}$  and  $Z$ . Let  $F \in \mathbb{C}[z][x]$  be a bi-homogeneous polynomial in the variables  $x = (x_0, \dots, x_d)$  and parameters  $z = (z_0, \dots, z_m)$ . Let  $Z \subseteq \mathbb{P}^m$  be any irreducible quasi-projective subvariety. We define

$$\mathcal{X}_F = \{(x, z) \in \mathbb{P}^d \times Z : F(x, z) \neq 0\}. \quad (3.2.6)$$

We want to stratify the family of hypersurface complements given by the projection

$$\pi_F : \mathcal{X}_F \rightarrow Z, \quad (x, z) \mapsto z. \quad (3.2.7)$$

We denote the fibres of this map by  $\mathcal{X}_{F,z} = \pi_F^{-1}(z)$ . We assume that the set  $\{z \in Z : F(x, z) \equiv 0\}$  is empty, so that all fibres are  $d$ -dimensional.

**Remark 3.2.4.** Alternatively, we may consider the family of hypersurfaces

$$\mathcal{X}_F^c = \{(x, z) \in \mathbb{P}^d \times Z : F(x, z) = 0\}$$

with closed fibres  $\mathcal{X}_{F,z}^c = \mathbb{P}^d \setminus \mathcal{X}_{F,z}$ . The equality  $\chi(\mathcal{X}_{F,z}) = \chi(\mathbb{P}^d) - \chi(\mathcal{X}_{F,z}^c) = d + 1 - \chi(\mathcal{X}_{F,z}^c)$  implies that an Euler stratification of  $\pi_F$  is one of  $\pi_F^c : \mathcal{X}_F^c \rightarrow Z$  and vice versa.

**Example 3.2.5** (Conics in  $\mathbb{P}^2$ ). Consider the family of conics  $\mathcal{X}_F = \{(x, z) \in \mathbb{P}^2 \times \mathbb{P}^5 : F(x; z) \neq 0\}$ , where  $F$  is as in (3.2.1), with its coordinate projection  $\pi : \mathcal{X}_F \rightarrow \mathbb{P}^5$ . By the discussion in Example 3.2.1, the poset  $\mathcal{S}$  consists of three strata, of dimensions five, four, and two. The closed four-dimensional stratum is given by the discriminant  $\Delta = 0$ , with  $\Delta$  as in (3.2.3). The smallest stratum is, up to scaling coordinates, the second Veronese embedding of  $\mathbb{P}^2$ , whose binomial ideal is given by the  $2 \times 2$ -minors of  $M(z)$  from (3.2.2). Notice that there is no stratum of dimension three, and the Euler characteristic drops by one on the discriminant.  $\diamond$

Example 3.2.5 illustrates that the loci  $Z_k = \{z \in Z : \chi(\mathcal{X}_{F,z}) = k\}$  of constant Euler characteristic are not necessarily quasi-projective, even when  $\mathcal{X}_F$  and  $Z$  are quasi-projective. We will show that they are always *constructible*. A constructible set is a finite union of quasi-projective varieties  $(V_1 \setminus W_1) \cup \dots \cup (V_\ell \setminus W_\ell)$ .

**Proposition 3.2.6.** *Let  $F, \mathcal{X}_F, Z$  be as in (3.2.6) and let  $\pi_F : \mathcal{X}_F \rightarrow Z$  be the coordinate projection (3.2.7) with fibres  $\mathcal{X}_{F,z} = \pi_F^{-1}(z)$ . For any integer  $k$ , the set  $Z_k = \{z \in Z : \chi(\mathcal{X}_{F,z}) = k\}$  is constructible. In particular, there exists an Euler stratification of the map  $\pi_F$ .*

We prove Proposition 3.2.6 using a result of Dimca and Papadima [DP03, Thm. 1].

**Theorem 3.2.7.** *Let  $F \in \mathbb{C}[x_0, \dots, x_d]$  be any non-constant homogeneous polynomial, and let  $D(F) = \{x \in \mathbb{P}^d : F(x) \neq 0\}$  be its hypersurface complement. Consider the Gauss map*

$$\nabla F : D(F) \rightarrow \mathbb{P}^d, \quad x \mapsto \left( \frac{\partial F}{\partial x_0}(x) : \dots : \frac{\partial F}{\partial x_d}(x) \right).$$

We have  $\deg \nabla F = (-1)^d \cdot \chi(D(F) \setminus H)$ , where  $H \subset \mathbb{P}^d$  is a generic hyperplane.

*Proof of Proposition 3.2.6.* The formula for  $\chi(D(F) \setminus H)$  from Theorem 3.2.7 implies that

$$\begin{aligned} (-1)^d \cdot \deg \nabla F(x, z) &= \chi(\mathcal{X}_{F,z} \setminus H) \\ &= \chi(\mathcal{X}_{F,z}) - \chi(\mathcal{X}_{F,z} \cap H) \\ &= \chi(\mathcal{X}_{F,z}) - (\chi(\mathbb{P}^{d-1}) - \chi(\mathcal{X}_{F,z}^c \cap H)) \end{aligned}$$

for all  $z \in Z$  (notice that  $\mathcal{X}_{F,z} = D(F(x,z))$  for any  $z \in Z$ ). Using  $\chi(\mathbb{P}^{d-1}) = d$  we find

$$\chi(\mathcal{X}_{F,z}) = d - \chi(\mathcal{X}_{F,z}^c \cap H) + (-1)^d \cdot \deg \nabla F(x,z). \quad (3.2.8)$$

For any  $k$ , we claim that the set  $W_k = \{z \in Z : \deg \nabla F(x,z) = k\}$  is constructible. To see this, note that the ramification locus of  $\{(x,z,b) \in \mathcal{X}_F \times \mathbb{P}_b^d : \nabla F(x,z) = b\} \rightarrow Z \times \mathbb{P}_b^d$  is closed, and its image  $B \subset Z \times \mathbb{P}_b^d$  is constructible by Chevalley's Theorem. Let  $B' \subset Z$  be the constructible subset consisting of the points  $z \in Z$  whose fibre along the projection  $B \rightarrow Z$  is  $d$ -dimensional. By construction, we have  $Z \setminus B' = W_{n^*}$ , where  $n^*$  is the generic degree of  $\nabla F(x,z)$  for  $z \in Z$ . The same argument applies when we replace  $Z$  by any of the irreducible components of  $\overline{B'} \subset Z$ . The claim follows by iterating this process.

For  $d = 1$ , we have  $\mathcal{X}_{F,z}^c \cap H = \emptyset$  and  $Z_k$  is constructible by (3.2.8). Proceed by induction on  $d$ : if the statement is true in dimension  $d - 1$ , then  $\{z \in Z : \chi(\mathcal{X}_{F,z}^c \cap H) = k\}$  is constructible for any  $k$  (Remark 3.2.4), which implies by (3.2.8) that  $Z_k$  is constructible.  $\square$

As mentioned in the beginning, the applications we have in mind require Euler stratifications of families of *very affine* hypersurface complements. That is, we are interested in the Euler characteristic of the complement of the zero locus of a polynomial  $f$  in the torus

$$T = D(x_0 \cdots x_d) = \{(x_0 : \cdots : x_d) \in \mathbb{P}^d : x_0 \cdots x_d \neq 0\} \cong (\mathbb{C}^*)^d.$$

We continue to assume that  $Z \subseteq \mathbb{P}^m$  is an irreducible quasi-projective variety, and we set

$$\mathcal{X}_f^* = \{(x,z) \in T \times Z : f(x,z) \neq 0\} = \mathcal{X}_F, \quad (3.2.9)$$

where  $F = x_0 \cdots x_d f$ . The following is a consequence of Proposition 3.2.6 and  $\mathcal{X}_f^* = \mathcal{X}_F$ .

**Lemma 3.2.8.** *Let  $f \in \mathbb{C}[z][x]$  be any bihomogeneous polynomial and let  $Z, \mathcal{X}_f^*$  be as in (3.2.9). The coordinate projection  $\pi_f : \mathcal{X}_f^* \rightarrow Z$  admits an Euler stratification.*

**Whitney stratifications.** Proposition 3.2.6 can also be proved via the theory of *Whitney stratifications*, originally developed by Thom [Tho64] and Whitney [Whi65]. This proof strategy works more generally when  $\pi : \mathcal{X} \rightarrow Z$  is a proper morphism of (abstract) varieties. A statement that fits our scope precisely is hard to find in the literature, and a self-contained explanation would be too much of a digression. Since the identity (3.2.8) is useful in later sections, we chose to include the proof above and limit ourselves to a sketch of the general argument.

In [Whi65, Thm. 19.2], Whitney states that every *variety* admits what is now called a Whitney stratification. Here, variety means complex analytic variety. The fact that every constructible set admits a Whitney stratification whose strata are again constructible is stated explicitly and with an outline of proof in [Wal06, pp. 336–337]. A Whitney stratification of a proper morphism  $\pi : \mathcal{X} \rightarrow Z$  (such as our morphism  $\pi_F^c : \mathcal{X}_F^c \rightarrow Z$ ) is a Whitney stratification of  $\mathcal{X}$  and  $Z$  such that  $\pi$  maps each stratum  $S_{\mathcal{X}}$  of  $\mathcal{X}$  into a stratum  $S_Z$  of  $Z$ , and the restriction  $\pi_{S_{\mathcal{X}}} : S_{\mathcal{X}} \rightarrow S_Z$  is a submersion [HN23, Def. 6.1]. By Thom's first isotopy lemma [Mat12, Prop. 11.1], the topology of the fibres of  $\pi$  is constant on each stratum  $S_Z$ . In particular, the Euler characteristic is constant. It follows that the Whitney stratification of  $\pi$  is an Euler stratification. The converse is not true: a priori, Euler strata are unions of Whitney strata. We provide examples in the projective and very affine case.

**Example 3.2.9.** Consider the following family of cuspidal cubic plane curves:

$$\mathcal{X}_F^c = \left\{ (x,z) \in \mathbb{P}^2 \times \mathbb{P}^1 : z_0 x_0^3 + z_1 x_1^2 x_2 = 0 \right\}.$$

For  $z \notin \{(1 : 0), (0 : 1)\}$ , the fibre  $(\pi_F^c)^{-1}(z)$  has a cusp singularity at the origin, implying  $\chi((\pi_F^c)^{-1}(z)) = 2$  (see the paragraph on multisingularity strata and in particular Example 3.2.12 below). For  $z = (1 : 0)$ , the cubic degenerates into a triple line, also with Euler characteristic 2. Thus, an Euler stratification for  $\pi_F^c$  is given by  $\{(0 : 1)\} \sqcup (\mathbb{P}^1 \setminus \{(0 : 1)\})$ , whereas  $\{(1 : 0)\}$  necessarily forms a separate Whitney stratum.  $\diamond$

**Example 3.2.10.** Consider the family of very affine curve complements

$$(\mathcal{X}_f^*)^c = \left\{ (x, z) \in T \times \mathbb{P}^1 : (x_1 + x_2)(z_0 x_1 + z_1 x_2) = 0 \right\}.$$

Here,  $T$  is the torus  $(\mathbb{C}^*)^2 \cong T \subset \mathbb{P}^2$ . For  $z \notin \{(1 : 0), (1 : 1), (0 : 1)\}$ , the fibre of  $(\mathcal{X}_f^*)^c \rightarrow \mathbb{P}^1$  consists of two lines intersecting at the origin. For  $z \in \{(1 : 0), (1 : 1), (0 : 1)\}$ , the fibre consists of only one line through the origin. The Euler characteristic is always zero, hence there is only a single Euler stratum. However, the topological types are different.  $\diamond$

**Multisingularity strata.** As seen in previous examples, Euler stratifications are closely related to the singularity structure of  $\mathcal{X}$ . In the following we recall some basic notions of singularity theory following [GLS07] and elaborate on the relation to Euler stratifications.

Let  $U \subseteq \mathbb{C}^d$  be an analytic open subset,  $f: U \rightarrow \mathbb{C}$  a holomorphic function and  $X = \mathcal{V}(f) \subset U$  the hypersurface defined by  $f$ . The set of singular points of  $X$  is given by

$$X_{\text{sing}} = \left\{ x \in U : f(x) = \frac{\partial f}{\partial x_1}(x) = \dots = \frac{\partial f}{\partial x_d}(x) = 0 \right\}.$$

A point  $x \in X_{\text{sing}}$  is called an isolated singularity if there exists a neighbourhood  $V$  of  $x$  such that  $(X_{\text{sing}} \cap V) \setminus \{x\} = \emptyset$ . In singularity theory, isolated singularities are typically studied up to right equivalence (analytic change of coordinates) or contact equivalence (isomorphism of factor algebras). Let  $J(f)$  denote the ideal sheaf  $\langle \frac{\partial f}{\partial x_1}, \dots, \frac{\partial f}{\partial x_d} \rangle \cdot \mathcal{O}(U)$  and define the algebra  $M_{f,x} := \mathcal{O}_{\mathbb{C}^d,x}/J(f)\mathcal{O}_{\mathbb{C}^d,x}$ , called the *Milnor algebra* of  $f$  at  $x$ . Its dimension as a  $\mathbb{C}$ -vector space is the *Milnor number* of  $f$  at  $x$ . We write  $\mu(f, x) = \dim_{\mathbb{C}} M_{f,x}$ . This number is a topological invariant of the singularity and plays a crucial role in our considerations. Note that  $\mu(f, x) > 0$  if and only if  $x$  is a singular point of  $f$  and that  $\mu(f, x)$  is finite by an application of the Hilbert–Rückert Nullstellensatz. If  $X$  has only isolated singularities  $y_1, \dots, y_m$ , then we define the *total Milnor number* to be  $\mu(X) := \sum_{i=1}^m \mu(f, y_i)$ . The connection to Euler stratifications is established by the following statement.

**Proposition 3.2.11** ([Par88, Cor. 1.7]). *Let  $M$  be a  $d$ -dimensional smooth complex projective variety and let  $\mathcal{L}$  be a line bundle on  $M$ . For two sections  $s_1, s_2 \in H^0(M, \mathcal{L})$ , let  $X_1 = \mathcal{V}(s_1)$  and  $X_2 = \mathcal{V}(s_2)$  denote their respective zero loci. If  $X_1$  and  $X_2$  have only isolated singularities, then we have the equality  $\mu(X_1) - \mu(X_2) = (-1)^d(\chi(X_1) - \chi(X_2))$ .*

**Example 3.2.12.** A smooth cubic plane curve has Euler characteristic zero by the genus-degree formula. A cusp singularity has Milnor number two. Therefore, a cuspidal plane cubic has Euler characteristic two. The full stratification of plane cubics is shown in Figure 3.3. Each stratum of  $\mathcal{X}_F^c \rightarrow Z = \mathbb{P}^9 \cong \text{Proj}(\mathbb{C}[x_0, x_1, x_2]_3)$  corresponds to a circle in the figure. Strata with isolated singularities are labelled by their singularity type. Below each circle, we record the Euler characteristic of  $\mathcal{X}_{F,z}^c$  on the stratum. The edges between strata indicate the poset relation. The Euler characteristic of a fibre  $\mathcal{X}_{F,z}^c$  is invariant under the action of  $\text{GL}(3, \mathbb{C})$  on ternary cubics. The ring of invariants is generated by  $I_4$  and  $I_6$ , which are homogeneous polynomials of degree four, respectively six, in the ten coefficients  $z_0, \dots, z_9$ . These generators are unique up to scaling, and  $I_4$  is called the *Aronhold invariant* [MS21, Ex. 11.12]. The closed stratum of nodal cubics ( $A_1$ ) is defined by the discriminant  $I_4^3 - I_6^2$ , and the cuspidal cubics ( $A_2$ ) are defined by  $I_4 = I_6 = 0$ .  $\diamond$

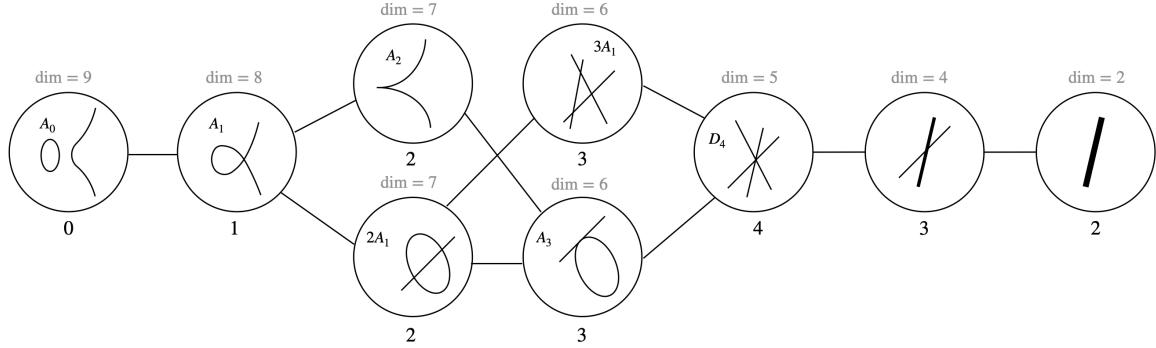


Figure 3.3: The Euler stratification of plane cubics.

This shows that for families with only isolated singularities, any Euler stratum is a union of *multisingularity strata*, i.e. loci where the general member of the hypersurface family has prescribed topological types of singularities. However, multisingularity strata are extremely hard to compute. Even for the case of plane curves, a complete description of multisingularity strata are out of reach [DH88]. In [Est17], Esterov provides formulae for the tropicalisation of the strata for two  $A_1$ -singularities and one  $A_2$ -singularity.

### 3.2.2. Points on the line

This subsection studies the case  $d = 1$ , corresponding to points on the projective line. In this setting, it is possible to enumerate and parametrise all Euler strata. However, finding implicit equations is challenging. Let  $n \in \mathbb{Z}_{>0}$ ; simplifying notation, we set

$$\mathcal{X}^c = \left\{ (x, z) \in \mathbb{P}^1 \times \mathbb{P}^n : z_0 x_0^n + z_1 x_0^{n-1} x_1 + \cdots + z_n x_1^n = 0 \right\}. \quad (3.2.10)$$

We study the family given by the projection  $\pi: \mathcal{X}^c \rightarrow Z$ . The topological Euler characteristic of a fibre  $\mathcal{X}_z^c = \pi^{-1}(z)$  is equal to the number of points of  $\mathcal{X}_z^c$ , ignoring multiplicities. Hence, Euler strata are loci where certain roots of  $F = z_0 x_0^n + \cdots + z_n x_1^n$  coincide. We label these strata by integer partitions of  $n$  as follows. Let  $\lambda = (\lambda_1, \dots, \lambda_k)$  be such a partition. For any  $k > 0$ , the set  $Z_k = \{z \in Z : \chi(\mathcal{X}_z^c) = k\}$  is a disjoint union of strata  $Z_k = \coprod_{|\lambda|=k} S_\lambda$  ranging over all partitions of  $n$  with length  $k$ . Here,  $S_\lambda$  denotes the set

$$S_\lambda = \{z \in \mathbb{P}^n : \mathcal{X}_z^c \text{ has } k \text{ distinct points with multiplicities } \lambda_1, \dots, \lambda_k\}. \quad (3.2.11)$$

These are called *coincident* or *multiple root loci*, or *Brill–Gordan loci* in the literature, see e.g. [Chi03, FNR06, Kur12, LS16, Wey89]. In the following we review some important results.

Let  $\prec$  denote the partial order on the set of partitions given by refinement. That is,  $\lambda = (\lambda_1, \dots, \lambda_k)$  refines  $\mu = (\mu_1, \dots, \mu_l)$ , written  $\lambda \prec \mu$ , if and only if there exists a partition  $(I_1, \dots, I_l)$  of  $[k]$  such that for any  $1 \leq i \leq l$ :  $\mu_i = \sum_{j \in I_i} \lambda_j$ . An example of the resulting lattice for partitions of five can be seen in Figure 3.4. The partitions are labelled by Young diagrams, and by strings like  $21^3$ , representing  $\lambda = (2, 1, 1, 1)$ . Each circle represents an irreducible Euler stratum of  $\pi: \mathcal{X}^c \subset \mathbb{P}^1 \times \mathbb{P}^5 \rightarrow \mathbb{P}^5$ . The following theorem is well-known.

**Theorem 3.2.13.** *Let  $\mathcal{X}^c$  be as in (3.2.10) and consider the projection  $\pi: \mathcal{X}^c \rightarrow Z = \mathbb{P}^n$ .*

- (i) *The set of coincident root loci  $\{S_\lambda\}_\lambda$  ranging over all partitions  $\lambda$  of  $n$ , ordered by refinement so that  $S_\lambda \prec S_{\lambda'} \Leftrightarrow \lambda' \prec \lambda$ , forms an Euler stratification of  $\pi$ .*

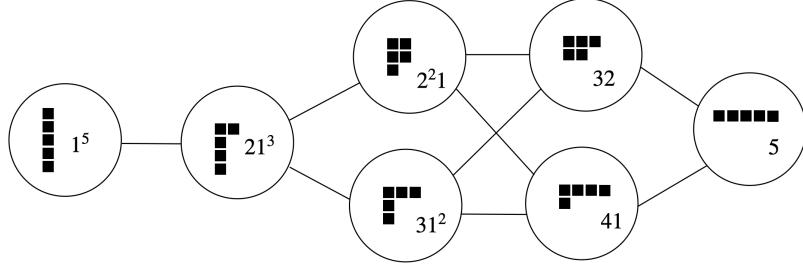


Figure 3.4: Young diagrams with five boxes index the strata of  $\pi: X^c \subset \mathbb{P}^1 \times \mathbb{P}^5 \rightarrow \mathbb{P}^5$ .

- (ii) The Zariski closure  $\nabla_\lambda$  of  $S_\lambda$  is an irreducible projective variety. It equals the disjoint union  $\nabla_\lambda = \coprod_{\lambda \prec \lambda'} S_{\lambda'}$ , where  $\prec$  denotes refinement of partitions.
- (iii) For  $z \in S_\lambda$ , the Euler characteristic of the fibre is  $\chi(\mathcal{X}_z^c) = |\lambda|$ . This equals the dimension  $\dim S_\lambda = \dim \nabla_\lambda = |\lambda|$  and the height of the corresponding Young diagram.

Hilbert proved in [Hil87] that the degree of the  $k$ -dimensional projective variety  $\nabla_\lambda$  is

$$\deg(\nabla_\lambda) = \frac{k!}{m_1! m_2! \dots m_n!} \cdot \lambda_1 \lambda_2 \dots \lambda_k,$$

where  $m_j := \#\{i: \lambda_i = j\}$ . In general, not much is known about the structure of the ideal  $I(\nabla_\lambda)$  defining  $\nabla_\lambda$ . When  $\lambda$  is of the form  $\lambda = (1^{n-a}, a)$  for  $a \geq \lfloor n/2 \rfloor + 2$ , Weyman showed that  $I(\nabla_\lambda)$  is generated in degree  $\leq 4$  [Wey89]. Moreover, for  $|\lambda| = 2$ , Abdesselam and Chipalkatti proved that  $I(\nabla_\lambda)$  is generated in degree  $\leq 4$  [AC06, Prop. 20].

Chipalkatti describes a generating set for  $I(\nabla_\lambda)$  in terms of covariant forms [Chi04]. We can deduce an upper bound on the number of generators of  $I(\nabla_\lambda)$  using [Chi04, Thm. 3.5 & Prop. 3.6] as follows: Let  $\mathcal{C}_\lambda$  be the set of all  $n$ -partitions  $\mu \not\prec \lambda$  that are minimal with this property, i.e.  $\mathcal{C}_\lambda = \{\mu \not\prec \lambda: \nu \prec \mu \Rightarrow \nu \prec \lambda\}$ . Then  $I(\nabla_\lambda)$  can be generated by  $n! \cdot |\mathcal{C}_\lambda|$  many polynomials. For example, consider the partition  $\lambda = (3, 2)$ . Then  $\mathcal{C}_\lambda = \{(4, 1)\}$  and hence  $I(\nabla_\lambda)$  can be generated by  $\leq 5!$  polynomials. This bound is not optimal, as  $I(\nabla_\lambda)$  is minimally generated by 28 polynomials, see [LS16, Tbl. 1].

Computing defining equations for  $\nabla_\lambda$  quickly becomes a challenging task. For example,  $I(\nabla_{2^{1^3}})$  (appearing in Figure 3.4) is the zero locus of the discriminant of the binary quintic, which is a degree eight polynomial with 59 terms:

$$z_1^2 z_2^2 z_3^2 z_4^2 - 4 z_0 z_2^3 z_3^2 z_4^2 - 4 z_1^3 z_3^3 z_4^2 + 18 z_0 z_1 z_2 z_3^3 z_4^2 - 27 z_0^2 z_3^4 z_4^2 - 4 z_1^2 z_2^3 z_4^3 + \dots + 3125 z_0^4 z_5^4.$$

In [LS16], the authors present a table with the degrees of the minimal generators for the ideals  $I(\nabla_\lambda)$  for  $n \leq 7$ . This can be done efficiently (and heuristically) by using finite field computations. In Section 3.2.6, we extend these results to  $n = 8$ . Moreover, we compute all generators for  $n = 1, \dots, 8$  over  $\mathbb{Q}$ . For this, we use the fact that  $S_\lambda$  is parametrised by

$$(\mathbb{P}^1)^k \dashrightarrow S_\lambda, \quad ((a_1 : b_1), \dots, (a_k : b_k)) \longmapsto (a_1 x_0 + b_1 x_1)^{\lambda_1} \cdots (a_k x_0 + b_k x_1)^{\lambda_k}. \quad (3.2.12)$$

As indicated earlier, applications in statistics and physics require studying Euler stratifications of families of very affine hypersurfaces in the algebraic torus. We modify the previous constructions by disregarding points at zero and infinity. That is, we consider

$$(\mathcal{X}^*)^c = \{(t, z) \in \mathbb{C}^* \times \mathbb{P}^n : z_0 + z_1 t + \dots + z_n t^n = 0\}, \quad (3.2.13)$$

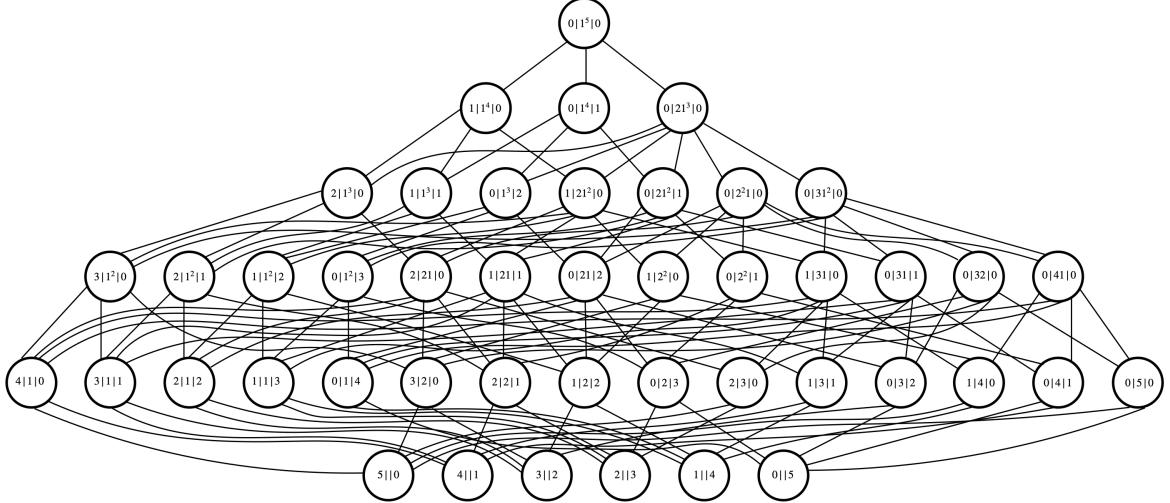


Figure 3.5: The Euler stratification of five points in  $\mathbb{C}^*$ .

with projection  $\pi: (\mathcal{X}^*)^c \rightarrow Z = \mathbb{P}^n$ . Notice that, while previously all Euler strata were necessarily stable under the action of  $\mathrm{GL}(2, \mathbb{C})$  on binary forms of degree  $n$ , this symmetry is now broken. This leads to a significant increase in the number of strata.

Strata are now labelled by two nonnegative integers  $m_0, m_\infty$  such that  $m_0 + m_\infty \leq n$ , and a partition of  $n - m_0 - m_\infty$ . For instance, the stratum corresponding to  $m_0 = 0, m_\infty = 2$  and the partition 21 of 3 consists of binary quintics with a root of multiplicity two at infinity, and two roots in  $\mathbb{C}^*$ , one of which has multiplicity two and one of which is simple. In Figure 3.5, we label such a stratum for  $n = 5$  by the corresponding partition, with the integers  $m_0$  and  $m_\infty$  written to the left, respectively to the right of it. The above stratum is 0|21|2. The poset from Figure 3.4 appears on the right side of Figure 3.5, considering only strata with  $m_0 = m_\infty = 0$ .

Much like the strata  $S_\lambda$  in (3.2.12), the stratum  $S_{m_0|\lambda|m_\infty}$  is parametrised by

$$(\mathbb{P}^1)^k \dashrightarrow S_{m_0|\lambda|m_\infty}, \quad ((a_1 : b_1), \dots, (a_k : b_k)) \mapsto x_1^{m_0} (a_1 x_0 + b_1 x_1)^{\lambda_1} \cdots (a_k x_0 + b_k x_1)^{\lambda_k} x_0^{m_\infty}.$$

Equations for its closure  $\nabla_{m_0|\lambda|m_\infty}$  are thus easily deduced from equations for  $\nabla_\lambda$ . To define the partial ordering of the strata  $S_{m_0|\lambda|m_\infty}$ , let  $(m_0, \lambda, m_\infty)$  be the unique partition of  $n$  for which  $S_{m_0|\lambda|m_\infty} \subseteq S_{(m_0, \lambda, m_\infty)}$ . The following definition of “ $\prec$ ” is natural:

$$m_0|\lambda|m_\infty \prec m'_0|\lambda'|m'_\infty \iff (m_0, \lambda, m_\infty) \prec (m'_0, \lambda, m'_\infty), \quad m_0 \leq m'_0 \text{ and } m_\infty \leq m'_\infty. \quad (3.2.14)$$

**Theorem 3.2.14.** Let  $(\mathcal{X}^*)^c$  be as in (3.2.13) and consider the projection  $\pi: (\mathcal{X}^*)^c \rightarrow Z = \mathbb{P}^n$ .

- (i) The set of strata  $\{S_{m_0|\lambda|m_\infty}\}$  ranging over all  $m_0 + m_\infty \leq n$  and all partitions  $\lambda$  of  $n - m_0 - m_\infty$ , ordered by (3.2.14), forms an Euler stratification of  $\pi$ .
- (ii) The Zariski closure  $\nabla_{m_0|\lambda|m_\infty}$  of  $S_{m_0|\lambda|m_\infty}$  is an irreducible projective variety. It equals the disjoint union  $\nabla_{m_0|\lambda|m_\infty} = \coprod_{m_0|\lambda|m_\infty \prec m'_0|\lambda'|m'_\infty} S_{m'_0|\lambda'|m'_\infty}$ .
- (iii) For  $z \in S_{m_0|\lambda|m_\infty}$ , the Euler characteristic of the fibre is  $\chi((\mathcal{X}^*)^c_z) = |\lambda|$ . This equals the dimension  $\dim S_{m_0|\lambda|m_\infty} = \dim \nabla_{m_0|\lambda|m_\infty} = |\lambda|$ .

As a consequence of Theorem 3.2.14, the number of strata is  $\sum_{i=0}^n (n+1-i)\mathcal{P}(i)$ , where  $\mathcal{P}(i)$  is the number of partitions of  $i$ . This is large compared to the  $\mathcal{P}(n)$  strata for (3.2.10).

---

**Algorithm 2** Euler stratification of  $X \cap H_z$ 


---

**Input:** homogeneous generators  $f_1, \dots, f_\ell \in \mathbb{Q}[x_0, \dots, x_N]$  of the vanishing ideal of  $X = \mathcal{V}(f_1, \dots, f_\ell)$  satisfying the assumptions of Proposition 3.2.15 and an integer  $1 \leq k \leq \deg X$

**Output:** ideal  $I \in \mathbb{Q}[z_0, \dots, z_N]$  such that:  $z \in \mathcal{V}(I)$  generic  $\Rightarrow \chi(X \cap H_z) = \deg X - k$

- 1:  $\Delta_X \leftarrow$  defining equation of  $X^\vee$  in  $\mathbb{Q}[X_0, \dots, X_N]$
  - 2:  $\mathfrak{m} \leftarrow$  maximal ideal generated by  $\langle X_0 - z_0, \dots, X_N - z_N \rangle \subset \mathbb{Q}(z_0, \dots, z_N)[X_0, \dots, X_N]$
  - 3:  $Q \leftarrow$  quotient ring  $\mathbb{Q}(z_0, \dots, z_N)[X_0, \dots, X_N]/\mathfrak{m}^k$
  - 4:  $M \leftarrow$  companion matrix representing multiplication by  $\Delta_X$  in  $Q$
  - 5: **return** ideal generated by entries of  $M$
- 

To end the subsection, we replace  $\mathbb{P}^1$  by a smooth projective curve  $X \subset \mathbb{P}^N$  and consider families of points arising as the intersection of  $X$  with a hyperplane  $H_z$ . Here, the index  $z$  stands for the coefficients of the defining linear equation of  $H$ :  $z \in Z = (\mathbb{P}^N)^\vee$ . Our goal is to compute the Euler stratification of  $Z$ , i.e. the loci in  $Z$  where  $X \cap H_z$  consists of a constant number of points. Note that if  $X \subset \mathbb{P}^n$  is the rational normal curve of degree  $n$ , then we recover the Euler stratification of (3.2.10). Our algorithm is deduced from the following expression for the Euler characteristic.

**Proposition 3.2.15.** *Let  $X$  be a one-dimensional smooth projective variety such that the dual variety  $X^\vee$  is a hypersurface. Let  $H \in X^\vee$  be such that  $(X \cap H)_{\text{sing}}$  is finite. Then*

$$\chi(X \cap H) = \deg X - \text{mult}_H X^\vee.$$

Here,  $\text{mult}_H X^\vee$  denotes the multiplicity of  $H \in X^\vee$ . Proposition 3.2.15 is a consequence of Theorem 3.2.16, which summarises [Tev06, Thm. 10.8 & Thm. 10.9].

**Theorem 3.2.16.** *Let  $X \subset \mathbb{P}^N$  be a smooth  $d$ -dimensional projective variety such that  $X^\vee$  is a hypersurface. Let  $H \in X^\vee$ , and let  $\mathcal{L}$  be a line bundle on  $X$ .*

(i) *If  $(X \cap H)_{\text{sing}}$  is finite, then the multiplicity of  $X^\vee$  at  $H$  is given by*

$$\text{mult}_H X^\vee = \sum_{p \in (X \cap H)_{\text{sing}}} \mu(X \cap H, p).$$

(ii) *For a global section  $s \in H^0(X, \mathcal{L}) \setminus \{0\}$  of  $\mathcal{L}$  with zero locus  $V \subset X$ , we have*

$$\mu(V) = (-1)^d(\chi(V) - \chi(X, \mathcal{L})),$$

where  $\chi(X, \mathcal{L})$  is the Euler characteristic of the zero locus of a generic section.

*Proof of Proposition 3.2.15.* Choose  $s \in \mathcal{O}_X(1)$  so that  $X \cap H = s^{-1}(0) = V$ . For  $X$  one-dimensional, we have  $\chi(X, \mathcal{O}_X(1)) = \deg X$  and, using Theorem 3.2.16(ii), we obtain

$$\mu(X \cap H) = \sum_{p \in (X \cap H)_{\text{sing}}} \mu(X \cap H, p) = \deg X - \chi(X \cap H). \quad (3.2.15)$$

Combining (3.2.15) with Theorem 3.2.16(i), we arrive at the desired statement.  $\square$

Proposition 3.2.15 gives rise to Algorithm 2, which we illustrate in the following.

**Example 3.2.17.** Let  $X = \mathcal{V}(f) \subset \mathbb{P}^2$  be the elliptic curve defined by  $f = x_0^3 - x_0 x_2^2 - x_1^2 x_2 + x_2^3$ . The output of Algorithm 2 for  $k = 1$  is the defining equation of the dual curve  $X^\vee$ :

$$4z_0^6 + 4z_0^5 z_2 + z_0^4 z_1^2 + 36z_0^3 z_1^2 z_2 - 4z_0^3 z_2^3 + 18z_0^2 z_1^2 + 30z_0^2 z_1^2 z_2^2 + 24z_0 z_1^4 z_2 - 23z_1^6 + 54z_1^4 z_2^2 - 27z_1^2 z_2^4.$$

This vanishes on  $z \in (\mathbb{P}^2)^\vee$  for which  $H_z = \{z_0x_0 + z_1x_1 + z_2x_2 = 0\}$  intersects  $X$  in at most two points. For  $k = 2$ , the algorithm returns an ideal whose radical is generated by

$$z_0^3 - 9z_0z_1^2 - 3z_0z_2^2 - 6z_1^2z_2, \quad z_0^2z_2^2 - 50z_0z_1^2z_2 - 18z_0z_2^3 + 28z_1^4 - 63z_1^2z_2^2, \\ 7z_0^2z_1z_2 + 56z_0z_1^3 + 18z_0z_1z_2^2 + 45z_1^3z_2 - 9z_1z_2^3, \quad 28z_0^2z_1^2 + 9z_0^2z_2^2 + 54z_0z_1^2z_2 + 6z_0z_2^3 + 21z_1^2z_2^2.$$

This defines nine points  $z \in (\mathbb{P}^2)^\vee$  for which  $H_z$  intersects  $X$  in a single point.  $\diamond$

### 3.2.3. Projective hypersurfaces

We switch to the hypersurface setting. Throughout the subsection,  $F(x, z) \in \mathbb{C}[z][x]$  is a bihomogeneous polynomial in variables  $x = (x_0, \dots, x_d)$  and parameters  $z = (z_0, \dots, z_m)$ . We develop an algorithm for computing the Euler stratification of the family

$$\pi_F : \mathcal{X}_F \rightarrow Z, \quad \text{where } \mathcal{X}_F = \{(x, z) \in \mathbb{P}^d \times Z : F(x, z) \neq 0\}. \quad (3.2.16)$$

For simplicity, we assume that  $Z$  is a closed irreducible subvariety of  $\mathbb{P}^m$ . The first task is to determine for which  $z \in Z$  the Euler characteristic of  $\mathcal{X}_{F,z} = \pi_F^{-1}(z)$  differs from the generic value. This is made precise by the *Euler discriminant* of the family. We point out that this term was coined by Esterov in [Est13, §1.2] to mean something slightly different.

**Definition 3.2.18.** By Proposition 3.2.6, there exists an integer  $\chi^*$  (the generic Euler characteristic) and a dense open Euler stratum  $S^* \subseteq Z$  such that  $\chi(\mathcal{X}_{F,z}) = \chi^*$  for all  $z \in S^*$ . The *Euler discriminant variety* of  $\pi_F : \mathcal{X}_F \rightarrow Z$  is the Zariski closure of  $Z \setminus S^*$  in  $Z$ . We denote this variety by  $\nabla_\chi(\pi_F)$ , or simply  $\nabla_\chi$ .

---

#### Algorithm 3 Euler stratification of (3.2.16)

---

**Input:** bihomogeneous polynomial  $F \in \mathbb{Q}[x, z]$  and the defining ideal  $I_Z$  of  $Z$   
**Output:** prime ideals of closed strata in the Euler stratification of  $\pi_F : \mathcal{X}_F \rightarrow Z$

```

1: strata0 ← { I_Z }
2: $i \leftarrow 0$
3: while strata i ≠ { $\mathbb{Q}[z]$ } do ▷ iteration ends when computed strata are empty
4: $i \leftarrow i + 1$
5: strata i ← {} ▷ initialise an empty list of strata
6: for $I \in \text{strata}_{i-1}$ do
7: $I(\nabla_\chi) \leftarrow$ Euler discriminant ideal of $\mathcal{X}_F \cap (\mathbb{P}^d \times \mathcal{V}(I))$
8: strata i ← strata i ∪ {minimal primes of $I(\nabla_\chi)$ }
9: end for
10: end while
11: return $\bigcup_i \text{strata}_i$

```

---

Computing an Euler stratification comes down to an iterated computation of Euler discriminants. This is clarified by Algorithm 3, which uses the notation  $I_Z = I(Z)$  for the vanishing ideal of  $Z \subseteq \mathbb{P}^m$ . The algorithm computes the Euler discriminant for  $z \in \mathcal{V}(I_Z) = Z$  first. After that, it restricts the family to each of the irreducible components of  $\nabla_\chi$  to find deeper strata, and so on. Notice that one can easily adapt the algorithm to keep track of the partial order relations between strata. The important step in the algorithm is Line 8, the computation of  $\nabla_\chi$ . The rest of the subsection is devoted to that step.

Our algorithm uses a formula from [Huh12] which expresses  $\chi(D(F))$  in terms of the cohomology class of the graph of the Gauss map  $\nabla F$ . This is the closure  $\Gamma = \overline{\Gamma^\circ}$  of

$$\Gamma^\circ = \{(x, y) \in D(F) \times (\mathbb{P}^d)^\vee : y = \nabla F(x)\} \quad (3.2.17)$$

in  $\mathbb{P}^d \times (\mathbb{P}^d)^\vee$ . In what follows, let  $H_k, H_k^\vee$  be generic  $k$ -planes in  $\mathbb{P}^d$  and  $(\mathbb{P}^d)^\vee$  respectively, and let  $[H_k \times H_\ell^\vee]$  be the generators of the Chow ring of  $\mathbb{P}^d \times (\mathbb{P}^d)^\vee$ . We state a version of Huh's result [Huh12, Thm. 9] and outline a geometric proof.

**Theorem 3.2.19.** *Let  $F \in \mathbb{C}[x_0, \dots, x_d]$  be any non-constant homogeneous polynomial, and let  $D(F) = \{x \in \mathbb{P}^d : F(x) \neq 0\}$  be its hypersurface complement. Suppose that*

$$[\Gamma] = \sum_{i=0}^d e_i [H_{d-i} \times H_i^\vee] \in A_d(\mathbb{P}^d \times (\mathbb{P}^d)^\vee)$$

is the class of  $\Gamma$  from (3.2.17) in the Chow ring of  $\mathbb{P}^d \times (\mathbb{P}^d)^\vee$ . Setting  $H_{-1} = \emptyset$ , we have

$$\chi(D(F)) = \sum_{i=0}^d \chi((D(F) \cap H_i) \setminus (D(F) \cap H_{i-1})) = \sum_{i=0}^d (-1)^i \deg \nabla F|_{H_i} = \sum_{i=0}^d (-1)^i e_i.$$

*Proof.* The first equality follows from excision. The second equality is Theorem 3.2.7:

$$\deg \nabla F|_{H_i} = (-1)^i \chi((D(F) \cap H_i) \setminus (D(F) \cap H_{i-1})).$$

It remains to show that  $\deg \nabla F|_{H_i}$  equals  $e_i$ . That is,  $\deg \nabla F|_{H_i}$  equals the number of intersection points in  $\Gamma \cap (H_i \times H_{d-i}^\vee)$ . The linear space  $H_i$  can be represented as the projectivised row span of an  $(i+1) \times (d+1)$  matrix  $\mathbb{L}_i$ . Let  $\tilde{x}_0, \dots, \tilde{x}_i$  be coordinates on  $H_i$ , so that  $x = \mathbb{L}_i^t \cdot \tilde{x}$ . We set  $F|_{H_i} = F(\mathbb{L}_i^t \cdot \tilde{x})$ , and the chain rule implies that

$$\nabla F|_{H_i}(\tilde{x}) = \left( \frac{\partial F|_{H_i}}{\partial \tilde{x}_0} : \dots : \frac{\partial F|_{H_i}}{\partial \tilde{x}_i} \right) = \mathbb{L}_i \left( \left( \frac{\partial F}{\partial x_0} : \dots : \frac{\partial F}{\partial x_d} \right) \Big|_{x=\mathbb{L}_i^t \cdot \tilde{x}} \right).$$

Here,  $\mathbb{L}_i(y)$  denotes the projection  $\mathbb{L}_i : (\mathbb{P}^d)^\vee \dashrightarrow (\mathbb{P}^i)^\vee$  represented by the matrix  $\mathbb{L}_i$ . This establishes a bijection between the set  $\{\tilde{x} \in D(F) \cap H_i : \nabla F|_{H_i}(\tilde{x}) = \tilde{y}\}$  and

$$\{(x, y) \in \Gamma : x \in H_i, \mathbb{L}_i(y) = \tilde{y}\} = \Gamma \cap (H_i \times H_{d-i}^\vee)$$

for generic  $\tilde{y} \in (\mathbb{P}^i)^\vee$ . The condition  $\mathbb{L}_i(y) = \tilde{y}$  determines a  $(d-i)$ -dimensional linear space  $H_{d-i}^\vee$  in which  $y$  is contained. The intersection  $\Gamma \cap (H_i \times H_{d-i}^\vee)$  is transverse, so that

$$\deg \nabla F|_{H_i} = |\{\tilde{x} \in D(F) \cap H_i : \nabla F|_{H_i}(\tilde{x}) = \tilde{y}\}| = |\Gamma \cap (H_i \times H_{d-i}^\vee)| = e_i. \quad \square$$

The numbers  $e_i$  can be read from the coefficients of the (bigraded) Hilbert polynomial of  $\Gamma$ , see [Huh12, §2]. They can also be computed numerically, without computing the ideal of  $\Gamma$ , by intersecting  $\Gamma^\circ$  from (3.2.17) with products of random linear spaces  $H_i \times H_{d-i}^\vee$ . The number of intersection points is  $e_i$ . Alternatively, one computes the cardinality of a generic fibre of  $\nabla F|_{H_i} : D(F) \cap H_i \rightarrow \mathbb{P}^i$ . We have implemented these approaches in Julia. For this, we rely on the numerical homotopy continuation algorithms implemented in [BT18].

**Example 3.2.20.** We consider the cuspidal plane curves from Example 3.2.9. For  $z = (1 : 1)$ , our implementation computes the following information, which is easily verifiable:

$$e_0 = \deg \nabla F|_{H_0} = 1, \quad e_1 = \deg \nabla F|_{H_1} = 2, \quad e_2 = \deg \nabla F|_{H_2} = 2. \quad \diamond$$

By Theorem 3.2.19, the Euler discriminant  $\nabla_\chi(\pi_F)$  is contained in the union of loci where the degrees of the restrictions of the Gauss map  $\nabla F(x, z)|_{H_i}$  drop. Let  $n^*$  be the

degree of  $\nabla F(x, z)$  for generic  $z \in Z$ . We define the *polar discriminant* of a nonzero homogeneous polynomial  $F \in \mathbb{C}[z][x]$  as the Zariski closure of

$$\{z \in Z : \deg \nabla F(x, z) < n^*\} \subset Z.$$

We now propose a randomised algorithm for computing the polar discriminant.

Let  $b \in (\mathbb{P}^d)^\vee$  be a generic point. The degree of  $\nabla F$  is the number of points  $x \in D(F)$  satisfying  $\nabla F(x) = b$ . Adding parameters  $z \in Z$ , we consider the incidence variety

$$\mathcal{Y}_b^\circ = \left\{ (x, z) \in \mathcal{X}_F : \text{rank} \begin{pmatrix} \frac{\partial F(x, z)}{\partial x_0} & \frac{\partial F(x, z)}{\partial x_1} & \dots & \frac{\partial F(x, z)}{\partial x_d} \\ b_0 & b_1 & \dots & b_d \end{pmatrix} \leq 1 \right\}. \quad (3.2.18)$$

Here,  $\mathcal{X}_F$  is as in (3.2.16). The closure of  $\mathcal{Y}_b^\circ$  in  $\mathbb{P}^d \times Z$  is denoted by  $\mathcal{Y}_b$ . By assumption, the projection  $\pi_Z : \mathcal{Y}_b \rightarrow Z$  has generically finite fibres of cardinality  $n^*$ . The polar discriminant is contained in the image of  $\mathcal{Y}_b \setminus \mathcal{Y}_b^\circ = \mathcal{Y}_b \cap \mathcal{V}(F(x, z))$  under that projection.

**Proposition 3.2.21.** *Let  $\mathcal{X}_F$  be the family from (3.2.16), with  $Z \subseteq \mathbb{P}^m$  any irreducible quasi-projective variety. Let  $\mathcal{Y}_b^\circ$  be as in (3.2.18), where  $b \in \mathbb{P}^d$  is generic. The variety  $\mathcal{Y}_b \cap \mathcal{V}(F)$  has dimension at most  $\dim Z - 1$ . Hence,  $\pi_Z(\mathcal{Y}_b \cap \mathcal{V}(F)) \subsetneq Z$  is a strict containment.*

*Proof.* The following incidence variety is irreducible of dimension  $d + \dim Z$ :

$$\mathcal{Y}^\circ = \{(x, z, b) \in \mathcal{X}_F \times \mathbb{P}_b^d : \nabla F(x, z) = b\}. \quad (3.2.19)$$

This is seen from the obvious parametrisation  $\mathcal{X}_F \rightarrow \mathcal{Y}^\circ$ . A general fibre of the dominant map  $\mathcal{Y}^\circ \rightarrow \mathbb{P}_b^d$  is of pure dimension  $\dim Z$ . That fibre is  $\mathcal{Y}_b^\circ$ , so its boundary in the closure  $\mathcal{Y}_b = \overline{\mathcal{Y}_b^\circ}$  has dimension at most  $\dim \mathcal{Y}_b - 1 = \dim Z - 1$ .  $\square$

The containment of the polar discriminant in  $\pi_Z(\mathcal{Y}_b \cap \mathcal{V}(F))$  might be strict. For instance, there might be components which depend on the specific choice of  $b$ , see Example 3.2.22. We can discard such spurious loci by repeating this procedure for several values of  $b$  and intersecting the results. Here,  $\dim Z + 1 \leq m + 1$  random choices for  $b$  suffice. The discussion is summarised in Algorithm 4. Lines 3 and 4 can be executed using standard Gröbner basis techniques.

---

#### Algorithm 4 Polar discriminant of $\nabla F(x, z)$

---

**Input:** bihomogeneous polynomial  $F \in \mathbb{Q}[z][x]$  and the defining ideal  $I_Z$  of  $Z$

**Output:** minimal primes of the ideal defining the polar discriminant of  $F$

- 1: **for**  $i = 0, \dots, \dim Z$  **do**
  - 2:    $b_i \leftarrow$  a random point in  $(\mathbb{P}_\mathbb{Q}^d)^\vee$
  - 3:   compute the vanishing ideal of the closure  $\mathcal{Y}_{b_i}$  of  $\mathcal{Y}_{b_i}^\circ$  in (3.2.18)
  - 4:    $I_i \leftarrow$  the defining ideal of the projection  $\pi_Z(\mathcal{Y}_{b_i} \cap \mathcal{V}(F(x, z)))$
  - 5: **end for**
  - 6: **return** minimal primes of  $I_0 + I_1 + \dots + I_{\dim Z}$
- 

**Example 3.2.22.** The polar discriminant of the ternary quadric in (3.2.1) is (3.2.3). It is computed efficiently with our code, see [TW24b], choosing uniformly random integers  $b_i$  between  $-100$  and  $100$ , a total number of  $\dim Z + 1 = 6$  times. Each of the ideals  $I_i$  in Algorithm 4 decomposes as  $\langle \Delta \rangle \cap \langle G(z, b) \rangle$ , where  $G(z, b)$  defines a hypersurface in  $Z = \mathbb{P}^5$  which depends on  $b$ .  $\diamond$

---

**Algorithm 5** Upper bound for the Euler discriminant of  $\pi_F : \mathcal{X}_F \rightarrow Z$ 


---

**Input:** bihomogeneous polynomial  $F \in \mathbb{Q}[z][x]$  and the defining ideal  $I_Z$  of  $Z$   
**Output:** minimal prime ideals of a variety containing the Euler discriminant of  $\pi_F : \mathcal{X}_F \rightarrow Z$

- 1: **for**  $i = 0, \dots, d - 1$  **do**
- 2:    $I_i \leftarrow \langle 0 \rangle$
- 3:   **for**  $j = 0, \dots, \dim Z$  **do**
- 4:      $H_{i,j} \leftarrow$  random  $i$ -plane in  $\mathbb{P}^d$
- 5:      $I_i \leftarrow I_i + \text{ideal of the polar discriminant of } (\nabla F|_{H_{i,j}}, Z)$
- 6:   **end for**
- 7: **end for**
- 8:  $I_d \leftarrow \text{polar discriminant of } (\nabla F, Z)$
- 9: **return** minimal primes of  $I_0 \cap I_1 \cap \dots \cap I_d$

---

The polar discriminant characterises for which  $z \in Z$  the degree of  $\nabla F$  drops. By Theorem 3.2.19, for computing the Euler discriminant, we must also check the degrees of the restrictions  $\nabla F|_{H_i}$  for generic  $i$ -planes  $H_i \subseteq \mathbb{P}^d$ . The polar discriminant of  $\nabla F|_{H_i} : D(F) \cap H_i \rightarrow \mathbb{P}^i$  may depend on  $H_i$ . Similarly as above, we can eliminate unwanted contributions by computing the polar discriminant repeatedly, for  $m + 1$  choices of  $H_i$ . This happens in lines 1–7 in Algorithm 5. The union of all these polar discriminants may contain the Euler discriminant strictly (see Example 3.2.23). Nonetheless, by Proposition 3.2.21, using Algorithm 5 in line 7 of Algorithm 3 leads to a correct but possibly redundant Euler stratification.

**Example 3.2.23.** We have seen in Example 3.2.9 that the Euler discriminant for  $F(x, z) = z_0 x_0^3 + z_1 x_1^2 x_2$  is given by  $z_0 = 0$ . The fibre of  $\nabla F$  over  $(b_0 : b_1 : b_2)$  consists of two points

$$\left( \sqrt{\frac{b_0}{3z_0}} : \sqrt{\frac{b_2}{z_1}} : \frac{b_1}{2z_1} \sqrt{\frac{z_1}{b_2}} \right), \quad \left( \sqrt{\frac{b_0}{3z_0}} : -\sqrt{\frac{b_2}{z_1}} : -\frac{b_1}{2z_1} \sqrt{\frac{z_1}{b_2}} \right).$$

When  $(z_0 : z_1) \rightarrow (0 : 1)$ , these points approach  $(1 : 0 : 0) \times (0 : 1)$ . The point  $(1 : 0 : 0) \times (0 : 1)$  is indeed contained in  $\mathcal{Y}_b \cap \mathcal{V}(F(x, z))$ . A similar argument shows that  $(1 : 0)$  is also contained in the polar discriminant of  $\nabla F$ . However,  $(1 : 0)$  does not belong to the Euler discriminant: for  $z = (1 : 0)$ , the alternating sum  $\sum_{i=0}^2 (-1)^i e_i = 1$ , as for generic  $z$ .  $\diamond$

Example 3.2.23 raises the question to what extent the polar and Euler discriminant agree. For plane curve families with isolated singularities, it turns out that they coincide.

**Proposition 3.2.24.** Let  $F(x_0, x_1, x_2, z)$  be a squarefree bihomogeneous polynomial. If  $Z \subseteq \mathbb{P}^m$  is such that the degree of the (reduced) plane curve  $\mathcal{X}_{F,z}^c \subset \mathbb{P}^2$  is constant for all  $z$ , then the Euler discriminant of  $\mathcal{X}_F \rightarrow Z$  coincides with the polar discriminant of  $F$ .

*Proof.* It follows from Theorem 3.2.7 that for any  $z, z' \in Z$ , we have

$$\chi(\mathcal{X}_{F,z}) - \chi(\mathcal{X}_{F,z'}) = \deg \nabla F(x, z) - \deg \nabla F(x, z') - (\deg \mathcal{X}_{F,z}^c - \deg \mathcal{X}_{F,z'}^c). \quad \square$$

### 3.2.4. Very affine hypersurfaces

Our original motivation for studying Euler stratifications comes from statistics and particle physics. In these applications, the families of interest are those from Equation (3.2.9):

$$\mathcal{X}_f^* = \{(x, z) \in T \times Z : f(x, z) \neq 0\} = \mathcal{X}_F, \quad \text{where } F(x, z) = x_0 \cdots x_d f(x, z). \quad (3.2.20)$$

As in the Introduction,  $T = D(x_0 \cdots x_d)$  is the dense torus of  $\mathbb{P}^d$ . In physics, the signed Euler characteristic of  $\mathcal{X}_{F,z}$  measures the number of *master integrals*, which form a basis of a vector space generated by Feynman integrals [AFST24, BBKP19]. The dimension of that vector space depends on kinematic parameters, here captured by  $z$ . In algebraic statistics, very affine hypersurface complements arise as toric models, i.e. discrete exponential families, see Section 2.2. The general result Theorem 2.2.12 implies that the maximum likelihood degree of such models equals the signed Euler characteristic of  $\mathcal{X}_{F,z}$ . In this context,  $z$  are model parameters. Euler stratifications aim to understand completely how the ML degree depends on  $z$ , thereby providing an answer to Problem 3.0.1.

We start by considering families in which the monomial support of  $f$  is fixed, and all coefficients vary freely. More precisely, we fix a matrix  $A$  with nonnegative integer entries of size  $(d+1) \times (m+1)$  with no repeated columns, and with constant column sum  $n > 0$ . In Example 3.2.2, the parameters are  $d = 2$ ,  $m = 5$  and  $n = 2$ . The columns of  $A$  are the integer vectors  $a_0, \dots, a_m \in \mathbb{N}^{d+1}$ . They serve as exponents in the formula for  $f(x, z)$ :

$$f(x, z) = z_0 x^{a_0} + z_1 x^{a_1} + \cdots + z_m x^{a_m}. \quad (3.2.21)$$

The Euler discriminant from Definition 3.2.18 of the family  $\mathcal{X}_F \rightarrow Z = \mathbb{P}^m$  is well-understood. We summarise this in Theorem 3.2.25 below, which needs some more notation. The convex hull  $\text{conv}(A)$  of the columns of  $A$  is a polytope in  $\mathbb{R}^{d+1}$  of dimension at most  $d$ . Its lattice volume is denoted by  $\text{vol}(A)$ . The *A-discriminant variety*  $\nabla_A$  is the closure of all parameter values  $z^*$  for which  $\mathcal{V}(f(x, z^*)) \cap T$  is singular. If  $\nabla_A$  has codimension one, then its defining equation is denoted by  $\Delta_A$  (this is defined up to a nonzero scalar multiple). If  $\nabla_A$  has codimension at least two, we set  $\Delta_A = 1$ . The *principal A-determinant* is defined as the following *A*-resultant:

$$E_A(z) = \text{Res}_A \left( f, x_1 \frac{\partial f}{\partial x_1}, \dots, x_d \frac{\partial f}{\partial x_d} \right) \in \mathbb{Q}[z_0, \dots, z_m].$$

For more background, see [GKZ08, Ch. 10]. The polynomial  $E_A(z)$  is known to factor into a product of discriminants. For each face  $Q \preceq \text{conv}(A)$ , let  $A_Q$  be the submatrix of  $A$  consisting of columns which lie on  $Q$ . By [GKZ08, Ch. 10, Thm. 1.2], we have

$$E_A(z) = \prod_{Q \preceq \text{conv}(A)} \Delta_{A_Q}^{e_Q}. \quad (3.2.22)$$

Here,  $e_Q \geq 1$  is the multiplicity of the toric variety  $X_A$  along its torus orbit corresponding to  $Q$ . The next theorem summarises [ABB<sup>+</sup>19, Thm. 13] and [Est13, Thm. 2.36].

**Theorem 3.2.25.** *The signed Euler characteristic  $(-1)^d \cdot \chi(\mathcal{X}_{f,z}^*) = (-1)^d \cdot \chi(\mathcal{X}_{F,z})$  with  $f(x, z)$  as in (3.2.21) is  $\text{vol}(A)$  if and only if  $E_A(z) \neq 0$ . Moreover, for a generic point  $z$  on a codimension-one discriminant  $\nabla_{A_Q}$ , we have  $\text{vol}(A) - (-1)^d \cdot \chi(\mathcal{X}_{F,z}) = e_Q$ , with  $e_Q$  as in (3.2.22).*

**Example 3.2.26.** The principal *A*-determinant  $E_A$  from Example 3.2.2 has seven factors, one for each face of the triangle  $\text{conv}(A)$ . The toric variety of  $A$  is the 2-uple embedding of  $\mathbb{P}^2$ . Its multiplicity along each torus orbit is one. All exponents in (3.2.22) are one, and the Euler characteristic differs from its generic value by one on each component of

$$\{E_A(z) = 0\}. \quad \diamond$$

Before switching back to the case where  $Z$  is any irreducible subvariety of  $\mathbb{P}^m$ , we discuss useful formulae for  $\chi(\mathcal{X}_{F,z^*}) = \chi(D(F(x, z^*)))$  when  $z = z^*$  is fixed and arbitrary. In Theorem 2.2.12, the signed Euler characteristic of  $D(F)$  is expressed as the number of critical points of the log-likelihood function. We rephrase this theorem as follows.

**Theorem 3.2.27.** Let  $F(x) = x_0 \cdots x_d f(x) \neq 0$ . There is a dense open subset  $U \subseteq \mathbb{P}^d$  such that for  $(v_0 : v_1 : \dots : v_d) \in U$ , the following equations have precisely  $(-1)^d \cdot \chi(D(F))$  nondegenerate isolated solutions in  $D(F) = T \setminus \mathcal{V}(f)$ :

$$\frac{v_1}{x_1} + v_0 \frac{\partial_{x_1} f}{f} = \dots = \frac{v_d}{x_d} + v_0 \frac{\partial_{x_d} f}{f} = 0. \quad (3.2.23)$$

In Theorem 3.2.27, a *nondegenerate isolated solution* is a point  $x \in D(F)$  satisfying (3.2.23) at which the Jacobian determinant of these  $d$  equations does not vanish. The notation  $\partial_{x_i} f$  is short for  $\frac{\partial f}{\partial x_i}$ . Using coordinates  $t_i = \frac{x_i}{x_0}$  on  $T$ , one writes the equations (3.2.23) as

$$\mathrm{d}\log f(1, t_1, \dots, t_d)^{v_0} t_1^{v_1} \cdots t_d^{v_d} = 0. \quad (3.2.24)$$

Theorem 3.2.27 has important practical implications: it turns out that computing Euler characteristics of very affine hypersurfaces via critical points is efficient and reliable in practice, see e.g. [ABF<sup>+</sup>23, AFST24, MHMT23]. Below, we will use Theorem 3.2.27 to compute Euler discriminants.

Theorem 3.2.7 expresses the degree of  $\nabla F$  as a sum of Euler characteristics:

$$\deg \nabla F = (-1)^d \cdot \chi(D(F)) + (-1)^{d-1} \cdot \chi(D(F) \cap H). \quad (3.2.25)$$

We explain how this is a decomposition of critical points according to a *likelihood degeneration*, in the sense of [ABF<sup>+</sup>23]. First, we show  $\deg \nabla F$  is indeed a critical point count.

**Proposition 3.2.28.** Let  $0 \neq f \in \mathbb{C}[x_0, \dots, x_d]_n$  be a form of degree  $n$  and let  $L = b_0 x_0 + \cdots + b_d x_d \in \mathbb{C}[x_0, \dots, x_d]_1$  be a generic linear form. The degree of the Gauss map  $\nabla F$  with  $F = x_0 x_1 \cdots x_d f$  equals the number of solutions in  $D(F \cdot L)$  to the equations

$$\frac{1}{x_i} + \frac{\partial_{x_i} f}{f} - (d+n+1) \frac{b_i}{L} = 0, \quad i = 1, \dots, d. \quad (3.2.26)$$

*Proof.* By definition, the degree of  $\nabla F$  is the number of points  $x \in D(F)$  satisfying

$$\mathrm{rank} \begin{pmatrix} \frac{\partial F}{\partial x_0} & \frac{\partial F}{\partial x_1} & \cdots & \frac{\partial F}{\partial x_d} \\ b_0 & b_1 & \cdots & b_d \end{pmatrix} = \mathrm{rank} \begin{pmatrix} \frac{1}{x_0} + \frac{\partial_{x_0} f}{f} & \frac{1}{x_1} + \frac{\partial_{x_1} f}{f} & \cdots & \frac{1}{x_d} + \frac{\partial_{x_d} f}{f} \\ b_0 & b_1 & \cdots & b_d \end{pmatrix} = 1,$$

for a generic point  $b \in \mathbb{P}^d$ . Here we divided the first row of the matrix by  $F = x_0 x_1 \cdots x_d f$ , which does not change the rank. Multiplying from the right with the matrix obtained by replacing the first column of the identity matrix with  $(x_0, \dots, x_d)$ , we obtain the equations

$$\mathrm{rank} \begin{pmatrix} d+n+1 & \frac{1}{x_1} + \frac{\partial_{x_1} f}{f} & \cdots & \frac{1}{x_d} + \frac{\partial_{x_d} f}{f} \\ b_0 x_0 + \cdots + b_d x_d & b_1 & \cdots & b_d \end{pmatrix} = 1.$$

Note that this last step is allowed because  $F$  has the special form  $F = x_0 x_1 \cdots x_d f$ . Since  $b$  is generic, we find that  $L(x) = b_0 x_0 + \cdots + b_d x_d \neq 0$ , and we obtain (3.2.26).  $\square$

In  $t$ -coordinates, the equations (3.2.26) are the critical point equations given by

$$\mathrm{d}\log f(1, t_1, \dots, t_d) t_1 \cdots t_d L(1, t_1, \dots, t_d)^{-(d+n+1)} = 0. \quad (3.2.27)$$

The decomposition (3.2.25) is seen from Theorem 3.2.27 and Proposition 3.2.28 by introducing a parameter  $\varepsilon$  in the exponents, and taking the limit  $\varepsilon \rightarrow 0$ . Explicitly, we set

$$\mathcal{L}(\varepsilon) = f(1, t_1, \dots, t_d)^{(1-\varepsilon)v_0+\varepsilon} t_1^{v_1(1-\varepsilon)+\varepsilon} \cdots t_d^{v_d(1-\varepsilon)+\varepsilon} L(1, t_1, \dots, t_d)^{-\varepsilon(d+n+1)}.$$

The solutions to  $\text{dlog } \mathcal{L}(\varepsilon) = 0$  are tuples of Puiseux series in  $\varepsilon$ . For  $\varepsilon = 1$ , they evaluate to the  $(\deg \nabla F)$ -many solutions to (3.2.27). The equations are explicitly given by

$$\frac{\nu_i(1-\varepsilon)+\varepsilon}{t_i} + ((1-\varepsilon)\nu_0+\varepsilon) \frac{\partial_{t_i}\tilde{f}}{\tilde{f}} - \varepsilon(d+n+1) \frac{b_i}{\tilde{L}} = 0, \quad i = 1, \dots, d, \quad (3.2.28)$$

where  $\tilde{f} = f(1, t)$  and  $\tilde{L} = L(1, t)$ . When  $\varepsilon \rightarrow 0$ , the solutions split up into two groups. The first group of solutions is of the form  $t(\varepsilon) = t^{(0)} + \text{higher order terms in } \varepsilon$ , with the lowest order term  $t^{(0)} \in (\mathbb{C}^*)^d$  satisfying  $L(1, t^{(0)}) \neq 0$ . Vanishing of the lowest order term in (3.2.28) implies that these solutions converge to the critical points of (3.2.24). The leading term  $t^{(0)} \in (\mathbb{C}^*)^d$  of the second group of solutions satisfies  $L(1, t^{(0)}) = 0$ , and  $L(1, t(\varepsilon)) = \varepsilon L^{(0)} + \text{higher order terms}$ , with  $L^{(0)} \in \mathbb{C}^*$ . In particular, these solutions converge to points on  $H = \mathcal{V}(L)$ . Plugging these ansätze for  $t$  and  $L$  into (3.2.28) and setting the lowest order term to zero shows that  $t^{(0)}$  are the critical points of the restriction of  $t_1^{\nu_1} \cdots t_d^{\nu_d} \tilde{f}^{\nu_0}$  to  $H$ . By Theorem 3.2.27, the first group of solutions consists of  $(-1)^d \cdot \chi(D(F))$  points, and the second group has  $(-1)^{d-1} \cdot \chi(D(F) \cap H)$  solutions. They sum up to the number of solutions to (3.2.27) by Equation (3.2.25).

**Remark 3.2.29.** The discussion above amounts to solving the equations (3.2.28) *tropically*. We have described the valuation of the coordinates  $t_i, \tilde{L}$  on  $D(L)$  for each  $\varepsilon$ -series solution. For more, see [ABF<sup>+</sup>23, §7].

**Example 3.2.30.** We work out an example with  $d = n = 1$ . Let  $f = x_1 - x_0$ ,  $L = x_1 - 2x_0$ . We are interested in the critical points of  $\log \mathcal{L}(\varepsilon)$ , given in the coordinate  $t = x_1/x_0$  by

$$\mathcal{L}(\varepsilon) = (t-1)^{\nu_0(1-\varepsilon)+\varepsilon} t^{\nu_1(1-\varepsilon)+\varepsilon} (t-2)^{-3\varepsilon}.$$

For concreteness, we choose  $\nu_0 = -11, \nu_1 = 17$ . The solutions of (3.2.28) are

$$t_{\pm}(\varepsilon) = \frac{29 - 27\varepsilon \pm \sqrt{25 + 154\varepsilon - 167\varepsilon^2}}{2(6 - 7\varepsilon)}.$$

The two solutions of (3.2.27) are  $t_{\pm}(1) = -1 \mp \sqrt{3}$ . When  $\varepsilon \rightarrow 0$ , the solutions converge to  $t_+(0) = 17/6$  and  $t_-(0) = 2$ . Here  $\{17/6\}$  is the unique critical point of (3.2.24) (it is rational since  $|\chi(D(x_0x_1f))| = 1$ ), and  $\{2\} = \mathcal{V}(L)$ . Equation (3.2.25) reads  $2 = 1 + 1$ .  $\diamond$

Theorem 3.2.27 can be used to show that the Euler characteristic of very affine hypersurfaces is semicontinuous. This is not the case for general projective hypersurface families, see Example 3.2.1.

**Theorem 3.2.31.** *Let  $\mathcal{X}_F$  be the family from (3.2.20), with  $Z \subseteq \mathbb{P}^m$  any irreducible quasi-projective variety. The set  $Z_{\leq k} = \bigcup_{q \leq k} Z_q$ , with  $Z_q = \{z \in Z : (-1)^d \cdot \chi(\mathcal{X}_{F,z}) = q\}$ , is closed in  $Z$ .*

*Proof.* By Lemma 3.2.8, the set  $Z_q$  is constructible. That is,  $Z_q$  has a decomposition

$$Z_q = (V_1^q \setminus W_1^q) \cup \cdots \cup (V_{m_q}^q \setminus W_{m_q}^q),$$

where  $V_i^q = \overline{V_i^q \setminus W_i^q}$  are irreducible closed subvarieties of  $Z$ . By Theorem 3.2.27, the Euler characteristic of  $\mathcal{X}_{F,z}$  for  $z \in V_i^q$  is the generic number of nondegenerate isolated solutions to the following system of  $d+1$  parametric polynomial equations on  $(\mathbb{C}^*)^{d+1}$ :

$$\nu_1 x_1^{-1} + \nu_0 y \partial_1 f = \cdots = \nu_d x_d^{-1} + \nu_0 y \partial_d f = yf - 1 = 0.$$

By [SW<sup>+</sup>05, Thm. 7.1.4], this quantity is semicontinuous, and the maximum  $q$  is attained for generic  $\nu, z$  in the dense open subset  $V_i^q \setminus W_i^q$ . We conclude that  $V_i^q \subseteq Z_{\leq q} \subseteq Z_{\leq k}$ . As a consequence,  $\overline{Z_q} \subseteq Z_{\leq k}$ , and we are done. This proof appeared in [FMT24, Thm. 3.1].  $\square$

---

**Algorithm 6** Euler discriminant of  $\pi_{x_0x_1\cdots x_d}: \mathcal{X}_{x_0x_1\cdots x_d} \rightarrow Z$ 


---

**Input:** bihomogeneous polynomial  $f \in \mathbb{Q}[x, z]$  and the defining ideal  $\mathcal{I}_Z$  of  $Z$   
**Output:** minimal prime ideals of the Euler discriminant of  $\pi_{x_0x_1\cdots x_d}: \mathcal{X}_{x_0x_1\cdots x_d} \rightarrow Z$

- 1: **for**  $i = 0, \dots, \dim Z$  **do**
  - 2:    $v \leftarrow$  a random point in  $\mathbb{P}^d$
  - 3:   compute the vanishing ideal of the closure  $\mathcal{Y}_v$  of  $\mathcal{Y}_v^\circ$  in (3.2.29)
  - 4:    $I_i \leftarrow$  the defining ideal of the projection  $\pi_Z(\mathcal{Y}_v \cap \mathcal{V}(F))$
  - 5: **end for**
  - 6: **return** minimal primes of  $I_0 + I_1 + \dots + I_{\dim Z}$
- 

Algorithm 5 can in principle be used to compute the Euler discriminant of the family (3.2.20). However, there is a more efficient (but still randomised) algorithm based on Theorem 3.2.27. We fix random parameters  $v \in \mathbb{P}^d$  and, much like in (3.2.18), we define the incidence

$$\mathcal{Y}_v^\circ = \left\{ (x, z) \in \mathcal{X}_F : \frac{v_i}{x_i} + v_0 \frac{\partial_i f(x, z)}{f(x, z)} = 0, i = 1, \dots, d \right\}. \quad (3.2.29)$$

The closure in  $\mathbb{P}^d \times Z$  is denoted by  $\mathcal{Y}_v$ . The number of critical points of (3.2.24) drops when a solution is contained in the boundary  $\mathcal{Y}_v \setminus \mathcal{Y}_v^\circ$ . Hence, the Euler discriminant is contained in the projection  $\pi_Z(\mathcal{Y}_v \setminus \mathcal{Y}_v^\circ) = \pi_Z(\mathcal{Y}_v \cap \mathcal{V}(F))$ . This is strictly contained in  $Z$ .

**Proposition 3.2.32.** *Let  $\mathcal{X}_F$  be the family from (3.2.20), with  $Z \subseteq \mathbb{P}^m$  any irreducible quasi-projective variety. Let  $\mathcal{Y}_v^\circ$  be as in (3.2.29), where  $v \in \mathbb{P}^d$  is generic. The variety  $\mathcal{Y}_v \cap \mathcal{V}(F)$  has dimension at most  $\dim Z - 1$ . Hence,  $\pi_Z(\mathcal{Y}_v \cap \mathcal{V}(F)) \subsetneq Z$  is a strict containment.*

*Proof.* The proof is similar to that of Proposition 3.2.21. In this case, the incidence is

$$\mathcal{Y}^\circ = \left\{ (x, z, v) \in \mathcal{X}_F \times \mathbb{P}_v^d : \frac{v_i}{x_i} + v_0 \frac{\partial_i f(x, z)}{f(x, z)} = 0, i = 1, \dots, d \right\}. \quad (3.2.30)$$

It is again parametrised by  $\mathcal{X}_F$ , and hence irreducible of dimension  $d + \dim Z$ .  $\square$

Algorithm 6 is based on ideas similar to those in Section 3.2.3. Components in  $\pi_Z(\mathcal{Y}_v \cap \mathcal{V}(F))$  which depend on  $v$  are eliminated by performing the computation  $\dim Z + 1$  times, and adding up the ideals. Importantly, in practice we observe that this repeated computation is not needed. This suggests that the projection of  $\overline{\mathcal{Y}^\circ} \cap (\mathcal{V}(F) \times \mathbb{P}_v^d) \subset \mathbb{P}_x^d \times Z \times \mathbb{P}_v^d$  to  $Z \times \mathbb{P}_v^d$  only has components of the form  $V \times \mathbb{P}_v^d$  or  $Z \times W$ . Here  $\mathcal{Y}^\circ$  is as in (3.2.30). The components of the form  $V \times \mathbb{P}_v^d$  contribute to the Euler discriminant. Among the components of the form  $Z \times W$  there are the hyperplanes identified in [SvdV23].

**Example 3.2.33.** Our Julia implementation of Algorithm 6 is used as in Algorithm 3 to compute an Euler stratification of  $\mathcal{X}_f^*$ , with  $f$  the ternary quadric from (3.2.1). As predicted by (3.2.5), there are seven closed strata of dimension four. Among the deeper strata, there are 21 of dimension three, 27 of dimension two, and 15 of dimension one. The latter consist of twelve lines

$$\langle z_5, z_2, z_1, z_0 \rangle, \langle z_5, z_3, z_2, z_0 \rangle, \langle z_5, z_4, z_2, z_0 \rangle, \langle z_3, z_2, z_1, z_0 \rangle, \langle z_4, z_3, z_2, z_0 \rangle, \langle z_4, z_3, z_1, z_0 \rangle, \\ \langle z_5, z_3, z_1, z_0 \rangle, \langle z_5, z_4, z_1, z_0 \rangle, \langle z_5, z_4, z_3, z_0 \rangle, \langle z_5, z_3, z_2, z_1 \rangle, \langle z_5, z_4, z_3, z_2 \rangle, \langle z_5, z_4, z_3, z_1 \rangle,$$

and three quadric curves:  $\langle 4z_3z_5 - z_4^2, z_2, z_1, z_0 \rangle, \langle 4z_0z_5 - z_2^2, z_4, z_3, z_1 \rangle, \langle 4z_0z_3 - z_1^2, z_5, z_4, z_2 \rangle$ .

$\diamond$

Table 3.1: Degrees of generators for the ideal of coincident root loci of a binary octic.

| $\lambda$   | generators of $I(\nabla_\lambda)$ | $\lambda$     | generators of $I(\nabla_\lambda)$ |
|-------------|-----------------------------------|---------------|-----------------------------------|
| (8)         | $2^{28}$                          |               |                                   |
| (7,1)       | $2^{15}$                          | (4,4)         | $3^{74}$                          |
| (6,1,1)     | $2^6, 3^1$                        | (3,1,1,1,1)   | $6^1, 7^6, 8^6$                   |
| (6,2)       | $2^6, 3^1, 4^{46}$                | (3,2,1,1,1)   | $6^1, 7^6, 8^6, 10^{46}$          |
| (5,1,1,1)   | $2^1, 3^6, 4^6$                   | (3,2,2,1)     | $6^1, 7^6, 8^{546}$               |
| (5,2,1)     | $2^1, 3^6, 4^6, 6^{54}$           | (3,3,1,1)     | $5^{84}$                          |
| (5,3)       | $2^1, 3^{19}, 4^{64}$             | (3,3,2)       | $4^1, 5^{166}, 6^{100}$           |
| (4,1,1,1,1) | $4^5, 5^9, 6^1$                   | (2,1,1,1,1,1) | $14^1$                            |
| (4,2,1,1)   | $4^6, 5^9, 6^1, 8^{54}$           | (2,2,1,1,1,1) | $11^{19}$                         |
| (4,2,2)     | $4^{15}, 5^{221}$                 | (2,2,2,1,1)   | $8^{120}$                         |
| (4,3,1)     | $4^{45}$                          | (2,2,2,2)     | $5^{286}$                         |

### 3.2.5. Applications

In the following subsections we present several case studies and applications of Euler stratifications. We compute the full stratification of binary octics in Subsection 3.2.6 ( $n = 8$  in (3.2.10)). In Subsection 3.2.7, we recall the relation between the Euler stratification of bilinear forms and the matroid stratification of the Grassmannian. Subsection 3.2.8 discusses applications in physics, while Subsection 3.2.9 focusses on statistics, tying back to the guiding topic of this chapter, Problem 3.0.1. Finally, in Subsection 3.2.10, we stratify hyperplane sections of Hirzebruch surfaces. Our code uses the packages `Oscar.jl` (v1.0.4) [OSC24] and `HomotopyContinuation.jl` (v2.9.4) [BT18] and is presented at [TW24b].

### 3.2.6. Binary octics

We come back to binary forms and their coincident root loci as discussed in Section 3.2.2. Our goal is to compute the full Euler stratification for a binary octic ( $n = 8$ ). According to Theorem 3.2.13, this requires the computation of ideals  $I(\nabla_\lambda)$  of coincident root loci for partitions  $\lambda$  of up to eight. By Theorem 3.2.14, the Euler stratification of eight points in  $\mathbb{C}^*$  is easily deduced from the same ideals. We describe our method of computing  $I(\nabla_\lambda)$ .

First, we compute the degrees of a minimal generating set of  $I(\nabla_\lambda)$ . In [LS16, Tbl. 1], these numbers are reported for partitions of up to seven. We extend this table to the case  $\lambda \vdash 8$  using finite field computations, see Table 3.1. Here, the entry for  $\lambda = (2^2, 1^4)$  is conjectural; the finite field computation did not terminate within seven days. Using the parametrisation of  $\nabla_\lambda$  given by (3.2.12), we sample points on  $\nabla_\lambda$  and interpolate with polynomials of the computed degrees using linear algebra over  $\mathbb{Q}$ . For this, we exploit the homogeneity of  $I(\nabla_\lambda)$  with respect to the bigrading given by the exponent matrix

$$\begin{pmatrix} 8 & 7 & \dots & 1 & 0 \\ 0 & 1 & \dots & 7 & 8 \end{pmatrix}.$$

That grading is obtained from the  $\mathbb{C}^*$ -action on the space of binary octics  $\mathbb{P}^8$ . It divides the set of degree  $d$  monomials in  $\mathbb{C}[z_0, \dots, z_8]$  into  $8d + 1$  many buckets of monomials with the same bidegree. The resulting linear system that needs to be solved becomes substantially smaller. For example, the maximal number of monomials of degree eleven in one bucket

is 2430, compared to 75582 monomials overall. The implementation as well as a database with all ideals for the Euler stratification of a binary octic in  $\mathbb{C}^*$  can be found at [TW24b].

### 3.2.7. Matroid stratification of bilinear forms

We start with very affine hypersurfaces defined by bilinear forms. For  $d_1, d_2 \geq 1$ , consider

$$f(x, y, z) = \sum_{i=0}^{d_1} \sum_{j=0}^{d_2} z_{ij} x_i y_j. \quad (3.2.31)$$

The parameter space  $Z = \mathbb{P}^{(d_1+1)(d_2+1)-1}$  is that of  $(d_1 + 1) \times (d_2 + 1)$  matrices  $z = (z_{ij})_{i,j}$ . Since  $f$  is bilinear, it is natural to consider its zero locus in the torus of  $\mathbb{P}^{d_1} \times \mathbb{P}^{d_2}$ . In this subsection, we set  $(\mathbb{C}^*)^{d_1+d_2} \cong T \subset \mathbb{P}^{d_1} \times \mathbb{P}^{d_2}$ , and we consider the following family:

$$\mathcal{X}_f^* = \{(x, y, z) \in T \times Z : f(x, y, z) \neq 0\}.$$

The fibres  $\mathcal{X}_{f,z}^*$  of  $\mathcal{X}_f^* \rightarrow Z$  represent independence models in algebraic statistics. Their Euler characteristic was studied in [CHKO24]. One of the main results [CHKO24, Thm. 1.3] states that for any  $z \in Z$ , the signed Euler characteristic  $(-1)^{d_1+d_2} \cdot \chi(\mathcal{X}_{f,z}^*)$  equals the Möbius invariant of the rank- $(d_1 + 1)$  matroid represented by the  $(d_1 + 1) \times (d_1 + d_2 + 2)$  matrix  $[\text{Id}_{d_1+1} \ z]$ .

A corollary is that an Euler stratification of  $\mathcal{X}_f^* \rightarrow Z$  is induced by the matroid stratification of the Grassmannian  $\text{Gr}(d_1 + 1, d_1 + d_2 + 2)$ . In particular, the Euler discriminant is the union of all hypersurfaces defined by the  $(d_1 + 1) \times (d_1 + 1)$  minors of  $[\text{Id}_{d_1+1} \ z]$ . Equivalently, these are the hypersurfaces defined by all minors of the matrix  $z$ .

If  $z \in \mathbb{R}^{(d_1+1) \times (d_2+1)}$  has real entries, then the Möbius invariant  $(-1)^{d_1+d_2} \cdot \chi(\mathcal{X}_{f,z}^*)$  equals the number of bounded cells in the following hyperplane arrangement complement:

$$\left\{ t \in \mathbb{R}^{d_1} : t_1 \cdots t_{d_1} \left( z_{00} + \sum_{i=1}^{d_1} z_{i0} t_i \right) \cdots \left( z_{0d_2} + \sum_{i=1}^{d_1} z_{id_2} t_i \right) \neq 0 \right\}.$$

**Example 3.2.34.** For  $d_1 = 2, d_2 = 1$ , the polynomial  $f$  and the matrix  $[\text{Id}_{d_1+1} \ z]$  are

$$f = z_{00} x_0 y_0 + z_{10} x_1 y_0 + z_{20} x_2 y_0 + z_{01} x_0 y_1 + z_{11} x_1 y_1 + z_{21} x_2 y_1, \quad [\text{Id}_{d_1+1} \ z] = \begin{pmatrix} 1 & 0 & 0 & z_{00} & z_{01} \\ 0 & 1 & 0 & z_{10} & z_{11} \\ 0 & 0 & 1 & z_{20} & z_{21} \end{pmatrix}.$$

For generic choices of  $z_{ij} \in \mathbb{R}$ , the lines  $\{z_{0i} + z_{1i}t_1 + z_{2i}t_2 = 0\}, i = 0, 1$ , together with  $\{t_1 = 0\}$  and  $\{t_2 = 0\}$ , define three bounded cells in  $\mathbb{R}^2$ . The Euler discriminant of  $\mathcal{X}_f^*$  vanishes when at least one of these cells collapses and is the product of the minors of  $z$ :

$$\Delta_\chi = z_{00} z_{10} z_{20} z_{01} z_{11} z_{21} (z_{00} z_{11} - z_{01} z_{10}) (z_{00} z_{21} - z_{01} z_{20}) (z_{10} z_{21} - z_{11} z_{20}). \quad \diamond$$

### 3.2.8. Feynman integrals

Let  $\mathcal{X}_f^*$  be the family of very affine hypersurfaces from (3.2.20). In particle physics, the signed Euler characteristic of the fibre  $\mathcal{X}_{f,z}^*$  counts the number of *master integrals* [BBKP19]. These are (regularised) Feynman integrals which form a basis for the *twisted cohomology* of  $\mathcal{X}_{f,z}^*$  with respect to the dlog form (3.2.24), see [BBKP19, MM19]. We consider Feynman integrals in parametric representation [Wei22, §2.5.4]. In that context, the polynomial  $f$  is

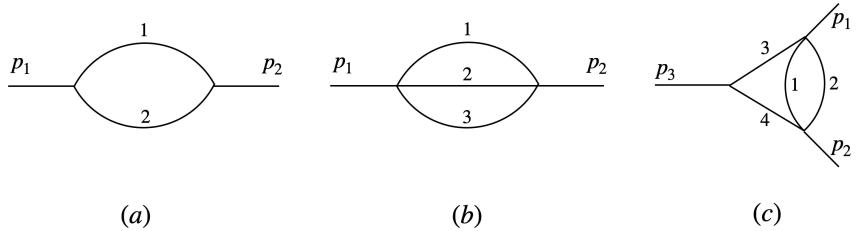


Figure 3.6: Three Feynman diagrams: bubble, sunrise and parachute.

the *graph polynomial* associated to a Feynman graph. The variables  $x$  are the integration variables, called *Schwinger parameters*, and the parameters  $z$  are *kinematic data*, such as momenta and masses of fundamental particles. The theory of twisted cohomology extends to the more general class of *Euler integrals* studied in [AFST24, MHMT23], where  $f$  can be any polynomial. Feynman graphs also appear in Section 5.2 in a combinatorial context.

Since  $x$  is integrated out, a Feynman integral represents a multivalued function of  $z$ . The goal of *Landau analysis* is to detect branch and pole loci of this multivalued function [MT22]. By [Bro09, Thm. 56], these loci are contained in the union of codimension-one strata in a Whitney stratification of  $\mathcal{X}_f^* \rightarrow Z$ . Brown defines the *Landau variety* as this union [Bro09, Def. 54]. Algorithms for computing the Landau variety can be found in [HPT24, Pan15]. The paper [FMT24] conjectures that the branch and pole loci are detected by a drop in the number of master integrals, i.e. by the Euler discriminant of the family  $\mathcal{X}_f^* \rightarrow Z$ . However, no algorithm for computing the Euler discriminant is presented in [FMT24]. Instead, an approximation of  $\nabla \chi$  is computed, called the *principal Landau determinant*, using efficient techniques from numerical algebraic geometry inspired by [MT22]. Here, we use Algorithm 6 to compute the Euler discriminant of several Feynman integrals, including some appearing in [HPT24]. In all examples we checked, the Euler discriminant coincides with Brown's Landau variety. In future work, we aspire to develop numerical versions of our algorithms to tackle more challenging diagrams.

**Example 3.2.35.** We start with the bubble graph (Figure 3.6(a)), with graph polynomial

$$f = (x_0 - m_1 x_1 - m_2 x_2)(x_1 + x_2) + s x_1 x_2.$$

The variety  $\mathcal{X}_f^* \subset \mathbb{P}^2 \times Z$  is a family of curve complements over  $Z = \mathbb{C}^3$ , where  $Z$  has coordinates  $(m_1, m_2, s)$ . The parameters represent the squared masses  $m_i$  of particles travelling along the internal edges of the diagram, and the *Mandelstam invariant*  $s$  depending on momenta of external particles. Algorithm 3 computes the following strata:

$$\begin{aligned} \text{codim 1 : } & \langle m_1 \rangle, \langle m_2 \rangle, \langle s \rangle, \langle m_1^2 - 2m_1 m_2 - 2m_1 s + m_2^2 - 2m_2 s + s^2 \rangle \\ \text{codim 2 : } & \langle m_2 - s, m_1 \rangle, \langle s, m_1 \rangle, \langle m_2, m_1 \rangle, \langle s, m_1 - m_2 \rangle, \langle m_2, m_1 - s \rangle, \langle s, m_2 \rangle \\ \text{codim 3 : } & \langle m_1, m_2, s \rangle \end{aligned}$$

This Euler stratification agrees with the Whitney stratification in [HPT24, Sec. V, §A].  $\diamond$

**Example 3.2.36.** Next, we consider the sunrise graph, shown in part (b) of Figure 3.6. The (homogenised) graph polynomial  $f$  defines a family of surface complements over  $Z = \mathbb{C}^4$ :

$$f = (x_0 - m_1 x_1 - m_2 x_2 - m_3 x_3)(x_1 x_2 + x_1 x_3 + x_2 x_3) + s x_1 x_2 x_3.$$

The Euler discriminant agrees with the principal Landau determinant in [FMT24, Ex. 3.7]:

$$\begin{aligned}\Delta_\chi = & m_1 \cdot m_2 \cdot m_3 \cdot s \cdot (m_1^4 - 4m_1^3m_2 - 4m_1^3m_3 - 4m_1^3s + 6m_1^2m_2^2 + 4m_1^2m_2m_3 \\ & + 4m_1^2m_2s + 6m_1^2m_3^2 + 4m_1^2m_{3s} + 6m_1^2s^2 - 4m_1m_2^3 + 4m_1m_2^2m_3 + 4m_1m_2^2s \\ & + 4m_1m_2m_3^2 - 40m_1m_2m_{3s} + 4m_1m_2s^2 - 4m_1m_3^3 + 4m_1m_3^2s + 4m_1m_3s^2 \\ & - 4m_1s^3 + m_2^4 - 4m_2^3m_3 - 4m_2^3s + 6m_2^2m_3^2 + 4m_2^2m_{3s} + 6m_2^2s^2 - 4m_2m_3^3 \\ & + 4m_2m_3^2s + 4m_2m_3s^2 - 4m_2s^3 + m_3^4 - 4m_3^3s + 6m_3^2s^2 - 4m_3s^3 + s^4).\end{aligned}\quad \diamond$$

**Example 3.2.37.** The graph polynomial of the parachute diagram (Figure 3.6(c)) reads

$$f = (x_0 - \sum_{i=1}^4 m_i x_i) \cdot ((x_1 + x_2)(x_3 + x_4) + x_1 x_2) + x_1 x_2 (M_1 x_3 + M_2 x_4) + M_3 x_3 x_4 (x_1 + x_2),$$

where  $M_i$  is the squared mass of the  $i$ -th external particle. With the specialisation  $m_1 = 1$ ,  $m_4 = 2$ ,  $M_1 = -1$  and all other parameters except  $M_3$  equal to zero, we compute that the Euler discriminant agrees with the Landau variety in [HPT24, Sec. V, §C]:

$$\Delta_\chi = M_3(M_3 - 1)(M_3 + 1)(M_3 - 2)(M_3 + 2).$$

In particular, the Euler discriminant contains the components which are missing in the principal Landau determinant, see [FMT24, Eq. (3.18)]. These components lie on a component of the Landau variety identified in [BP22, Eq. (6.15)].  $\diamond$

### 3.2.9. Maximum likelihood estimation for toric models

Recall from Section 2.2 that in algebraic statistics, a log-affine model with  $m$  possible outcomes is represented by a scaled projective toric variety  $X_{A,z} \subseteq \mathbb{P}^{m-1}$ . By Theorem 2.2.12, the maximum likelihood degree of  $X_{A,z}$  is given by the signed Euler characteristic of the very affine variety  $X_{A,z} \setminus \mathcal{H}$ , where  $\mathcal{H}$  is the distinguished hyperplane arrangement  $\mathcal{H} = \{p_1 \dots p_m (p_1 + \dots + p_m) = 0\}$ . Here,  $p_i$  are the coordinates of  $\mathbb{P}^{m-1}$ . We now explain how the dependence of  $\text{MLdeg}(X_{A,z})$  on  $z$  is governed by Euler stratifications.

Let  $A$  be an integer matrix of size  $d \times m$  corresponding to  $X_{A,z}$ , i.e.  $X_{A,z}$  is the image closure of the monomial map  $\phi_{A,z}: (\mathbb{C}^*)^d \rightarrow \mathbb{P}^{m-1}$ ,  $x \mapsto (z_1 x^{a_1} : \dots : z_m x^{a_m})$  as in Definition 2.1.1 (although we do allow some, but not all,  $z_i$  to be zero). Moreover, without loss of generality as per Proposition 2.1.3, assume that  $\phi_{A,z}$  is one-to-one. Let  $f$  be the polynomial  $f = \sum_{i=1}^m z_i x^{a_i}$  as in (3.2.21). Then the restriction of  $\phi_{A,z}$  to  $\mathcal{X}_{f,z}^*$  gives an isomorphism of very affine varieties  $\mathcal{X}_{f,z}^* \cong X_{A,z} \setminus \mathcal{H}$ . Concordantly, the Euler stratification of  $\mathcal{X}_f^* \rightarrow \mathbb{P}^{m-1}$  from (3.2.20) completely describes the dependence of the ML degree of the discrete exponential family on the model parameters  $z$ .

The description of the ML degree as the Euler characteristic of a very affine variety also plays a crucial role for the study of the parametric ML degree from the perspective of hypersurface arrangements. See Subsection 3.3.2 for a treatment from that point of view.

We can also take the following perspective: the map  $\phi_{A,z}$  identifies  $(\mathcal{X}_{f,z}^*)^c \subset (\mathbb{C}^*)^d$  with the hyperplane section  $H_z \cap \text{im } \phi_{A,z} \subset H_z \cap X_{A,z}$ , where  $H_z = \mathcal{V}(z_1 p_1 + \dots + z_m p_m)$ . Since  $\chi(\mathcal{X}_{f,z}^*) = -\chi((\mathcal{X}_{f,z}^*)^c)$ , we are stratifying  $(\mathbb{P}^{m-1})^\vee$  according to the Euler characteristic of  $H_z$  intersected with the dense torus of  $X_{A,z}$ .

**Example 3.2.38.** Following [AKK20, AO24], we consider matrices  $A$  whose columns are lattice points of *reflexive* polygons, meaning that the dual of these polygons are again lattice polygons. The corresponding toric surfaces are the Gorenstein toric Fano surfaces, of which there exist 16 isomorphism classes [AKK20, Fig. 1], i.e. there are 16 reflexive

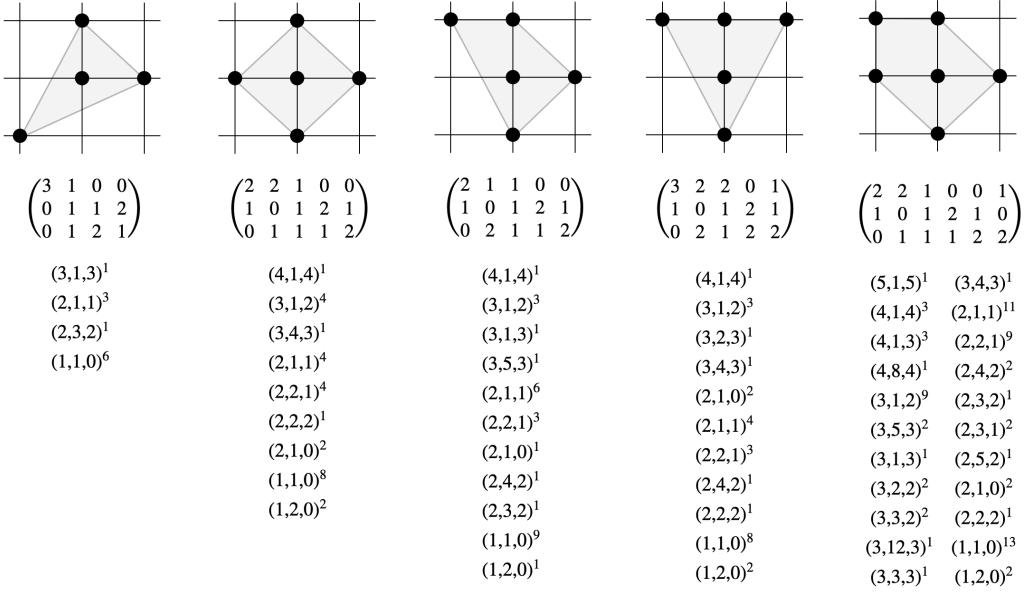


Figure 3.7: Euler stratifications for hyperplane sections of five toric Fano surfaces.

polygons up to unimodular transformation. Algorithm 3 computes the Euler stratification for the first five of these using Algorithm 6 for the Euler discriminant. These are the polygons appearing in [AKK20, Tbl. 2]. The results are summarised in Figure 3.7. The figure contains the polygons, the corresponding \$A\$-matrices, and a list of closed, irreducible strata in \$\mathbb{P}^m\$. The notation for a closed stratum \$\bar{S}\$ is \$(\dim \bar{S}, \deg \bar{S}, \chi(\mathcal{X}\_{f,z}^\*))^e\$, where \$z \in S\$ is a generic point on the stratum and the exponent \$e\$ counts the number of strata of that type. The generators for the defining ideal of each stratum and the code for reproducing these numbers can be found online at [TW24b]. Our results can be used to compute the *ML degree drop* from [AO24, Def. 2.7] for any member of the respective model families. \$\diamond\$

### 3.2.10. Hirzebruch surfaces

A Hirzebruch surface is a toric surface \$X\_A \subset \mathbb{P}^{n\_0+n\_1+1}\$ obtained from a matrix of the form

$$A = \begin{pmatrix} n_1 + 1 & n_1 & \cdots & n_1 - n_0 + 1 & n_1 & n_1 - 1 & \cdots & 0 \\ 0 & 1 & \cdots & n_0 & 0 & 1 & \cdots & n_1 \\ 0 & 0 & \cdots & 0 & 1 & 1 & \cdots & 1 \end{pmatrix},$$

where \$n\_0 \leq n\_1\$ are positive integers. As explained in Subsection 3.2.9, a hyperplane section \$H\_z \cap \text{im } \phi\_A\$ is isomorphic to a curve in the torus \$(\mathbb{C}^\*)^2 \cong T \subset \mathbb{P}^2\$. That curve is defined by

$$f = z_{00}x_0^{n_1+1} + z_{10}x_0^{n_1}x_1 + \cdots + z_{n_00}x_0^{n_1-n_0+1}x_1^{n_0} + z_{01}x_0^{n_1}x_2 + z_{11}x_0^{n_1-1}x_1x_2 + \cdots + z_{n_11}x_1^{n_1}x_2.$$

This gives rise to a family of curves \$\mathcal{X}\_f^c \subset T \times \mathbb{P}^{n\_0+n\_1+1}\$. The following Lemma already appeared as [ABB<sup>+</sup>19, Thm. 17]. Here, we give a geometric proof of the statement.

**Lemma 3.2.39.** *Let \$f\$ be as above. For any \$z \in \mathbb{P}^{n\_0+n\_1+1}\$, the signed Euler characteristic of \$(\mathcal{X}\_{f,z}^\*)^c = \{x \in T : f(x; z) = 0\}\$ equals the number of roots \$t \in \mathbb{C}^\*\$ of*

$$(z_{00} + z_{10}t + z_{20}t^2 + \cdots + z_{n_00}t^{n_0})(z_{01} + z_{11}t + z_{21}t^2 + \cdots + z_{n_11}t^{n_1}) = 0. \quad (3.2.32)$$

*Proof.* Fix  $z \in \mathbb{P}^{n_0+n_1+1}$ . Let  $(t_1, t_2) = (x_0^{-1}x_1, x_0^{-1}x_2)$  be coordinates on  $T$ . The curve  $(\mathcal{X}_{f,z}^*)^c$  is the zero locus of  $x_0^{-(n_1+1)}f(x; z) = g_0(t_1; z) + t_2 \cdot g_1(t_1; z)$ . Here the polynomials  $g_0, g_1$  are precisely the factors appearing in Equation (3.2.32). Let  $\pi_1 : (\mathcal{X}_{f,z}^*)^c \rightarrow \mathbb{C}^*$  be the projection to the  $t_1$ -coordinate. Let  $U \subset \mathbb{C}^*$  be the open subset of points  $t_1$  satisfying  $g_0(t_1; z)g_1(t_1; z) \neq 0$ . The restriction of  $\pi_1$  to  $\pi_1^{-1}(U)$  is a fibration whose fibres consist of a single point. The complement  $\mathbb{C}^* \setminus U$  is a disjoint union  $U_0 \sqcup U_1 \sqcup U_{01}$ , where

$$\begin{aligned} U_0 &= \{t_1 \in \mathbb{C}^* : g_0(t_1; z) = 0, g_1(t_1; z) \neq 0\}, \\ U_1 &= \{t_1 \in \mathbb{C}^* : g_0(t_1; z) \neq 0, g_1(t_1; z) = 0\}, \\ U_{01} &= \{t_1 \in \mathbb{C}^* : g_0(t_1; z) = g_1(t_1; z) = 0\}. \end{aligned}$$

The fibre of  $\pi_x$  over any point of  $U_0 \sqcup U_1$  is empty, and the fibre over any point of  $U_{01}$  is  $\mathbb{C}^*$ . Using the excision property and multiplicativity along fibrations of  $\chi(\cdot)$ , we obtain

$$\chi((\mathcal{X}_{f,z}^*)^c) = \chi(U) \cdot \chi(\{\text{pt}\}) + \chi(U_0) \cdot \chi(\emptyset) + \chi(U_1) \cdot \chi(\emptyset) + \chi(U_{01}) \cdot \chi(\mathbb{C}^*) = \chi(U).$$

Using excision once more we see that

$$\chi((\mathcal{X}_{f,z}^*)^c) = -\chi(\{t_1 \in \mathbb{C}^* : g_0(t_1; z)g_1(t_1; z) = 0\}). \quad \square$$

Lemma 3.2.39 reduces the Euler stratification of  $\mathcal{X}_f^*$  to the case  $d = 1$  as in Section 3.2.2.

**Theorem 3.2.40.** *Let  $\pi_f : \mathcal{X}_f^* \rightarrow \mathbb{P}^{n_0+n_1+1}$  be a family of curves with  $f$  as above. There exists an Euler stratification of  $\pi_f$  with  $\sum_{i=0}^{n_0+n_1} (n_0 + n_1 + 1 - i)\mathcal{P}(i) + 1$  many strata, where  $\mathcal{P}(i)$  is the number of partitions of  $i$ . All but one of these strata correspond bijectively to the strata in the Euler stratification of  $\mathbb{P}^{n_0+n_1}$  induced by a univariate polynomial of degree  $n_0 + n_1$  on  $\mathbb{C}^*$ .*

*Proof.* Consider the surjective morphism  $\rho : \mathbb{P}^{n_0+n_1+1} \setminus E \rightarrow \mathbb{P}^{n_0+n_1}$  sending the coefficients of  $f(x; z) = f_0(x_0, x_1, z) + x_2 f_1(x_0, x_1, z)$  to the coefficients of  $f_0 f_1$ . Here,  $E$  is the base locus  $E = \{z_{00} = \dots = z_{n_0 0} = 0\} \cup \{z_{01} = \dots = z_{n_1 1} = 0\}$ . By Lemma 3.2.39, the preimages of the Euler strata of  $f_0 f_1$  in  $\mathbb{P}^{n_0+n_1}$  under  $\rho$  give an Euler stratification of  $\mathbb{P}^{n_0+n_1+1} \setminus E$ . The remaining observation is that the Euler characteristic of  $\mathcal{X}_{f,z}^*$  along the closed set  $E$  is constant and equal to zero. The number of strata is a consequence of Theorem 3.2.14.  $\square$

### 3.3. Arrangements and parametric likelihood

The purpose of this section is twofold: Firstly, we establish connections between hypersurface arrangements [Dup15, OT13] and likelihood geometry [HS14]. Thereby arises a new description, summarised in Theorem 3.3.1, of the prime ideal  $I(\mathcal{A})$  of the likelihood correspondence of a parametrised statistical model. The description rests on the Rees algebra of the *likelihood module*  $M(\mathcal{A})$  of the arrangement  $\mathcal{A}$ , a module that is closely related to the module of logarithmic derivations introduced by Saito [Sai80] for a general arrangement. Our new description is often computationally advantageous compared to previous methods for computing the ideal  $I(\mathcal{A})$  [HKS05]. Secondly, in contrast to most previous work in algebraic statistics, our perspective considers statistical models *parametrically*, which is more natural to many statistical applications. The following is our main result.

**Theorem 3.3.1.** *The quotient  $R[s]/I(\mathcal{A})$  is the Rees algebra of the likelihood module  $M(\mathcal{A})$ .*

In Subsection 3.3.1, we start by introducing the likelihood module and other key players, and then prove Theorem 3.3.1. The nicest scenario arises when the Rees algebra agrees with the symmetric algebra. In this case, the module is of linear type. We call an arrangement  $\mathcal{A}$  *gentle* if the likelihood module  $M(\mathcal{A})$  has this property. For gentle arrangements, the ideal of the likelihood correspondence is easy to compute, and the maximum likelihood degree is determined by  $M(\mathcal{A})$ . Being gentle is a new concept that is neither implied nor implies known properties of a nonlinear arrangement  $\mathcal{A}$ , like being *free* or *tame*.

The literature on the ML degree [CHKS06, HKS05] has focused mostly on implicitly defined models. We here emphasise the parametric description that is more common in statistics, and also seen for scattering equations in physics [Lam24, ST21]. We develop these connections in Subsection 3.3.2 and compare implicit and parametric MLE.

In Subsection 3.3.3 we relate gentleness to the familiar notions of free and tame arrangements. Theorem 3.3.19 offers a concise statement. Subsection 3.3.4 considers generic arrangements, working towards a conjecture that every generic arrangement is gentle. In Subsection 3.3.5 we turn to the linear case when the hypersurfaces are hyperplanes. We study the likelihood correspondence for graphic arrangements, that is, subarrangements of the braid arrangement. The edge graph of the octahedron yields the smallest graphical arrangement which is not gentle; see Theorem 3.3.38. In Subsection 3.3.6 we present software in MACAULAY2 [GS02] for computing the likelihood correspondence of  $\mathcal{A}$ . The code we develop along with examples is available at the repository [KKM<sup>+</sup>24b].

#### 3.3.1. Arrangements and modules

Following Section 2.4, let  $f_1, \dots, f_m \in R = \mathbb{C}[x_1, \dots, x_n]$  be homogeneous polynomials and let  $\mathcal{A} = \{f_1, \dots, f_m\}$  denote a hypersurface arrangement in projective space  $\mathbb{P}^{n-1}$ .

For any complex vector  $s = (s_1, s_2, \dots, s_m) \in \mathbb{C}^m$ , we consider the *likelihood function*

$$f^s = f_1^{s_1} f_2^{s_2} \cdots f_m^{s_m}.$$

This is known as the *master function* in the arrangement literature [CDFV11]. Its logarithm

$$\ell_{\mathcal{A}} = s_1 \log(f_1) + s_2 \log(f_2) + \cdots + s_m \log(f_m)$$

is the *log-likelihood function*, similar to (2.2.3). In the context of particle physics, this function is the *scattering potential*. After choosing appropriate branches of the logarithm, the function  $\ell_{\mathcal{A}}$  is well-defined on the arrangement complement  $\mathbb{P}^{n-1} \setminus \bigcup_{f_i \in \mathcal{A}} \mathcal{V}(f_i)$ .

For us, it is natural to assume  $m > n$  and that the arrangement is *essential*, i.e. the lowest dimensional intersections are isolated points. With these hypotheses, the complement of the arrangement is a very affine variety, see [Huh13], and from now on we assume this is the case. When the  $f_i$  are linear forms, one recovers the theory of hyperplane arrangements. This is included in our setup as an important special case.

In likelihood inference one wishes to maximise  $\ell_{\mathcal{A}}$  for given  $s_1, \dots, s_m$ , see Section 2.2. Due to the logarithms, the critical equations  $\nabla \ell_{\mathcal{A}} = 0$  are not polynomial. Of course, these rational functions can be made polynomial by clearing denominators. But, multiplying through with a high degree polynomial is a very bad idea in practice. A key observation in this section is that the various modules of (log)-derivations that have been considered in the theory of hyperplane arrangements correctly solve the problem of clearing denominators.

We now define graded modules over the polynomial ring  $R$  which are associated to the arrangement  $\mathcal{A}$ . To this end, consider the following  $m \times (m+n)$  matrix

$$Q = \begin{bmatrix} f_1 & 0 & \dots & 0 & \frac{\partial f_1}{\partial x_1} & \dots & \frac{\partial f_1}{\partial x_n} \\ 0 & f_2 & \dots & 0 & \frac{\partial f_2}{\partial x_1} & \dots & \frac{\partial f_2}{\partial x_n} \\ \vdots & \ddots & & \vdots & & & \vdots \\ 0 & 0 & \dots & f_m & \frac{\partial f_m}{\partial x_1} & \dots & \frac{\partial f_m}{\partial x_n} \end{bmatrix} \in R^{m \times (m+n)}. \quad (3.3.1)$$

Each vector in the kernel  $\ker(Q)$  is naturally partitioned as  $\begin{pmatrix} a \\ b \end{pmatrix}$ , where  $a \in R^m$  and  $b \in R^n$ . With this partition, let  $\begin{pmatrix} A \\ B \end{pmatrix} \in R^{(m+n) \times l}$  be a matrix whose columns generate  $\ker(Q)$ .

We shall distinguish four graded  $R$ -modules associated with the arrangement  $\mathcal{A}$ :

- The *Terao module* of  $\mathcal{A} = \{f_1, \dots, f_m\}$  is  $\ker(Q)$ . This is a submodule of  $R^{m+n}$ .
- The *Jacobian syzygy module*  $J(\mathcal{A})$  is  $\text{im}(B)$ . This is a submodule of  $R^n$ .
- The *log-derivation module*  $D(\mathcal{A})$  is  $\text{im}(A)$ . This is a submodule of  $R^m$ .
- The *likelihood module*  $M(\mathcal{A})$  is  $\text{coker}(A)$ . This has  $m$  generators and  $l$  relations.

The first three modules are isomorphic and are often identified, see Lemma 3.3.3.

**Example 3.3.2** (Braid arrangement). Let  $m = 6, n = 4$  and let  $\mathcal{A}$  be the graphic arrangement associated with the complete graph  $K_4$ . Writing  $x, y, z, w$  for the variables, we have

$$Q = \begin{bmatrix} x-y & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & x-z & 0 & 0 & 0 & 0 & 1 & 0 & -1 & 0 \\ 0 & 0 & x-w & 0 & 0 & 0 & 1 & 0 & 0 & -1 \\ 0 & 0 & 0 & y-z & 0 & 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 & y-w & 0 & 0 & 1 & 0 & -1 \\ 0 & 0 & 0 & 0 & 0 & z-w & 0 & 0 & 1 & -1 \end{bmatrix}.$$

The Terao module  $\ker(Q) \subset R^{10}$  is free. It is generated by the  $l = 4$  rows of the matrix

$$\begin{bmatrix} A \\ B \end{bmatrix}^T = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & -1 & -1 & -1 & -1 \\ 1 & 1 & 1 & 1 & 1 & 1 & -x & -y & -z & -w \\ x+y & x+z & x+w & y+z & y+w & z+w & -x^2 & -y^2 & -z^2 & -w^2 \\ x^2+xy+y^2 & x^2+xz+z^2 & \dots & \dots & \dots & z^2+zw+w^2 & -x^3 & -y^3 & -z^3 & -w^3 \end{bmatrix}.$$

The Vandermonde matrix in the last four columns represents the syzygies on

$$\nabla f = [\partial_x f, \partial_y f, \partial_z f, \partial_w f],$$

where  $f$  is the sextic  $(x-y)(x-z)(x-w)(y-z)(y-w)(z-w)$ . This is the module  $J(\mathcal{A}) \subset R^4$ . The module  $D(\mathcal{A}) \subset R^6$  is free of rank three and generated by the three nonzero rows of  $A^T$ . This arrangement  $\mathcal{A}$  has all the nice features in Subsection 3.3.3.  $\diamond$

Recall the definition of the module of logarithmic  $\mathcal{A}$ -derivations  $\text{Der}(\mathcal{A})$  (see Definition 2.4.1). The condition  $\theta(f) \in \langle f \rangle$  in (2.4.1) ensures that the derivation  $\theta$ , when applied to the log-likelihood function  $\ell_{\mathcal{A}}$ , yields an honest polynomial rather than a rational function with  $f_i$  in the denominators. This is expressed in Theorem 3.3.12 via an injective  $R$ -module homomorphism  $\text{Der}(\mathcal{A}) \rightarrow R[s_1, \dots, s_m]$  which evaluates  $\theta$  on  $\ell_{\mathcal{A}}$ .

Using modules instead of ideals one can store more refined information, namely how each  $\theta \in \text{Der}(\mathcal{A})$  acts on the individual factors  $f_i$  or their logarithms. While at first it might seem natural to store elements of  $\text{Der}(\mathcal{A})$  as coefficient vectors in  $R^n$ , it is more efficient to store their values on the  $f_i$ . This yields the log-derivation module  $D(\mathcal{A})$ . This representation has been used in computer algebra systems like MACAULAY2, together with the matrix  $M$  from above. In the likelihood context, it appears in [HKS05, Alg. 18].

**Lemma 3.3.3.** *Let  $\mathcal{A}$  be an arrangement in  $\mathbb{P}^{n-1}$ , defined by coprime polynomials  $f_1, \dots, f_m$ .*

1. *The Terao module, the Jacobian syzygy module  $J(\mathcal{A})$ , the log-derivation module  $D(\mathcal{A})$ , and the module of logarithmic  $\mathcal{A}$ -derivations  $\text{Der}(\mathcal{A})$  are all isomorphic as  $R$ -modules.*
2. *We have  $J(\mathcal{A}) \cong J_0(\mathcal{A}) \oplus R\theta_E$ , where the second direct summand is the free rank 1 module spanned by the Euler derivation  $\theta_E = \sum_{i=1}^n x_i \partial_{x_i}$ , and  $J_0(\mathcal{A}) = \ker(R^n \xrightarrow{\nabla_f} R)$ .*
3. *The four modules above are isomorphic to the first syzygy module of the likelihood module. In particular,  $\text{pd}(M(\mathcal{A})) = \text{pd}(D(\mathcal{A})) + 1$  holds for their projective dimensions.*

*Proof.* The isomorphisms exist because the condition  $\theta(f) \in \langle f \rangle$  is equivalent to the simultaneous conditions  $\theta(f_i) \in \langle f_i \rangle$  for  $i = 1, \dots, m$ . Here we use that  $f_1, \dots, f_m$  are coprime. Item 2 follows directly from the first item in Proposition 2.4.3. For item 3 we consider free resolutions over the ring  $R$ . Let  $A \in R^{m \times l}$  be the matrix whose image equals  $D(\mathcal{A})$ . A free resolution of  $\text{coker}(A)$  uses  $A$  as the map  $F_0 \leftarrow F_1$ , i.e.

$$0 \leftarrow M(\mathcal{A}) \leftarrow R^m \xleftarrow{A} R^l \xleftarrow{A_2} F_2 \leftarrow \dots$$

The image of  $A$  is a submodule of  $R^m$ , and its free resolution looks like this:

$$0 \leftarrow D(\mathcal{A}) \leftarrow R^l \xleftarrow{A_2} F_2 \leftarrow F_3 \leftarrow \dots$$

The module  $R^l$  sits in homological degree zero in the resolution of  $\text{im}(A) = D(\mathcal{A})$ , and it sits in homological degree one in the resolution of  $\text{coker}(A) = M(\mathcal{A})$ . The two resolutions agree from the map  $A$  on to the right, but the homological degree is shifted by one.  $\square$

Having introduced the various modules for an arrangement  $\mathcal{A}$ , we now turn our attention to likelihood geometry. This concerns the critical equations  $\nabla \ell_{\mathcal{A}} = 0$  of the log-likelihood function. To capture the situation for all possible data values  $s_i$ , we define the parametric likelihood correspondence below. This definition should be contrasted with the *implicitly defined* likelihood correspondence in Definition 2.2.9.

**Definition 3.3.4.** The *(parametric) likelihood correspondence*  $\mathcal{L}_{\mathcal{A}}$  is the Zariski closure of

$$\left\{ (x, s) \in \mathbb{C}^n \times \mathbb{C}^m : \frac{\partial \ell_{\mathcal{A}}}{\partial x_i}(x, s) = 0, i = 1, \dots, n, f^s(x) \neq 0, F(x) \in X_{\text{reg}} \right\} \subset \mathbb{P}^{n-1} \times \mathbb{P}^{m-1},$$

where  $X$  is the Zariski closure of the image of  $F: \mathbb{C}^n \rightarrow \mathbb{C}^m$ ,  $x \mapsto (f_1(x), \dots, f_m(x))$ , and  $X_{\text{reg}}$  is its set of nonsingular points. The *likelihood ideal*  $I(\mathcal{A})$  is the vanishing ideal of  $\mathcal{L}_{\mathcal{A}}$ .

The likelihood correspondence is a key player in algebraic statistics [BCF23, HS14]. For example, the ML degree (see Definition 3.3.13) can be read off from its multidegree. Analogously to the implicit likelihood correspondence, we have the following result.

**Lemma 3.3.5.** *The likelihood ideal  $I(\mathcal{A})$  is prime and  $\mathcal{L}_{\mathcal{A}}$  is an irreducible variety.*

*Proof.* For each fixed vector  $x \in \mathbb{C}^n$ , the likelihood equations are linear in the  $s$ -variables. The locus where this linear system has the maximal rank is Zariski open and dense in  $\mathbb{C}^n$ . By our assumption  $m > n$ , the variety  $\mathcal{L}_{\mathcal{A}}$  is therefore a vector bundle of rank  $m - n$ . In particular,  $\mathcal{L}_{\mathcal{A}}$  is irreducible, and its radical ideal  $I(\mathcal{A})$  is prime.  $\square$

The second ingredient of Theorem 3.3.1 is the Rees algebra of the likelihood module. Recall the construction of  $\mathcal{R}(M(\mathcal{A}))$  from Definition 2.3.1.

**Definition 3.3.6.** Let  $\mathcal{A}$  be an arrangement and  $M(\mathcal{A}) = \text{coker}(A)$  its likelihood module. We call  $I_0(\mathcal{A}) = \langle (s_1, \dots, s_m) \cdot A \rangle$  the *pre-likelihood ideal* of  $\mathcal{A}$ . This is the ideal shown in (2.3.1), presenting the symmetric algebra of  $M(\mathcal{A})$ . Let  $I$  be the kernel of the composition

$$R[s_1, \dots, s_m] \rightarrow \text{Sym}(M(\mathcal{A})) \rightarrow \mathcal{R}(M(\mathcal{A})). \quad (3.3.2)$$

Thus,  $I$  is an ideal in the ring on the left containing the pre-likelihood ideal  $I_0(\mathcal{A})$ . We refer to  $I$  as the *Rees ideal* of the module  $M(\mathcal{A})$  because it presents the Rees algebra of  $M(\mathcal{A})$ .

Theorem 3.3.1 states that the Rees ideal of  $M(\mathcal{A})$  equals the likelihood ideal,  $I = I(\mathcal{A})$ . This will be proved below. The ambient polynomial ring  $R[s] = \mathbb{C}[x_1, \dots, x_n, s_1, \dots, s_m]$  is bigraded via  $\deg(x_i) = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$  for  $i = 1, \dots, n$  and  $\deg(s_i) = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$  for  $i = 1, \dots, m$ . The Rees ideal can be computed with general methods in MACAULAY2. See [Eis18] for a computational introduction. The output of these methods differ from ours as these tools usually work with minimal presentations of modules, thereby reducing the number of variables  $s_i$ . For us, it makes sense to preserve symmetry and accept non-minimal presentations.

The case where the likelihood module is of linear type is of particular interest to us, as then the computation of the likelihood ideal  $I(\mathcal{A})$  is very simple.

**Definition 3.3.7.** An arrangement  $\mathcal{A}$  is *gentle* if its likelihood module is of linear type, i.e. if its likelihood ideal  $I(\mathcal{A})$  equals the pre-likelihood ideal  $I_0(\mathcal{A})$ . This happens if and only if the map on the right in (3.3.2) is an isomorphism, in which case  $\text{Sym}(M(\mathcal{A})) = \mathcal{R}(M(\mathcal{A}))$ .

**Example 3.3.8.** The graphic arrangement of  $K_4$  is gentle. Fix the matrix  $A$  from Example 3.3.2. The pre-likelihood ideal has three generators, one for each nonzero column of  $A$ :

$$I_0(\mathcal{A}) = \langle [s_{12}, s_{13}, s_{14}, s_{23}, s_{24}, s_{34}] \cdot A \rangle \subset R[s_{12}, s_{13}, s_{14}, s_{23}, s_{24}, s_{34}]. \quad (3.3.3)$$

One generator is  $\sum_{ij} s_{ij}$ . The other two generators have bidegrees  $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$  and  $\begin{pmatrix} 2 \\ 1 \end{pmatrix}$ . Using MACAULAY2, we find that the pre-likelihood ideal  $I_0(\mathcal{A})$  is prime. Hence, by Proposition 3.3.10,  $I_0(\mathcal{A})$  equals the Rees ideal of  $M(\mathcal{A})$ , which is the likelihood ideal  $I(\mathcal{A})$ . It defines a complete intersection in  $\mathbb{P}^3 \times \mathbb{P}^5$ . This is the likelihood correspondence  $\mathcal{L}_{\mathcal{A}}$ .  $\diamond$

**Example 3.3.9** ( $n = 3, m = 4$ ). The arrangement  $\mathcal{A} = \{x, y, z, x^3 + y^3 + xyz\}$  is not gentle. It consists of the three coordinate lines and one cubic in  $\mathbb{P}^2$ . Its pre-likelihood ideal equals

$$I_0(\mathcal{A}) = \langle s_1 + s_2 + s_3 + 3s_4, xz \cdot s_2 - (3y^2 + xz) \cdot s_3, yz \cdot s_2 + (3x^2 + 2yz) \cdot s_3 + 3yz \cdot s_4, (x^3 + y^3) \cdot s_2 + (3y^3 + xyz) \cdot s_3 + (3y^3 + xyz) \cdot s_4 \rangle.$$

This ideal is radical but it is not prime. Its prime decomposition equals

$$\begin{aligned} I_0(\mathcal{A}) &= (I_0(\mathcal{A}) + \langle x, y \rangle) \cap I(\mathcal{A}), \quad \text{where } I(\mathcal{A}) = I_0(\mathcal{A}) + \langle q \rangle \\ \text{and } q &= z^2 \cdot s_2^2 + z^2 \cdot s_2 s_3 + 9xy \cdot s_3^2 - 2z^2 \cdot s_3^2 + 3z^2 \cdot s_2 s_4 - 3z^2 \cdot s_3 s_4. \end{aligned}$$

The extra generator  $q$  of the likelihood ideal is quadratic in the data variables  $s$ .  $\diamond$

For hyperplane arrangements, our ideals were introduced by Cohen et al. [CDFV11] who called them the *logarithmic ideal* and the *meromorphic ideal*, respectively. In spirit of Terao's freeness conjecture, one can ask whether gentleness is combinatorial, i.e. can the matroid decide whether an arrangement is gentle? Since all line arrangements in  $\mathbb{P}^2$  are gentle (Theorem 3.3.19), a counterexample would need to have rank at least four.

Our technique for computing likelihood ideals of arrangements rests on the following result. It transforms the pre-likelihood ideal  $I_0$  into the Rees ideal  $I$  via saturation.

**Proposition 3.3.10.** *Let  $p$  be an element in  $R$  such that  $M(\mathcal{A})[p^{-1}]$  is a free  $R[p^{-1}]$ -module. Then the likelihood ideal of the arrangement  $\mathcal{A}$  is the saturation  $I(\mathcal{A}) = (I_0(\mathcal{A}) : p^\infty)$ . In particular, the arrangement  $\mathcal{A}$  is gentle if and only if its pre-likelihood ideal  $I_0(\mathcal{A})$  is prime.*

*Proof.* The proof of the statement about  $p$  uses the fact that the Rees algebra construction commutes with localisation (Proposition 2.3.3). The likelihood ideal  $I(\mathcal{A})$  is always prime, since the Rees algebra is a domain by Proposition 2.3.2. Thus, if  $I_0(\mathcal{A})$  is not prime, then it is not the likelihood ideal and the arrangement  $\mathcal{A}$  is not gentle. If  $I_0(\mathcal{A})$  is prime, then picking any suitable  $p$  in the first part shows that it is the likelihood ideal.  $\square$

**Remark 3.3.11.** The existence of an element  $p$  as in Proposition 3.3.10 is guaranteed by generic freeness. In our case, we can take  $p$  as the product of the  $f_i$  and all maximal nonzero minors of the Jacobian matrix of  $F = (f_1, \dots, f_m)$ . This follows from the construction of the likelihood correspondence. There,  $F(x) \in X_{\text{reg}}$  is required, but the proof of Lemma 3.3.5 requires only that the Jacobian of  $F$  has maximal rank. We can replace  $F(x) \in X_{\text{reg}}$  by this latter condition without changing the closure. Computing the saturation tends to be a horrible computation. For practical purposes, it usually suffices to saturate  $I_0$  at just a few of these polynomials and checking primality after each step. In Example 3.3.9, we can take  $p$  to be any element in the ideal  $\langle x, y \rangle$  for the singular locus of the cubic  $x^3 + y^3 + xyz$ .

*Proof of Theorem 3.3.1.* Let  $I$  be the prime likelihood ideal and  $I_0$  the pre-likelihood ideal of an arrangement  $\mathcal{A}$ . Since the generators of  $I_0$  vanish on the likelihood correspondence  $\mathcal{L}_{\mathcal{A}}$ , we have  $I_0 \subseteq I$ . Let  $I'$  be the Rees ideal of the likelihood module  $M(\mathcal{A})$ . Clearly, also  $I_0 \subseteq I'$  and  $I'$  is prime by Proposition 2.3.2. Let  $p$  be an element as in Proposition 3.3.10, then  $I' = I_0 : p^\infty \subseteq I : p^\infty$ . Since  $p \in R$  does not contain any  $s$  variables,  $p \notin I$ . Hence,  $I : p^\infty = I$  and thus  $I' \subseteq I$ . Conversely, also  $I = I_0 : f$  where  $f$  equals a sufficiently high power of the product of the polynomials cutting out the singular locus of  $X$  and the forms  $f_i$ , another polynomial that is  $s$ -free and no such polynomial vanishes on  $\mathcal{L}_{\mathcal{A}}$ . Hence, also  $I = I_0 : f \subseteq I' : f = I'$  and thus  $I = I'$ .  $\square$

We conclude this subsection with a result directly linking arrangements and likelihood.

**Theorem 3.3.12.** *The evaluation of  $\mathcal{A}$ -derivations at the log-likelihood function*

$$\text{Der}(\mathcal{A}) \rightarrow I(\mathcal{A}) \subset R[s], \quad \theta \mapsto \theta(\ell_{\mathcal{A}})$$

*is an injective  $R$ -linear map onto  $I_0(\mathcal{A})$ . It is an isomorphism if and only if  $\mathcal{A}$  is gentle.*

*Proof.* Any derivation  $\theta$  maps  $\ell_{\mathcal{A}}$  to a rational function in  $\mathbb{C}[s](x)$ . The image is a polynomial in  $\mathbb{C}[s, x]$  if and only if  $\theta \in \text{Der}(\mathcal{A})$ . The isomorphism between  $\text{Der}(\mathcal{A})$  and  $D(\mathcal{A})$  in Lemma 3.3.3 ensures that the map is injective, and that these polynomials generate the ideal  $I_0(\mathcal{A})$ . If  $\mathcal{A}$  is gentle,  $I_0(\mathcal{A}) = I(\mathcal{A})$  and the map is an isomorphism.  $\square$

### 3.3.2. Parametric likelihood in statistics and physics

Our study of hypersurface arrangements offers new tools for statistics and physics. We now explain this point and the relation to implicit likelihood as explained in Section 2.2. Let  $\mathcal{A}$  be an arrangement given by homogeneous polynomials  $f_1, \dots, f_m \in \mathbb{R}[x_1, \dots, x_n]$  of the same degree. The unknowns  $x_1, \dots, x_n$  are model parameters and the polynomials  $f_1, \dots, f_m$  represent probabilities. Let  $X$  denote the Zariski closure of the image of the map

$$F: \mathbb{C}^n \dashrightarrow \mathbb{P}^{m-1}, x \mapsto (f_1(x) : f_2(x) : \dots : f_m(x)).$$

The algebraic variety  $X$  represents a statistical model for discrete random variables. Our model has  $m$  states. The parameter region consists of the points in  $\mathbb{R}^n$  where all  $f_i$  are positive. On that region, the rational function  $f_i / \sum_{j=1}^m f_j$  is the probability of observing the  $i$ th state. In other words, the statistical model is given by the intersection of  $X$  with the probability simplex  $\Delta_{m-1}^\circ$  in  $\mathbb{P}^{m-1}$ . Here, the  $f_i$  are rarely linear, and the  $s_i$  are nonnegative integers which summarise the data. Namely,  $s_i$  is the number of samples that are in state  $i$ .

In statistics, one maximises the log-likelihood function  $\ell_{\mathcal{A}}$  over all points  $x$  of the parameter region. Here, the  $s_i$  are given numbers and one considers the critical equations  $\nabla \ell_{\mathcal{A}} = 0$ . This is a system of rational function equations. Any algebraic approach will transform these into polynomial equations. Naïve clearing of denominators does not work because it introduces too many spurious solutions. The key challenge is to clear denominators in a manner that is both efficient and mathematically sound. That challenge is precisely the point of this section.

A key notion in likelihood geometry is the ML degree (Definition 2.2.8), counting critical points of the likelihood function. We introduce a notion of this in our parametric arrangement setup. The likelihood correspondence  $\mathcal{L}_{\mathcal{A}}$  lives inside  $\mathbb{P}^{n-1} \times \mathbb{P}^{m-1}$ . Its class in the Chow ring  $A(\mathbb{P}^{n-1} \times \mathbb{P}^{m-1}) \cong \mathbb{Z}[p, u]/\langle p^n, u^m \rangle$  is a binary form

$$[\mathcal{L}_{\mathcal{A}}] = c_d p^d + c_{d-1} p^{d-1} u + c_{d-2} p^{d-2} u^2 + \dots + c_1 p u^{d-1} + c_0 u^d, \quad (3.3.4)$$

where  $d = \text{codim}(\mathcal{L}_{\mathcal{A}})$ . This agrees with the *multidegree* of  $I(\mathcal{A})$  as in [MS05, Part II, §8.5].

**Definition 3.3.13.** The (*parametric*) *maximum likelihood* (ML) *degree*  $\text{MLdeg}(\mathcal{A})$  of the arrangement  $\mathcal{A}$  is the leading coefficient of  $[\mathcal{L}_{\mathcal{A}}]$ , i.e. it equals  $c_i$  where  $i$  is the largest index such that  $c_i > 0$  in (3.3.4).

If  $c_d > 0$  then  $\text{MLdeg}(\mathcal{A}) = c_d$  and Definition 3.3.13 gives a critical point count.

**Proposition 3.3.14.** If  $\text{MLdeg}(\mathcal{A}) = c_d$  then the set

$$\left\{ x \in \mathbb{P}^{n-1} : \nabla \ell_{\mathcal{A}}(x, s) = 0, f^s(x) \neq 0, F(x) \in X_{\text{reg}} \right\}, \quad (3.3.5)$$

is finite for generic choices of  $s$ . Its cardinality equals  $\text{MLdeg}(\mathcal{A})$  and does not depend on  $s$ .

*Proof.* Under the assumption  $c_d > 0$ , the projection  $\pi: \mathcal{L}_{\mathcal{A}} \rightarrow \mathbb{P}^{m-1}$  is finite-to-one. A general fibre has cardinality  $c_d$  and is described by (3.3.5).  $\square$

**Remark 3.3.15.** The above setup differs from the one common to algebraic statistics and described in Section 2.2 in several aspects: First, “generic choices of  $s$ ” means generic in a subspace of  $\mathbb{C}^m$ . This is usually  $\{s : \sum_{i=1}^m d_i s_i = 0\}$ . Second, Proposition 3.3.14 gives a *parametric* version of the ML degree, whereas Definition 2.2.8 defines the ML degree *implicitly*. Moreover, in Definition 2.2.8, the hypersurface defined by  $\sum_{i=1}^m f_i$  is added to the arrangement. Only this modification allows the interpretation of  $\mathcal{A}$  as a statistical model,

as described in the paragraph above. If this hypersurface is included in  $\mathcal{A}$  and we assume that the parametrisation is finite-to-one, then our parametric ML degree is an integer multiple of the implicit ML degree. Under these assumptions, there is a flat morphism from the parametric to the implicit likelihood correspondence in Definition 2.2.9. The induced map on Chow rings is injective, and the claim follows. Our definition via the multidegree of  $\mathcal{L}_{\mathcal{A}}$  allows for a sensible notion even in the case where the parametrisation is not finite-to-one. This appears for example in the formulation of toric models given below.

For illustration, we revisit the *coin model* from the introduction of [HKS05].

**Example 3.3.16.** A gambler has two biased coins, one in each sleeve, with unknown biases  $t_2, t_3$ . They select one of them at random, with probabilities  $t_1$  and  $1 - t_1$ , toss that coin four times, and record the number of times heads comes up. If  $p_i$  is the probability of  $i - 1$  appearances of heads then

$$\begin{aligned} p_1 &= t_1 \cdot (1 - t_2)^4 + (1 - t_1) \cdot (1 - t_3)^4, \\ p_2 &= 4t_1 \cdot t_2(1 - t_2)^3 + 4(1 - t_1) \cdot t_3(1 - t_3)^3, \\ p_3 &= 6t_1 \cdot t_2^2(1 - t_2)^2 + 6(1 - t_1) \cdot t_3^2(1 - t_3)^2, \\ p_4 &= 4t_1 \cdot t_2^3(1 - t_2) + 4(1 - t_1) \cdot t_3^3(1 - t_3), \\ p_5 &= t_1 \cdot t_2^4 + (1 - t_1) \cdot t_3^4. \end{aligned} \quad (3.3.6)$$

We homogenise these equations by setting  $t_j = x_j/x_4$  for  $j \in \{1, 2, 3\}$ . Let  $f_i(x)$  be the numerator of  $p_i(t)$  after this substitution. This is a homogeneous polynomial in four variables of degree  $d_i = 5$ . We finally set  $f_6(x) = x_4$  and  $d_6 = 1$ . If we now take  $s_6 = -d_1s_1 - d_2s_2 - \dots - d_5s_5$ , then we are in the setting of Subsection 3.3.1. Namely, we have an arrangement  $\mathcal{A}$  of  $m = 6$  surfaces in  $\mathbb{P}^3$ .

We observe  $N$  iterations of this game and record the outcomes in the data vector  $(s_1, s_2, s_3, s_4, s_5) \in \mathbb{N}^5$ , where  $s_i$  is the number of trials with  $i - 1$  heads. Hence,  $\sum_{i=1}^5 s_i = N$ . Our assignment  $s_6 = -5N$  ensures that  $d_1s_1 + \dots + d_6s_6$  lies in  $I_0(\mathcal{A})$ . The task in statistics is to infer the unknown parameters  $t_1, t_2$  and  $t_3$  from the data  $s_1, \dots, s_5$ . The ML degree of this model is 24. Indeed, the equations  $\nabla \ell_{\mathcal{A}}(x, s) = 0$  have 24 complex solutions  $x = (t, 1) \in \mathbb{P}^4$  for random data  $s_1, s_2, s_3, s_4, s_5$ , provided  $t_1(1 - t_1)(t_2 - t_3) \neq 0$ . In [HKS05] it is reported that the ML degree for this model is 12. The discrepancy of a factor two arises because of the two-to-one parametrisation (3.3.6).

In summary, our projective formulation realises the coin model as an arrangement  $\mathcal{A}$  in  $\mathbb{P}^3$  with  $n = 4, m = 6$ , and  $d_1 = \dots = d_5 = 5$  and  $d_6 = 1$ . The quintics  $f_1, f_2, f_3, f_4, f_5$  have 13, 12, 9, 6, 3 terms respectively. For instance, the homogenisation of  $p_4(t)$  yields

$$f_4(x) = 4(-x_1x_2^4 + x_1x_3^4 + x_1x_2^3x_4 - x_1x_3^3x_4 - x_2^4x_4 + x_3^3x_4^2).$$

The pre-likelihood ideal  $I_0(\mathcal{A})$  has six generators of bidegrees  $(0, 1), (2, 1), (10, 1)$ , and  $(13, 1)$  thrice. The first ideal generator is  $5(s_1 + s_2 + s_3 + s_4 + s_5) + s_6$ , and the second generator is

$$4s_6(x_1x_2 - x_1x_3 + x_3x_4) + 5(s_2 + 2s_3 + 3s_4 + 4s_5)x_4^2. \quad \diamond$$

We now turn to the two-parameter models seen in the Introduction of [CHKS06].

**Example 3.3.17.** Let  $n = 3, m = 5, d_1 = d_2 = d_3 = d_4 = 2$ , and  $d_5 = 1$ . This gives arrangements of four conics and the line at infinity in  $\mathbb{P}^2$ . One very special case is the independence model for two binary random variables, in a homogeneous formulation:

$$f_1 = x_1x_2, f_2 = (x_3 - x_1)x_2, f_3 = x_1(x_3 - x_2), f_4 = (x_3 - x_1)(x_3 - x_2), f_5 = x_3.$$

The arrangement is tame and free (see Subsection 3.3.3), but not gentle; the pre-likelihood ideal has a prime decomposition

$$\langle s_+, s_5, x_3 \rangle \cap \langle 2s_+ + s_5, s_+ x_1 - (s_1+s_3) x_3, s_+ x_2 - (s_1+s_2) x_3, (s_1+s_2) x_1 - (s_1+s_3) x_2 \rangle.$$

Here,  $s_+ = s_1 + s_2 + s_3 + s_4$  is the sample size. The likelihood ideal is the second intersectand above. Its four generators confirm that the ML degree equals one. The likelihood ideal is not a complete intersection since  $\text{codim}(I) = 3$ . For the implicit formulation of the same model see [BCF23, Ex. 2.4].

As in the Introduction of [CHKS06], we compare this with the arrangement given by four random ternary quadrics  $f_1, f_2, f_3, f_4$  and additionally  $f_5 = x_3$ . Such a generic arrangement is tame and gentle. The likelihood ideal equals the pre-likelihood ideal. It is minimally generated by seven polynomials: the linear form  $2(s_1 + s_2 + s_3 + s_4) + s_5$ , four generators of degree  $\binom{6}{1}$ , and two generators of degree  $\binom{7}{1}$ . The bidegree (3.3.4) of the likelihood correspondence  $\mathcal{L}_A \subset \mathbb{P}^4 \times \mathbb{P}^2$  equals  $25p^2 + 6pu + u^2$ . Hence, the ML degree equals 25, as predicted by [CHKS06, Thm. 1].  $\diamond$

We now explain how maximum likelihood estimation on *toric models* can be described in our hypersurface arrangement setup and how it relates to Subsection 3.2.9. Let  $A$  be an integer matrix of size  $n \times (N+1)$ , let  $z \in (\mathbb{C}^*)^{N+1}$ , and let  $X_{A,z} \subseteq \mathbb{P}^N$  be the scaled projective toric variety that arises as the image of the monomial map  $\phi_{A,z}$ . By Proposition 2.1.3, we can assume that  $\phi_{A,z}$  is one-to-one. Let  $y_i$  for  $i = 0, \dots, N$  denote the coordinates of  $\mathbb{P}^N$  and let  $\mathcal{H}$  be the distinguished hyperplane arrangement  $\mathcal{H} = \{y_0, \dots, y_N, y_0 + \dots + y_N\}$ . As explained in Subsection 3.2.9, the map  $\phi_{A,z}$  gives rise to an isomorphism of very affine varieties between  $\mathcal{X}_{f,z}^* = \{x \in (\mathbb{C}^*)^n : f(x) \neq 0\}$  and  $X_{A,z} \setminus \mathcal{H}$ , where  $f$  is the polynomial  $f = \sum_{i=0}^N z_i x^{a_i}$ . By Theorem 3.2.27, the signed Euler characteristic of  $X_{A,z} \setminus \mathcal{H}$  is equal to the number of critical points of the function

$$x_1^{s_1} x_2^{s_2} \dots x_n^{s_n} f^{s_{n+1}}, \quad (3.3.7)$$

for generic values  $s_1, \dots, s_n$  and  $s_{n+1} = -\frac{1}{d}(s_1 + \dots + s_n)$ , where  $d = \deg(f)$ . We can encode this in the arrangement setup by setting  $f_i = x_i$  for  $i = 1, \dots, n = m-1$  and  $f_m = f$ . The likelihood function of this arrangement  $\mathcal{A} = \{x_1, \dots, x_n, f\}$  agrees with (3.3.7). The ML degree of  $X_{A,z}$  is equal to the ML degree of  $\mathcal{A}$ . In situations where  $\phi_{A,z}$  is not one-to-one, the ML degree of  $\mathcal{A}$  is a product of the degree of the fibre with the ML degree of  $X_{A,z}$ . The question of how the ML degree depends on the parameters  $z \in (\mathbb{C}^*)^{N+1}$  brings us back to the topic of Euler stratifications, see Section 3.2.

For the convenience of the reader more akin to the arrangement literature, we would like to emphasise that the setup described above differs from the notion of a toric arrangement in the sense of e.g. [dD15].

One instance with  $n = 3$  was seen in Example 3.3.9. Our representation of a toric model depends on the choice of the parametrisation and so does gentleness of the arrangement  $\mathcal{A}$ . This is one reason why previous work on likelihood geometry emphasised the implicit representation. We illustrate the toric setup with the most basic model in statistics.

**Example 3.3.18.** We revisit the independence model for two binary random variables from Example 2.2.3. It is described by the four parameters  $a_0, a_1, b_0, b_1$  giving rise to probabilities

$$p_{00} = a_0 b_0, \quad p_{01} = a_0 b_1, \quad p_{10} = a_1 b_0, \quad p_{11} = a_1 b_1.$$

This parametrises the Segre surface  $\{p_{00}p_{11} = p_{01}p_{10}\}$  in  $\mathbb{P}^3$ . The model has ML degree 1 (Example 2.2.10). The formulation of this model given in Example 3.3.17 is not gentle.

We can also represent this independence model as a toric model by setting  $n = 4$  and

$$\mathcal{A} = \{a_0, a_1, b_0, b_1, f\} \quad \text{with } f = a_0b_0 + a_0b_1 + a_1b_0 + a_1b_1.$$

This is a gentle arrangement of  $m = 5$  surfaces in  $\mathbb{P}^3$ . Its likelihood ideal equals

$$I(\mathcal{A}) = I_0(\mathcal{A}) = \langle s_1 + s_2 + s_5, s_3 + s_4 + s_5, (b_0 + b_1)s_4 + b_1s_5, (a_0 + a_1)s_2 + a_1s_5 \rangle.$$

The arrangement  $\mathcal{A}$  is an overparametrisation. A minimal toric model would live in the plane  $\mathbb{P}^2$ . For instance, one can choose the arrangement  $\mathcal{A}' = \{x, y, z, xy + xz + yz + z^2\}$ . This arrangement is also gentle. Its multidegree is  $p^2u + 2pu^2 + u^3$ . One can verify  $I_0(\mathcal{A}') = I(\mathcal{A}')$  computationally with the tools described in Subsection 3.3.6.  $\diamond$

It is no surprise that the physics of *scattering amplitudes* is closely connected to hypersurface arrangements, since the connection to Euler characteristics of very affine varieties has already been pointed out in Subsection 3.2.8. We now turn to that setup.

In the CHY model [CHY14] one considers scattering equations on the moduli space  $\mathcal{M}_{0,n}$  of  $n$  labelled points in  $\mathbb{P}^1$ . The *scattering correspondence*, analogous to the likelihood correspondence, appears in [Lam24, Eq. (0.2)], and is studied in detail in [Lam24, § 3]. The formulation in [ST21, Eq. (3)] expresses the positive region  $\mathcal{M}_{0,n}^+$  of  $\mathcal{M}_{0,n}$  as a linear statistical model of dimension  $n-3$  on  $n(n-3)/2$  states. Adding another coordinate for the homogenisation, we have  $m = \binom{n-1}{2}$  in our setup. The ML degree equals  $(n-3)!$ . If the data  $s_1, \dots, s_m$  are real, then all  $(n-3)!$  complex critical points are in fact real by Varchenko's Theorem [ST21, Prop. 1]. The case  $n = 6$  is worked out in [ST21, Ex. 2]. This model has  $m-1 = 9$  states and the ML degree is six. The nine probabilities  $p_i$  are given in [ST21, Eq. (6)]. These  $p_i$  sum to 1 and all six critical points in [ST21, Eq. (9)] are real.

Usually, we think of  $\mathcal{M}_{0,n}$  as the set of points for which the  $2 \times 2$  minors of the matrix

$$\begin{bmatrix} 0 & 1 & 1 & \dots & 1 & 1 \\ -1 & 0 & y_1 & \dots & y_{n-3} & 1 \end{bmatrix}$$

are nonzero. If we homogenise the resulting equations by considering the  $2 \times 2$  minors of

$$\begin{bmatrix} 0 & 1 & 1 & \dots & 1 & 1 \\ -1 & x_1 & x_2 & \dots & x_{n-2} & x_{n-1} \end{bmatrix},$$

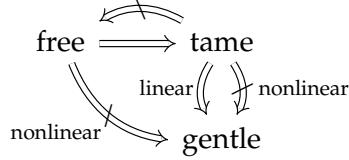
then  $\mathcal{M}_{0,n}$  becomes the complement of the braid arrangement. This is the graphic arrangement of the complete graph  $K_{n-1}$  on  $n-1$  vertices (see Section 2.4), defined by the  $\binom{n-1}{2}$  linear forms  $x_i - x_j$  for  $1 \leq i < j \leq n$ .

For example,  $\mathcal{M}_{0,5}$  can be viewed as the complement of the arrangement in Example 3.3.2. In this case, the image of the likelihood correspondence in  $\mathbb{P}^2 \times \mathbb{P}^5$  under the map to data space  $\mathbb{P}^5$  is the hyperplane  $\{s_{12} + s_{13} + s_{14} + s_{23} + s_{24} + s_{34} = 0\}$ . This map is two-to-one. By [ST21, § 2], the fibres are the two solutions to the scattering equations in the CHY model for five particles. A similar identification works for every graphic arrangement, when some edges of  $K_{n-1}$  are deleted. Physically, this corresponds to setting some Mandelstam invariants to zero. The article [EPS24] studies graphic arrangements of ML degree one from a physics perspective. For instance, in [EPS24, Ex. 1.3], the authors study  $K_5$  with three edges removed.

### 3.3.3. Gentle, free and tame arrangements

The concept of freeness (Definition 2.4.4) has received considerable attention in the theory of hyperplane arrangements, see e.g. [OT13, Thm. 4.15]. Also, the notion of tameness

(Definition 2.4.5) appeared in this context as a weakening of the freeness property. In this subsection we explore the relationship between these concepts and the gentleness of an arrangement. We shall explain the following (non)implications:



By Lemma 3.3.3,  $\mathcal{A}$  is free if and only if the likelihood module  $M(\mathcal{A})$  has projective dimension one if and only if  $D(\mathcal{A})$  is free. Clearly, every free arrangement is tame. The braid arrangement from Example 3.3.2 is free. We have already seen that the braid arrangement is also gentle. This holds more generally.

**Theorem 3.3.19.** *Tame linear arrangements are gentle.*

*Proof.* The statement follows from [CDFV11, Cor. 3.8] and Proposition 3.3.10. The ideal  $I$  in [CDFV11] is our pre-likelihood ideal  $I_0(\mathcal{A})$ , and their variety  $\bar{\Sigma}$  is our parametric likelihood correspondence  $\mathcal{L}_{\mathcal{A}}$ .  $\square$

In  $\mathbb{P}^2$ , every linear arrangement is free, as a direct application of Saito's criterion (Theorem 2.4.6), see [OT13, Ex. 4.20]. Thus, every linear arrangement in  $\mathbb{P}^2$  is gentle. Although freeness is a strong property for an arrangement, for hypersurfaces it does not necessarily imply gentleness. We saw a free arrangement that is not gentle in Example 3.3.17. We do not know whether the reverse implication " $\text{gentle} \Rightarrow \text{tame}$ " holds. To the best of our knowledge, this is unknown even for the linear case; see the Introduction of [CDFV11].

**Problem 3.3.20.** Is every gentle arrangement tame?

For a linear arrangement, freeness is equivalent to the (pre-)likelihood ideal being a complete intersection [CDFV11, Thm. 2.13]. As Example 3.3.17 shows, this is not necessarily true for hypersurfaces. However, under the additional assumption that  $\mathcal{A}$  is gentle, we can generalise [CDFV11, Thm. 2.13]. This connects to [HS14] where the authors ask for a characterisation of statistical models whose likelihood ideal is a complete intersection.

**Theorem 3.3.21.** *Let  $\mathcal{A}$  be a gentle arrangement of hypersurfaces. Then  $\mathcal{A}$  is free if and only if the likelihood ideal  $I(\mathcal{A})$  is a complete intersection.*

*Proof.* Suppose  $\mathcal{A}$  is free of rank  $l$ , i.e. the log-derivation module  $D(\mathcal{A})$  is a free module with generators  $\{D_1, \dots, D_l\}$ . These generators form the columns of the matrix  $A$  from Subsection 3.3.1. Consequently, the pre-likelihood ideal  $I_0(\mathcal{A})$  has  $l$  generators. By assumption,  $\mathcal{A}$  is gentle, so  $I_0(\mathcal{A}) = I(\mathcal{A})$ . Since  $\mathcal{L}_{\mathcal{A}}$  has codimension  $l$ , this shows that  $I(\mathcal{A})$  is a complete intersection.

Conversely, assume  $I(\mathcal{A})$  has  $l$  generators  $g_1, \dots, g_l$ . Similarly to Theorem 3.3.12, for  $1 \leq i \leq l$ , let  $\theta_i \in \text{Der}_S(\mathcal{A})$  be a derivation for which  $\theta_i(\ell_{\mathcal{A}}) = g_i$ . Here,  $S = \mathbb{C}[s_1, \dots, s_m]$  and  $\text{Der}_S(\mathcal{A})$  is the module of  $S$ -linear logarithmic derivations on  $S \otimes_{\mathbb{C}} R$ . The module  $\text{Der}_S(\mathcal{A})$  is generated by the  $\theta_i$  and has rank  $l$ , hence it is free. By extension of scalars,

$$\Omega_{R/\mathbb{C}}^1(\mathcal{A}) \otimes_R (S \otimes_{\mathbb{C}} R) \cong \Omega_{S \otimes R/S}^1(\mathcal{A}),$$

and  $\Omega_{S \otimes R/S}^1(\mathcal{A})$  is dual to  $\text{Der}_S(\mathcal{A})$ . Then, by tensor-hom adjunction, we obtain

$$\begin{aligned} \text{Der}_S(\mathcal{A}) &\cong \text{Hom}_{S \otimes R}(\Omega_{R/\mathbb{C}}^1(\mathcal{A}) \otimes_R (S \otimes_{\mathbb{C}} R), S \otimes_{\mathbb{C}} R) \\ &\cong \text{Hom}_R(\Omega_{R/\mathbb{C}}^1(\mathcal{A}), \text{Hom}_{S \otimes R}(S \otimes_{\mathbb{C}} R, S \otimes_{\mathbb{C}} R)) \\ &\cong \text{Hom}_R(\Omega_{R/\mathbb{C}}^1(\mathcal{A}), S \otimes_{\mathbb{C}} R). \end{aligned}$$

Since  $\Omega_{R/\mathbb{C}}^1(\mathcal{A}) = \Omega^1(\mathcal{A})$  is finitely presented and  $S \otimes_{\mathbb{C}} R$  is faithfully flat, it follows that  $\text{Der}(\mathcal{A}) = \text{Hom}(\Omega^1(\mathcal{A}), R)$  is free.  $\square$

In the case of a free and gentle arrangement, it is now easy to read off the ML degree.

**Corollary 3.3.22.** *Let  $\mathcal{A}$  be free and gentle. If the columns of  $A$  have degrees  $d_1, \dots, d_l$  then*

$$\text{MLdeg}(\mathcal{A}) = \prod_{i:d_i>0} d_i. \quad (3.3.8)$$

*Proof.* By definition, the ML degree is the leading coefficient in the multidegree of  $I(\mathcal{A})$ . Since  $\mathcal{A}$  is free and gentle, by Theorem 3.3.21, the likelihood ideal is a complete intersection, and it is linear in the  $s$  variables. Therefore, the class in (3.3.4) is the product

$$[\mathcal{L}_{\mathcal{A}}] = \prod_{i=1}^l (d_i p + u).$$

Our assertion now follows because (3.3.8) is the leading coefficient of this binary form.  $\square$

**Example 3.3.23.** For the braid arrangement in Example 3.3.2, the matrix  $A^T$  has two rows of positive degree, namely one and two. Hence, by (3.3.8),  $\text{MLdeg}(\mathcal{A}) = 1 \cdot 2 = 2$ . Analogously, for general  $n$ , the braid arrangement  $\mathcal{A}(K_n)$  has ML degree  $(n - 3)!$ , as already stated in our physics discussion about  $\mathcal{M}_{0,n}$  in Subsection 3.3.2.  $\diamond$

Symmetric algebras and Rees algebras are ubiquitous in commutative algebra. Many papers studied them, especially when  $M$  has a short resolution. The *Fitting ideals* of  $M$  play an essential role. Let  $I_t(A)$  be the ideal generated by the  $t \times t$ -minors of a matrix  $A \in R^{m \times l}$  with  $M = \text{coker}(A)$ . These ideals are independent of the presentation of  $M$  [Eis13, § 20.2].

Early work of Huneke [Hun81, Thm. 1.1] characterises when the symmetric algebra of a module  $M$  with  $\text{pd}(M) = 1$  is a domain, and thus when a free arrangement is gentle. This happens if and only if  $\text{depth}(I_t(A), R) \geq \text{rk}(A) + 2 - t$  for all  $t = 1, \dots, \text{rk}(A)$ . Huneke also showed that in this case the symmetric algebra is a complete intersection, one direction of our Theorem 3.3.21. Simis and Vasconcelos [SV81] obtained similar results concurrently.

In the 40+ years since these publications, many variants have been found. For example, it was studied for which  $k$  all inequalities  $\text{depth}(I_t(A)) \geq \text{rk}(A) + (1 + k) - t$  hold. Then  $M$  is said to have *property  $\mathcal{F}_k$* . Assuming  $\mathcal{F}_k$  and related hypotheses, properties (e.g. Cohen-Macaulay) of symmetric and Rees algebras of modules were studied.

A notable special case arises if the double dual  $M^{\vee\vee}$  of a module  $M$  is free. In [SUV03, § 5] such an  $M$  is called an *ideal module* because it behaves very much like an ideal. Every ideal module  $M$  is the image of a map of free modules, and various criteria for gentleness (i.e. linear type) of  $M$  can be derived. These might give rise to more efficient computational tests for gentleness. For example, the likelihood module of the octahedron in Example 3.3.37 is an ideal module.

### 3.3.4. Generic arrangements

In this subsection we study the pre-likelihood ideal of arrangements of generic hypersurfaces. Though less important for a statistical application, this is a very natural setting in algebraic geometry. For hyperplanes, genericity means that the associated matroid is uniform. Such arrangements are known to be tame and hence gentle, by Theorem 3.3.19. We conjecture that this holds true for general hypersurface arrangements.

**Conjecture 3.3.24.** Any generic hypersurface arrangement is gentle.

The goal of this subsection is to work towards this conjecture. The main results are that generic hypersurface arrangements are tame (Corollary 3.3.33), and a resolution of the module  $\text{Der}(\mathcal{A})$ , which gives access to the number of generators of  $I_0(\mathcal{A})$  (Theorem 3.3.30). Moreover, we show that  $I_0(\mathcal{A})$  is Cohen–Macaulay for generic arrangements.

We first construct a minimal free resolution of  $\text{Der}(\mathcal{A})$ . By Proposition 2.4.3,  $\text{Der}(\mathcal{A})$  contains the Euler derivation as a direct summand, i.e.  $\text{Der}(\mathcal{A}) = \text{Der}_0(\mathcal{A}) \oplus R\theta_E$ . It suffices to construct a resolution of  $\text{Der}_0(\mathcal{A})$ . The relation with the matrix  $Q$  from (3.3.1) is as follows: the cokernel of  $Q^T$  is the module of logarithmic differential forms  $\Omega^1(\mathcal{A}) = \text{coker}(Q^T)$ . The module  $\text{Der}(\mathcal{A})$  is its dual,  $\Omega^1(\mathcal{A}) = \text{Hom}(\text{Der}(\mathcal{A}), R)$ . Also,  $\Omega^1(\mathcal{A})$  contains a direct summand dual to the Euler derivation; we write  $\Omega^1(\mathcal{A}) = \Omega_0^1(\mathcal{A}) \oplus R\omega_E$ . The module  $\Omega_0^1(\mathcal{A})$  can be obtained as the cokernel of the transposed *pruned* matrix

$$Q_0 = \begin{pmatrix} 0 & \dots & 0 & \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \dots & \frac{\partial f_1}{\partial x_n} \\ f_2 & \dots & 0 & \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \dots & \frac{\partial f_2}{\partial x_n} \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & f_m & \frac{\partial f_m}{\partial x_1} & \frac{\partial f_m}{\partial x_2} & \dots & \frac{\partial f_m}{\partial x_n} \end{pmatrix},$$

which is constructed from  $Q$  by deleting the first column. We describe how to obtain a resolution of  $\text{Der}_0(\mathcal{A})$  from  $Q_0$  in Theorem 3.3.30. First, we need some preliminary results.

**Lemma 3.3.25.** The Fitting ideal  $I_m(Q_0)$  has depth at least  $n$ .

*Proof.* The Fitting ideal  $I_m(Q_0)$  is generated by the maximal minors of  $Q_0$ . We need to show that there exists a regular sequence in  $I_m(Q_0)$  of length  $n$ . For this, we can take the sequence defined by the minors of the first  $k - 1$  columns and one additional column among the last  $n$  columns of  $Q_0$ . Such a minor is of the form  $\pm \partial_{x_i} f_1 \prod_{j=2}^k f_j$ . By genericity of the  $f_j$ , these form a regular sequence of length  $n$ .  $\square$

**Lemma 3.3.26.** The sheaf  $\widetilde{\Omega_0^1(\mathcal{A})}$  associated to  $\Omega_0^1(\mathcal{A})$  is locally free.

*Proof.* Consider the two short exact sequences

$$0 \rightarrow \text{Der}_0(\mathcal{A}) \rightarrow R^n \rightarrow J(f) \rightarrow 0,$$

where  $J(f)$  denotes the Jacobian ideal  $J(f) = \left\langle \frac{\partial f}{\partial x_1}, \dots, \frac{\partial f}{\partial x_n} \right\rangle$ , and

$$0 \rightarrow J(f) \rightarrow R \rightarrow R/J(f) \rightarrow 0.$$

Applying  $\text{Hom}(-, R)$  to both of these sequences yields

$$\text{Ext}^i(\text{Der}_0(\mathcal{A}), R) \cong \text{Ext}^{i+2}(R/J(f), R). \quad (3.3.9)$$

Let  $\mathfrak{p} \in \text{Spec}(R)$  be a prime ideal that is not the irrelevant ideal. Then, for all  $i > 0$ ,

$$\left( \text{Ext}^{i+2}(R/J(f), R) \right)_{\mathfrak{p}} = \text{Ext}^{i+2}((R/J(f))_{\mathfrak{p}}, R_{\mathfrak{p}}) = 0,$$

as  $J(f)$  is supported in codimension two (using genericity of  $\mathcal{A}$ ). Therefore, by (3.3.9),

$$\left( \text{Ext}^i(\text{Der}_0(\mathcal{A}), R) \right)_{\mathfrak{p}} = 0 \quad \text{for all } i > 0$$

and  $\widetilde{\text{Der}_0(\mathcal{A})}$  is locally free. This implies local freeness of  $\widetilde{\Omega_0^1(\mathcal{A})}$ .  $\square$

The statement of Lemma 3.3.26 is slightly weaker than requiring  $\Omega_0^1(\mathcal{A})$  to be locally free as a module as this would also need the localisation at the irrelevant ideal to be free.

**Lemma 3.3.27.** *For a generic arrangement  $\mathcal{A}$ , the module  $\Omega^p(\mathcal{A})$  is given by*

$$\Omega^p(\mathcal{A}) \cong \bigwedge^p \Omega^1(\mathcal{A}).$$

*Proof.* We first show the identity

$$\Omega^p(\mathcal{A}) \cong \left( \bigwedge^p \Omega^1(\mathcal{A}) \right)^{\vee\vee}. \quad (3.3.10)$$

In the hyperplane case, this has appeared as Proposition 2.2 in [DS09]. The proof works similarly for hypersurfaces, as we now show. Let  $j_p: \Lambda^p \Omega^1(\mathcal{A}) \rightarrow \Omega^p(\mathcal{A})$  be the canonical inclusion from Proposition 2.4.3 and let  $E^p(\mathcal{A})$  denote its cokernel. This module is supported on the singular locus of  $\mathcal{A}$  and hence in codimension at least two (by genericity of  $\mathcal{A}$ ). Therefore,  $\text{Hom}_R(E^p, R) = \text{Ext}_R^1(E^p, R) = 0$ , and it follows that the dual map

$$j_p^\vee: \Omega^p(\mathcal{A})^\vee \rightarrow \left( \bigwedge^p \Omega^1(\mathcal{A}) \right)^\vee$$

is an isomorphism. Since  $\Omega^p(\mathcal{A})$  is reflexive by Proposition 2.4.3, the isomorphism (3.3.10) follows from dualising again.

As  $\widetilde{\Omega^1(\mathcal{A})}$  is locally free by Lemma 3.3.26, the double dualising is not necessary.  $\square$

**Definition 3.3.28.** The *pruned module of logarithmic p-forms* is  $\Omega_0^p(\mathcal{A}) := \Lambda^p \Omega_0^1(\mathcal{A})$ .

**Lemma 3.3.29.** *The module  $\text{Der}_0(\mathcal{A})$  is isomorphic to a pruned module of logarithmic forms*

$$\text{Der}_0(\mathcal{A}) \cong \Omega_0^{n-2}(\mathcal{A}).$$

*Proof.* By Item 3 in Proposition 2.4.3, there is an isomorphism  $\text{Der}(\mathcal{A}) \cong \Omega^{n-1}(\mathcal{A})$ . Using Lemma 3.3.27 it follows that

$$\text{Der}(\mathcal{A}) \cong \bigwedge^{n-1} \left( \Omega_0^1(\mathcal{A}) \oplus R\omega_E \right) \cong \bigoplus_{k=0}^{n-1} \left( \bigwedge^k \Omega_0^1(\mathcal{A}) \otimes \bigwedge^{n-1-k} R\omega_E \right) \cong \bigwedge^{n-1} \Omega_0^1(\mathcal{A}) \oplus \bigwedge^{n-2} \Omega_0^1(\mathcal{A}).$$

Since  $\text{rk}(\Omega_0^1(\mathcal{A})) = n - 1$  and  $\text{Der}(\mathcal{A}) = \text{Der}_0(\mathcal{A}) \oplus R\theta_E$ , the desired statement follows.  $\square$

We are now ready to construct a (minimal) free resolution of  $\text{Der}_0(\mathcal{A})$ .

**Theorem 3.3.30.** *Let  $\mathcal{A} = \{f_1, \dots, f_m\}$  be a generic arrangement and assume  $\deg(f_1) = \dots = \deg(f_m) \geq 2$ . Then the module  $\text{Der}_0(\mathcal{A})$  has the following minimal free resolution:*

$$0 \rightarrow \text{Sym}^{n-2} R^m \rightarrow \dots \rightarrow \text{Sym}^k R^m \otimes \bigwedge^{n-2-k} R^{m+n-1} \rightarrow \dots \rightarrow \bigwedge^{n-2} R^{m+n-1} \rightarrow \text{Der}_0(\mathcal{A}) \rightarrow 0 \quad (3.3.11)$$

The differential is given by

$$\begin{aligned} d: \text{Sym}^k R^m \otimes \bigwedge^{n-2-k} R^{m+n-1} &\rightarrow \text{Sym}^{k-1} R^m \otimes \bigwedge^{n-2-k+1} R^{m+n-1} \\ (a_1 \otimes \dots \otimes a_k) \otimes \omega &\mapsto \sum_{i=1}^k (a_1 \otimes \dots \otimes \widehat{a_i} \otimes \dots \otimes a_k) \otimes (Q_0^T(a_i) \wedge \omega), \end{aligned} \quad (3.3.12)$$

and the map  $\bigwedge^{n-2} R^{m+n-1} \rightarrow \text{Der}_0(\mathcal{A})$  comprises the isomorphism from Lemma 3.3.29.

**Remark 3.3.31.** The complex (3.3.11) is in fact a graded complex. This grading is obtained as follows. Let  $e$  denote the degree of  $f_1, \dots, f_m$ . Then the matrix  $Q_0$  turns the codomain  $R^{m+n-1}$  into a graded free module

$$R^{m+n-1} = R(e)^{m-1} \oplus R(e-1)^n.$$

Considering  $R^m$  to have zero degrees, this induces a grading on the modules  $\text{Sym}^k R^m \otimes \wedge^{n-2-k} R^{m+n-1}$ . In particular, the differential  $d$  in (3.3.12) is homogeneous of degree zero.

*Proof.* Since the arrangement  $\mathcal{A}$  is generic,

$$0 \rightarrow R^m \xrightarrow{Q_0^T} R^{m+n-1} \rightarrow \Omega_0^1(\mathcal{A}) \rightarrow 0 \quad (3.3.13)$$

is a free resolution of  $\Omega_0^1(\mathcal{A})$ . By Lemma 3.3.29, a free resolution of  $\text{Der}_0(\mathcal{A})$  is given by a free resolution of the  $(n-2)^{\text{nd}}$  exterior power of  $\Omega_0^1(\mathcal{A})$ . Such a complex is constructed by Lebelt in [Leb77, Beispiel (ii)], giving rise to the complex (3.3.11). Exactness of this complex follows from [Leb74, Satz 2], together with Lemma 3.3.25. Therefore, (3.3.11) is a free resolution of  $\Omega_0^{n-2}(\mathcal{A})$ . It remains to show minimality. By Remark 3.3.31, the complex (3.3.11) is graded and, since  $\deg(f_i) \geq 2 \forall i = 1, \dots, m$ , all entries of  $Q_0$  are contained in the maximal ideal of  $R$ . Hence, by Lemma 2.3.4, the resolution is minimal.  $\square$

**Example 3.3.32.** Consider the case of five generic quadrics in  $\mathbb{P}^3$ , i.e. we take  $n = 4$  and  $m = 5$ . Using the computational techniques presented in Subsection 3.3.6, we can compute a resolution of the log-derivation module  $D(\mathcal{A}) \cong \text{Der}(\mathcal{A})$  via the following commands (we do this over a finite field to speed up the computation).

```
R = ZZ/10007[x_1..x_4];
A = toList(1..5) / (i -> random(2, R));
D = logDerModule A;
betti res D
```

The result shows that we have a minimal free resolution

$$0 \rightarrow R^{15} \rightarrow R^{40} \rightarrow R^{29} \rightarrow D(\mathcal{A}) \rightarrow 0.$$

The ranks agree with the numbers  $\binom{m+k-1}{k} \binom{m+n-1}{n-2-k}$  we obtain from Theorem 3.3.30, except for  $29 = \binom{4+5-1}{4-2} + 1$ , where the  $+1$  comes from the additional Euler derivation.  $\diamond$

As a consequence of the proof of Theorem 3.3.30 we also obtain the following result.

**Corollary 3.3.33.** *Any generic arrangement  $\mathcal{A}$  is tame.*

*Proof.* Applying Lebelt's construction [Leb77, Beispiel (ii)] to the resolution (3.3.13) yields free resolutions of length  $p$  for any  $p^{\text{th}}$  exterior power of  $\Omega_0^1(\mathcal{A})$ . By Lemma 3.3.27, this implies  $\text{pd}_R(\Omega^p(\mathcal{A})) \leq p$ .  $\square$

A further consequence of Theorem 3.3.30 is that we obtain the numbers and degrees of minimal generators of the pre-likelihood ideal  $I_0(\mathcal{A})$ . The relation to the pre-likelihood ideal can be seen as follows. As before, let  $\begin{pmatrix} A \\ B \end{pmatrix} \in R^{(m+n) \times l}$  be a matrix whose columns generate  $\ker(Q)$ , so that  $I_0(\mathcal{A})$  is generated by the entries of  $(s_1, \dots, s_m) \cdot A$ . Therefore, to obtain the number of minimal generators of  $I_0(\mathcal{A})$  we need to know the number of minimal generators of  $\ker(Q)$ . It suffices to consider  $Q_0$  since the Euler derivation simply contributes a linear generator in the  $s$  variables. Dualising the sequence (3.3.13), we get

$$0 \rightarrow \text{Hom}(\Omega_0^1(\mathcal{A}), R) \rightarrow R^{m+n-1} \xrightarrow{Q_0} R^m \rightarrow \text{Ext}_R^1(\Omega_0^1(\mathcal{A}), R) \rightarrow 0,$$

which is again exact. The leftmost nonzero term is precisely  $\text{Der}_0(\mathcal{A})$ . Therefore,  $\ker(Q_0) = \text{im}(\text{Der}_0(\mathcal{A}) \hookrightarrow R^{m+n-1})$ , so the numbers of minimal generators of  $\ker(Q_0)$  and  $\text{Der}_0(\mathcal{A})$  agree. The latter can be read off from Theorem 3.3.30 to be  $\binom{m+n-1}{n-2}$ . Therefore,  $I_0(\mathcal{A})$  is minimally generated by  $\binom{m+n-1}{n-2} + 1$  many elements. We can make a more precise statement by considering the grading in Remark 3.3.31.

**Corollary 3.3.34.** *Assume that  $\mathcal{A} = \{f_1, \dots, f_m\}$  is generic and all polynomials have the same degree  $\deg(f_i) \geq 2$ . Then there are  $\binom{m-1}{n-2}$  minimal generators of  $I_0(\mathcal{A})$  of minimal degree (not considering the linear form corresponding to the Euler derivation).*

*Proof.* Let  $e = \deg(f_1)$ . The generators of minimal degree come from the minimal degree generators of  $\bigwedge^{n-2} R^{m+n-1}$  in the complex (3.3.11). The module  $R^{m+n-1}$  splits as  $R(e)^{m-1} \oplus R(e-1)^n$ , see Remark 3.3.31. Therefore, we have

$$\bigwedge^{n-2} R^{m+n-1} \cong \bigoplus_{k=0}^{n-2} \left( \bigwedge^k R(e)^{m-1} \otimes \bigwedge^{n-2-k} R(e-1)^n \right).$$

The summand with the lowest degree is obtained for  $k = 0$  and is  $\bigwedge^{n-2} R(e)^{m-1}$ . It has rank  $\binom{m-1}{n-2}$  and thus the claim follows.  $\square$

**Example 3.3.35.** In the setup of Example 3.3.32, we can compute that  $I_0(\mathcal{A})$  has 29 generators. One of them is linear in the  $s$  variables, there are 6 =  $\binom{5-1}{4-2}$  generators of degree six in the  $x$  variables, 16 of degree seven and six of degree eight.  $\diamond$

We now come back to the case where  $\mathcal{A} = \{f_1, \dots, f_m\}$  is any generic arrangement. A further step towards proving Conjecture 3.3.24 is the following statement.

**Proposition 3.3.36.** *For a generic  $\mathcal{A}$ , the pre-lielihood ideal  $I_0(\mathcal{A})$  is Cohen–Macaulay.*

The proof is similar to the proof of [CDFV11, Thm. 3.7] showing the statement above for tame linear arrangements, though the genericity assumption simplifies our proof a bit.

*Proof.* Firstly, we show that the following complex of  $(S \otimes R)$ -modules is exact:

$$0 \rightarrow S \otimes R \xrightarrow{\partial} \Omega_{S \otimes R/R}^1(\mathcal{A}) \xrightarrow{\partial} \dots \xrightarrow{\partial} \Omega_{S \otimes R/R}^n(\mathcal{A}) \rightarrow (S \otimes R)/I_0 \rightarrow 0. \quad (3.3.14)$$

Here, the differential  $\partial$  is given by  $\partial(\omega) = \omega_s \wedge \omega$ , where  $\omega_s$  is the logarithmic one-form

$$\omega_s = \sum_{i=1}^m s_i \frac{df_i}{f_i}.$$

We first show exactness at  $\Omega_{S \otimes R/R}^n(\mathcal{A})$ . For this, note that there is an isomorphism

$$S \otimes R \xrightarrow{\sim} \Omega_{S \otimes R/R}^n(\mathcal{A}), \quad 1 \mapsto \frac{1}{f} dx_1 \wedge \dots \wedge dx_n.$$

Moreover, we have  $\Omega_{S \otimes R/R}^{n-1}(\mathcal{A}) \cong \text{Der}_S(\mathcal{A})$  by Proposition 2.4.3. Under these identifications, exactness at  $\Omega_{S \otimes R/R}^n(\mathcal{A})$  follows. To show exactness at all other positions, we do this locally at all maximal ideals  $\mathfrak{m} \subset S \otimes R$ . First assume that  $\mathfrak{m}$  is not the irrelevant ideal. By Lemma 3.3.26,  $\Omega_{S \otimes R/R}^p(\mathcal{A})_{\mathfrak{m}}$  is a free  $(S \otimes R)_{\mathfrak{m}}$ -module for all  $0 \leq p \leq n$ . Then the complex (3.3.14) becomes a Koszul-like complex and exactness follows. To show exactness locally at the irrelevant ideal, we apply Corollary 3.3.33 together with Lemma 3.11 from [CDFV11] which does not require the arrangement to be linear.

By the Auslander–Buchsbaum formula,  $I_0$  is Cohen–Macaulay if and only if  $\text{pd}_{S \otimes R}((S \otimes R)/I_0) = \text{codim}((S \otimes R)/I_0)$ . Since the likelihood correspondence has codimension  $n$ , we have  $\text{codim}((S \otimes R)/I_0) = n$  (confer Lemma 3.3.5). It remains to show  $\text{pd}_{S \otimes R}((S \otimes R)/I_0) \leq n$ . Consider the complex  $\Omega_{S \otimes R/R}^{n-\bullet}(\mathcal{A})$  obtained from (3.3.14) by replacing the last map by zero, i.e. this complex is exact except for the zeroth homology which is  $(S \otimes R)/I_0$ . Let  $\mathfrak{p} \in \text{Spec}((S \otimes R)/I_0)$  and let  $\kappa = (S \otimes R)_{\mathfrak{p}}/\mathfrak{p}(S \otimes R)_{\mathfrak{p}}$  be the residue field. The first hypertor spectral sequence, Theorem 2.3.6, applied to the localised complex at  $\mathfrak{p}$  gives

$$E_{pq}^1 = \text{Tor}_q^{(S \otimes R)_{\mathfrak{p}}}(\Omega_{S \otimes R/R}^{n-p}(\mathcal{A})_{\mathfrak{p}}, \kappa) \Rightarrow \text{Tor}_{p+q}^{(S \otimes R)_{\mathfrak{p}}}(\Omega_{S \otimes R/R}^{n-\bullet}(\mathcal{A})_{\mathfrak{p}}, \kappa). \quad (3.3.15)$$

As  $\Omega_{S \otimes R/R}^{n-\bullet}(\mathcal{A})_{\mathfrak{p}}$  has only nontrivial homology in degree zero, we have

$$\text{Tor}_{p+q}^{(S \otimes R)_{\mathfrak{p}}}(\Omega_{S \otimes R/R}^{n-\bullet}(\mathcal{A})_{\mathfrak{p}}, \kappa) \cong \text{Tor}_{p+q}^{(S \otimes R)_{\mathfrak{p}}}(((S \otimes R)/I_0)_{\mathfrak{p}}, \kappa). \quad (3.3.16)$$

By Corollary 3.3.33, the arrangement  $\mathcal{A}$  is tame, i.e.  $\text{Tor}_q^{(S \otimes R)_{\mathfrak{p}}}(\Omega_{S \otimes R/R}^{n-p}(\mathcal{A})_{\mathfrak{p}}, \kappa) = 0$  for  $p + q > n$ . Then, combining (3.3.15) and (3.3.16), this implies  $\text{Tor}_{p+q}^{(S \otimes R)_{\mathfrak{p}}}(((S \otimes R)/I_0)_{\mathfrak{p}}, \kappa) = 0$  for  $p + q > n$  and all primes  $\mathfrak{p} \in \text{Spec}((S \otimes R)/I_0)$ . Therefore,  $\text{pd}_{S \otimes R}((S \otimes R)/I_0) \leq n$ .  $\square$

In the next subsection we turn to a different case of arrangements, namely hyperplane arrangements with structure encoded by graphs. Those arise in particle physics.

### 3.3.5. Graphic arrangements

Recall the notion of graphic arrangements from Section 2.4. They are subarrangements of the braid arrangement. In particle physics [EPS24, Lam24] they arise from the moduli space  $\mathcal{M}_{0,n}$ . Throughout this subsection, let  $G = (V, E)$  be a simple undirected graph with  $V = \{1, \dots, n\}$ , let  $R = \mathbb{C}[x_1, \dots, x_n]$  and let  $\mathcal{A}(G) = \{x_i - x_j : \{i, j\} \in E\}$  be the graphic arrangement. A classical result due to Stanley, Edelman and Reiner states that  $\mathcal{A}(G)$  is free if and only if the graph  $G$  is chordal (see [AKMM23] for further developments). For example, the complete graph  $G = K_4$  in Example 3.3.2 is chordal, and we saw that  $D(\mathcal{A}(K_4)) \cong R^3$ . The octahedron in Example 3.3.37 is not chordal.

In this subsection, we examine gentleness for graphic arrangements. A priori, it is not clear that there exist graphs whose arrangement is not gentle. We now show this.

**Example 3.3.37** (Octahedron). Consider the edge graph  $G$  of an octahedron, depicted in Figure 3.8. The graphic arrangement  $\mathcal{A}(G)$  consists of the twelve hyperplanes

$$x_1 - x_2, x_1 - x_3, x_1 - x_5, x_1 - x_6, x_2 - x_3, x_2 - x_4, x_2 - x_6, x_3 - x_4, x_3 - x_5, x_4 - x_5, x_4 - x_6, x_5 - x_6.$$

The likelihood module has twelve generators and six relations, of degrees one, two and three (four times), in addition to the Euler relation of degree zero. These relations correspond to the seven generators of the pre-likelihood ideal  $I_0$ . A computation in MACAULAY2 shows that  $I_0 : (x_1 - x_2) \neq I_0$ , so Proposition 3.3.10 tells us that  $\mathcal{A}(G)$  is not gentle. Another computation shows that the ideal quotient  $I = I_0 : (x_1 - x_2)$  is a prime ideal, and it hence equals the likelihood ideal  $I(\mathcal{A}(G))$ . The ideal  $I$  differs from  $I_0$  by only one additional generator  $f \in R$  of degree  $\binom{3}{3}$  with 3092 terms. Computing  $P = I_0 : f$  reveals the second minimal prime of the pre-likelihood ideal  $I_0$ , and we obtain the prime decomposition

$$I_0 = I \cap P \quad \text{where} \quad P = \left\langle \sum_{ij \in E} s_{ij}, x_1 - x_6, x_2 - x_6, x_3 - x_6, x_4 - x_6, x_5 - x_6 \right\rangle.$$

The linear forms  $x_i - x_6$  in  $P$  generate the irrelevant ideal for the ambient space  $\mathbb{P}^5$  of  $\mathcal{A}(G)$ . One further computes that  $\text{pd}(\Omega^1(\mathcal{A}(G))) = 2$ , so  $\mathcal{A}(G)$  is not tame either.  $\diamond$

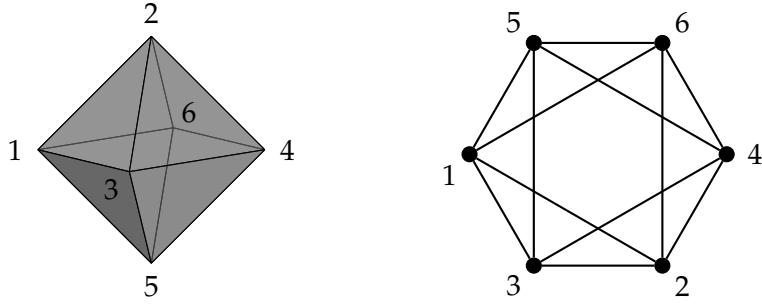


Figure 3.8: The octahedron and its edge graph.

Example 3.3.37 is uniquely minimal among non-gentle arrangements.

**Theorem 3.3.38.** *Consider the graphical arrangements  $\mathcal{A}(G)$  for all graphs  $G$  with  $n \leq 6$  vertices. With the exception of the octahedron graph, all of these arrangements are gentle.*

*Proof.* We prove this by exhaustive computation using our tools from Subsection 3.3.6.  $\square$

Except for the octahedron, all graphical arrangements on fewer than six vertices satisfy  $\text{pd}(\Omega^1(\mathcal{A}(G))) = 1$ . The octahedron gives rise to more non-gentle graphical arrangements.

**Corollary 3.3.39.** *Any graph that contains the octahedron as an induced subgraph is not gentle.*

This is a corollary of Proposition 3.3.40, which holds for all hyperplane arrangements  $\mathcal{A}$ , not just graphical ones. Any arrangement of a vertex-induced subgraph is a localisation at the intersection of the hyperplanes corresponding to the edges of the induced subgraph.

**Proposition 3.3.40.** *The localisation of a gentle hyperplane arrangement is gentle.*

*Proof.* Let  $\mathcal{A}$  be a gentle arrangement and  $X \in L(\mathcal{A})$ . Suppose that  $\mathcal{A}_X = \{f_1, \dots, f_k\}$  and  $\mathcal{A} \setminus \mathcal{A}_X = \{f_{k+1}, \dots, f_m\}$ . Since the  $f_i$  are linear, the following ideals are prime:

$$P = \langle f_1, \dots, f_k \rangle \subset R \quad \text{and} \quad \tilde{P} = P + \langle s_1, \dots, s_m \rangle \subset R[s_1, \dots, s_m].$$

Since  $I_0(\mathcal{A})$  is prime and  $I_0(\mathcal{A}) \subseteq \tilde{P}$ , the localisation  $I_0(\mathcal{A})_{\tilde{P}} \subset R[s]_{\tilde{P}}$  is prime. We claim

$$I_0(\mathcal{A})_{\tilde{P}} = \langle \theta(\ell_{\mathcal{A}}) : \theta \in \text{Der}(\mathcal{A})_P \rangle = \langle \theta(\ell_{\mathcal{A}}) : \theta \in \text{Der}(\mathcal{A}_X)_P \rangle. \quad (3.3.17)$$

The first equality is by Theorem 3.3.12 since localisation is exact. The second follows from  $\text{Der}(\mathcal{A})_P = \text{Der}(\mathcal{A}_X)_P$  which holds for localisations of arrangements [OT13, Ex. 4.123].

We now prove that  $s_i \in I_0(\mathcal{A})_{\tilde{P}}$  for all  $k+1 \leq i \leq m$ . To this end, fix  $s_i$ , its corresponding linear form  $f_i$  and hyperplane  $H_i = \{f_i = 0\}$  for  $k+1 \leq i \leq m$ . By Proposition 2.4.3, we have  $\text{Der}(\mathcal{A}) = R\theta_E \oplus \text{Der}_0(\mathcal{A})$ . As  $\text{Der}_0(\mathcal{A}) \subsetneq \text{Der}_0(\mathcal{A} \setminus f_i)$  we can choose  $\theta_{H_i} \in \text{Der}_0(\mathcal{A} \setminus f_i) \setminus \text{Der}_0(\mathcal{A})$ . Hence,  $\theta_{H_i}(f_i) = g$  for some nonzero  $g \in R$  and  $\theta_{H_i}(f_j) = 0$  for all  $j \neq i$ . The assumption  $f_i \notin \mathcal{A}_X$  yields  $\theta_{H_i} \in \text{Der}(\mathcal{A}_X)$ . Using (3.3.17) we obtain

$$\theta_{H_i}(\ell_{\mathcal{A}}) = s_i \frac{g}{f_i} \in I_0(\mathcal{A})_{\tilde{P}}.$$

As  $I_0(\mathcal{A})_{\tilde{P}}$  contains no polynomials that lie in  $R$ , we get  $g/f_i \notin I_0(\mathcal{A})_{\tilde{P}}$ . Thus,  $s_i \in I_0(\mathcal{A})_{\tilde{P}}$ . Then the quotient  $I_0(\mathcal{A})_{\tilde{P}}/\langle s_i : k+1 \leq i \leq m \rangle$  is also prime and by (3.3.17) equals

$$\langle \theta(\ell_{\mathcal{A}_X}) : \theta \in \text{Der}(\mathcal{A}_X)_P \rangle \subset R[s_1, \dots, s_k]_{P+\langle s_1, \dots, s_k \rangle}.$$

The preimage of this ideal in  $R[s_1, \dots, s_k]$  is the prime ideal  $I_0(\mathcal{A}_X)$ . Hence,  $\mathcal{A}_X$  is gentle.  $\square$

The proof above is independent of  $\mathcal{A}$  being linear. Hence, for any gentle arrangement of hypersurfaces  $\mathcal{A}$  and a prime ideal  $P \subset R$  the subarrangement  $\mathcal{A} \cap P$  is gentle.

Since induced subgraphs give rise to localisations, Proposition 3.3.40 is one ingredient in the following conjectural characterisation of graphic arrangements that are gentle.

**Conjecture 3.3.41.** *A graphic arrangement  $\mathcal{A}(G)$  is gentle if and only if the octahedron graph cannot be obtained from  $G$  by a series of edge contractions of an induced subgraph of  $G$ .*

This conjecture is supported by Theorem 3.3.38. Besides localisations, a proof would also require restrictions to a hyperplane which in the graphic case correspond to edge contraction. For general linear arrangements, restrictions do not preserve gentleness, though.

**Proposition 3.3.42.** *Restrictions of gentle hyperplane arrangements need not be gentle.*

*Proof.* Edelman and Reiner [ER93] constructed a free arrangement of 21 hyperplanes in  $\mathbb{P}^4$  with a restriction to 15 hyperplanes in  $\mathbb{P}^3$  which is not free. The linear forms in that nonfree arrangement  $\mathcal{A}$  are all subsums of  $x_1 + x_2 + x_3 + x_4$  which is the 4-dimensional resonance arrangement. This  $\mathcal{A}$  is not tame. The pre-likelihood ideal  $I_0(\mathcal{A})$  has five minimal generators. The ML degree is 51. Using the MACAULAY2 tools in Subsection 3.3.6, we find that the ideal quotient  $I_0(\mathcal{A}) : x_1$  strictly contains  $I_0(\mathcal{A})$ . Therefore,  $\mathcal{A}$  is not gentle.  $\square$

Restriction of  $\mathcal{A}(G)$  at a hyperplane models contraction of an edge in  $G$ . This preserves chordality. Thus, restrictions of free graphic arrangements are again free. Therefore, one might still hope that restrictions of a gentle graphic arrangement are gentle.

We proceed to the second main result in this subsection, a combinatorial construction of generators for the pre-likelihood ideal  $I_0(\mathcal{A}(G))$  of any graph  $G$ . Consider the derivations

$$\theta_k = x_1^k \partial_{x_1} + x_2^k \partial_{x_2} + \cdots + x_n^k \partial_{x_n} \quad \text{for } k = 0, 1, \dots, n-1.$$

It follows from Saito's criterion (Theorem 2.4.6) that  $\{\theta_0, \theta_1, \dots, \theta_{n-1}\}$  is a basis of the free module  $\text{Der}(\mathcal{A}(K_n))$ . Before removing edges from  $K_n$  to obtain an arbitrary graph, it is instructive to contemplate Theorem 3.3.12 for Saito's derivations  $\theta_k$ .

**Example 3.3.43.** The log-likelihood function for the braid arrangement  $\mathcal{A} = \mathcal{A}(K_n)$  equals

$$\ell_{\mathcal{A}} = \sum_{1 \leq i < j \leq n} s_{ij} \cdot \log(x_i - x_j). \quad (3.3.18)$$

By applying the derivation  $\theta_k$  to that function, we obtain a polynomial in  $\mathbb{C}[x, s]$ , namely

$$\theta_k(\ell_{\mathcal{A}}) = \sum_{1 \leq i < j \leq n} \left( \sum_{\ell=0}^{k-1} x_i^\ell x_j^{k-1-\ell} \right) \cdot s_{ij}. \quad (3.3.19)$$

We know from Theorem 3.3.12 that these polynomials generate  $I_0(\mathcal{A})$ , and hence also the likelihood ideal  $I(\mathcal{A})$  as  $\mathcal{A}$  is tame and thus gentle. For  $n = 4$  see Examples 3.3.2.  $\diamond$

Now let  $G = (V, E)$  be an arbitrary graph and let  $\mathcal{A} = \mathcal{A}(G)$  be its graphic arrangement. The log-likelihood function  $\ell_{\mathcal{A}}$  is the sum in (3.3.18) but now restricted to pairs  $\{i, j\}$  in  $E$ . The corresponding restricted sum in (3.3.19) still lies in the ideal  $I_0(\mathcal{A})$ .

A subset  $T$  of  $[n]$  is a *separator* of  $G$  if the induced subgraph on  $[n] \setminus T$  is disconnected. We denote this subgraph by  $G \setminus T$ , and we consider any connected component  $C$  of  $G \setminus T$ . Following [Müh], we define the *separator-based derivation* associated to the data above:

$$\theta_C^T = \sum_{i \in C} \prod_{t \in T} (x_i - x_t) \cdot \partial_{x_i}.$$

The following is implied by the main result in [Müh] along with Theorem 3.3.12.

**Theorem 3.3.44.** *The module  $\text{Der}(\mathcal{A}(G))$  is generated by  $\theta_0, \dots, \theta_{n-1}$  and a set of separator-based derivations. Hence,  $I_0(\mathcal{A})$  is generated by the images of  $\ell_{\mathcal{A}}$  under the derivations  $\theta_k$  and  $\theta_C^T$ .*

The generators in this theorem are redundant. We do not need  $\theta_k$  if  $k$  exceeds the connectivity of  $G$ , and not all separator-based derivations  $\theta_C^T$  are necessary to generate  $\text{Der}(\mathcal{A}(G))$  and  $I_0(\mathcal{A})$ . It remains an interesting problem to extract minimal generators.

**Example 3.3.45** (Octahedron revisited). Let  $G$  be the graph in Example 3.3.37. In this case it suffices to consider only (inclusionwise) minimal separators  $T$ ; these are  $\{2, 3, 5, 6\}$ ,  $\{1, 3, 4, 6\}$  and  $\{1, 2, 4, 5\}$ . The connectivity of the graph is four. The module  $\text{Der}(\mathcal{A}(G))$  is minimally generated by the following eight derivations:

$$\theta_0, \theta_1, \theta_2, \theta_3, \theta_4, \theta_{\{1\}}^{\{2,3,5,6\}}, \theta_{\{2\}}^{\{1,3,4,6\}}, \theta_{\{3\}}^{\{1,2,4,5\}}.$$

Setting  $z_{ij} := x_i - x_j$ , we infer the following set of minimal generators for the ideal  $I_0(\mathcal{A})$ :

$$\begin{aligned} \theta_k(\ell_{\mathcal{A}}) &= \sum_{(i,j) \in E} \left( \sum_{\ell=0}^{k-1} x_i^\ell x_j^{k-1-\ell} \right) \cdot s_{ij} \quad \text{for } k = 1, \dots, 4, \\ \theta_{\{1\}}^{\{2,3,5,6\}}(\ell_{\mathcal{A}}) &= z_{13}z_{15}z_{16} \cdot s_{12} + z_{12}z_{15}z_{16} \cdot s_{13} + z_{12}z_{13}z_{16} \cdot s_{15} + z_{12}z_{13}z_{15} \cdot s_{16}, \\ \theta_{\{2\}}^{\{1,3,4,6\}}(\ell_{\mathcal{A}}) &= z_{23}z_{24}z_{26} \cdot s_{12} + z_{21}z_{24}z_{26} \cdot s_{23} + z_{21}z_{23}z_{26} \cdot s_{24} + z_{21}z_{23}z_{24} \cdot s_{26}, \\ \theta_{\{3\}}^{\{1,2,4,5\}}(\ell_{\mathcal{A}}) &= z_{32}z_{34}z_{35} \cdot s_{13} + z_{31}z_{34}z_{35} \cdot s_{23} + z_{31}z_{32}z_{35} \cdot s_{34} + z_{31}z_{32}z_{34} \cdot s_{35}. \end{aligned}$$

These seven generators are linear in  $s$ , and they have the  $x$ -degrees stated in Example 3.3.37. Since  $\theta_0(\ell_{\mathcal{A}}) = 0$ , this derivation does not yield a generator of  $I_0(\mathcal{A})$ .  $\diamond$

### 3.3.6. Software and computations

We have implemented functions in MACAULAY2 to compute the pre-lielihood ideal  $I_0(\mathcal{A})$  and the likelihood ideal  $I(\mathcal{A})$  for any arrangement  $\mathcal{A}$ . The input consists of  $m$  homogeneous polynomials  $f_1, \dots, f_m$  in  $n$  variables  $x_1, \dots, x_n$ . Along the way, our code creates the four modules seen in Subsection 3.3.1, and it also computes the relevant multidegrees.

Our code is made available, along with various examples, in the MathRepo code repository hosted at MPI-MiS via [KKM<sup>+</sup>24b]. In this subsection we offer a guide on how to use the software, and present three short case studies.

We start with the function `preLikelihoodIdeal`. Its input is a list  $F$  of  $m$  homogeneous elements in a polynomial ring  $R$ . The list  $F$  defines an arrangement  $\mathcal{A}$  in  $\mathbb{P}^{n-1}$ . Our code augments the given ring  $R$  with additional variables  $s_1, s_2, \dots, s_m$ , one for each element in the list  $F$ , and it outputs generators for the pre-lielihood ideal  $I_0(\mathcal{A})$ . We can then analyse that output and test whether it is prime, in which case  $I_0(\mathcal{A}) = I(\mathcal{A})$ . Our code also has a function `likelihoodIdeal` which computes  $I(\mathcal{A})$  directly even if  $\mathcal{A}$  is not gentle.

**Example 3.3.46.** Revisiting Example 3.3.17, we consider an arrangement  $\mathcal{A}$  of four conics and one line in  $\mathbb{P}^2$ . We compute its pre-lielihood ideal  $I_0(\mathcal{A})$  as follows:

```
R = QQ[x,y,z];
F = {x^2+y^2+z^2, x^2+2*y*z-z^2, y^2+2*z*x-x^2, z^2+2*x*y-y^2, x+y+z};
I = preLikelihoodIdeal(F)
```

The ideal  $I_0(\mathcal{A})$  has seven minimal generators, starting with  $2(s_1 + \dots + s_4) + s_5$ . Our choice of  $\mathcal{A}$  exhibits the generic behaviour in Example 3.3.17. In particular, the ML degree is 25. Running `codim I, multidegree I, betti mingens I` computes the codimension three, the multidegree  $25p^2u + 6pu^2 + u^3$  and the total degrees of minimal generators. A following `isPrime I` returns `true`, so the arrangement  $\mathcal{A}$  is indeed gentle.  $\diamond$

We now turn to our case studies. The first is a non-gentle arrangement of planes in  $\mathbb{P}^3$ .

**Example 3.3.47.** The following arrangement is due to Cohen et al. [CDFV11, Ex. 5.3]:

```
R = QQ[x1,x2,x3,x4];
F = {x1,x2,x3,x1+x4,x2+x4,x3+x4,x1+x2+x4,x1+x3+x4,x2+x3+x4}
ass preLikelihoodIdeal F
I = likelihoodIdeal F;
codim I, multidegree I, betti mingens I, isPrime I
```

We obtain  $I(\mathcal{A})$  from  $I_0(\mathcal{A})$  by removing the associated prime  $\langle s_1 + \dots + s_9, x_1, x_2, x_3, x_4 \rangle$ . The likelihood ideal  $I(\mathcal{A})$  has six minimal generators, and multidegree

$$[\mathcal{L}_{\mathcal{A}}] = 5p^3u + 9p^2u^2 + 5pu^3 + u^4. \quad \diamond$$

**Example 3.3.48** (No 3-way interaction). A commonly studied model for three binary random variables (e.g. [Sul18, Ex. 7.3.12]) is given by

$$p_{ijk} = a_{ij}b_{ik}c_{jk} \quad \text{for } i, j, k \in \{0, 1\}.$$

This parametrises the toric hypersurface  $\{p_{000}p_{110}p_{101}p_{011} = p_{100}p_{010}p_{001}p_{111}\} \subset \mathbb{P}^7$ . This toric model fits into our setup by setting  $m = 9$ , and considering the  $n = 12$  parameters

$$x = (a_{00}, a_{10}, a_{01}, a_{11}, b_{00}, \dots, b_{11}, c_{00}, \dots, c_{11}).$$

We take  $\mathcal{A}$  to be the twelve coordinate hyperplanes  $a_{00}, a_{10}, \dots, c_{11}$  together with

$$\begin{aligned} f(x) = & a_{00}b_{00}c_{00} + a_{00}b_{01}c_{01} + a_{01}b_{00}c_{10} + a_{01}b_{01}c_{11} \\ & + a_{10}b_{10}c_{00} + a_{10}b_{11}c_{01} + a_{11}b_{10}c_{10} + a_{11}b_{11}c_{11}. \end{aligned}$$

The pre-likelihood ideal  $I_0(\mathcal{A})$  has 25 minimal primes, so the arrangement is far from gentle. The likelihood ideal  $I(\mathcal{A})$  can be computed by performing the saturation  $I_0(\mathcal{A}) : a_{00}f^2$  and checking that this ideal is prime. We found this to be the fastest method.

An alternative parametrisation of the model with only seven parameters  $x_i$  is given by

$$g(x) = x_1^6 + x_1^5x_2 + x_1^5x_3 + x_1^5x_4 + x_1^3x_2x_3x_5 + x_1^3x_3x_4x_6 + x_1^3x_2x_4x_7 + x_2x_3x_4x_5x_6x_7.$$

The arrangement  $\mathcal{A}' = \{x_1, \dots, x_7, g(x)\}$  is not gentle either. The ideal  $I_0(\mathcal{A}')$  has 19 generators. The likelihood ideal is  $I_0(\mathcal{A}') : x_1x_2x_3x_4x_5$ . It has 48 generators in various degrees, some of which are quartic in the  $s$ -variables. The multidegree

$$[\mathcal{L}_{\mathcal{A}'}] = 3p^6u + 13p^5u^2 + 25p^4u^3 + 30p^3u^4 + 18p^2u^5 + 6pu^6 + u^7$$

reveals the correct ML degree of three, known from [ABB<sup>+</sup>19, Ex. 32].  $\diamond$

**Example 3.3.49** (CEGM model). Consider the moduli space of six labelled points in linearly general position in  $\mathbb{P}^2$ . This very affine variety arises in the CEGM model in particle physics [CEGM19]. We write this as the projective arrangement  $\mathcal{A}$  with  $m = 15$  and  $n = 5$  given by the  $3 \times 3$  minors of the  $3 \times 6$  matrix

$$\begin{bmatrix} 1 & 0 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 & x_1 & x_2 \\ 0 & 0 & 1 & 1 & x_3 & x_4 \end{bmatrix}.$$

Using  $x_5$  for the homogenisation, we compute the pre-likelihood ideal  $I_0(\mathcal{A})$  as follows:

```
R = QQ[x1,x2,x3,x4,x5];
F = {x1,x2,x3,x4,x5,x1-x2,x1-x3,x1-x5,x2-x5,x2-x4,x3-x4,x3-x5,x4-x5,
 x1*x4-x2*x3,x1*x4-x2*x3-x1+x2+x3-x4};
I0 = preLikelihoodIdeal F;
```

The ideal  $I_0$  of this arrangement is simple to define, having only six generators of degrees  $\binom{2}{1}$  (twice) and  $\binom{3}{1}$  (four times). However, due to their size, computing one Gröbner basis of this ideal is already challenging. Using numerical irreducible decomposition from [BT18] we obtain that  $I_0$  has 25 associated primes, so  $\mathcal{A}$  is certainly not gentle.  $\diamond$

## 3.4. Conclusion

In this chapter we have studied Problem 3.0.1 concerning the maximum likelihood degree of scaled toric varieties from three different angles. First, in Section 3.1, we considered the case of toric varieties with ML degree one. The toric fibre product construction gives rise to new models with ML degree one. Our first main result, Theorem 3.1.8, gives an explicit description of the Horn parametrisation of a toric fibre product. Moreover, we connect the toric fibre product to geometric modelling. The second main result, Theorem 3.1.14, yields a construction for a set of blending functions on a toric fibre product patch satisfying the property of rational linear precision. A classification of toric varieties with ML deg one is currently only known in the two-dimensional case; it is an interesting problem for future work to extend this result to higher dimensional cases. The toric fibre product construction is likely to play a crucial role in this.

In Section 3.2 we introduced the notion of Euler stratifications, and connect it to Problem 3.0.1. Besides structural results on Euler stratifications, we developed Algorithms 3 and 6 to compute such stratifications for projective and very affine hypersurface families, respectively. This gives a complete computational answer to Problem 3.0.1. However, the complexity of the situation quickly explodes, and our algorithms become computationally infeasible. An interesting approach for future work is to compute tropicalisations of Euler strata. We expect to find more efficient algorithms for the computation of these combinatorial shadows. This also connects to work on tropical Severi varieties, see e.g. [DHT17].

Finally, Section 3.3 takes a parametric view on the ML degree. We connect this to the theory of logarithmic derivations of arrangements (confer Theorem 3.3.12). The main result, Theorem 3.3.1, provides a novel way to compute the ideal of the likelihood correspondence. This is particularly easy if the arrangement is gentle. A main direction for future studies is to prove that generic arrangement are gentle (Conjecture 3.3.41).

A central notion of this chapter was the ML degree, a critical point count measuring the (algebraic) complexity of maximum likelihood estimation for a specific model. In the next chapter on polynomial neural networks, we introduce a similar critical point count, the *learning degree* of a polynomial neural network. It measures the complexity of the optimisation landscape and informs the training process of the network.

## Chapter 4

# Polynomial neural networks

Feedforward neural networks are a ubiquitous tool in machine learning, see e.g. [GBC16, Ch. 6], [BN06, Ch. 5]. The nonlinearity in their activation functions allows modelling complex phenomena effectively. Despite their empirical triumph, the underlying theoretical behaviours of neural networks remain an open and active field of research. Polynomial neural networks constitute a class of feedforward neural networks where the activation function is given by exponentiation with an integer. Hence, the networks' outputs are polynomials in the input data. The set of all functions potentially learnable by the network is then a semialgebraic set and amenable to the techniques of algebraic geometry. In this chapter we focus on two main aspects: the expressivity (Section 4.2) and the optimisation process (Section 4.3) of a polynomial neural network. The former deals with the question of which functions can be represented by the network in principle; the latter is concerned with how many and what functions can be learned in a training process, e.g. using gradient descent, after different initialisations.

## 4.1. Introduction

Over the past decade, neural networks have achieved remarkable success, primarily driven by advancements in deep learning. With the increased computational power, availability of large datasets, and algorithmic innovations, deep learning has led to groundbreaking achievements in areas like image and speech recognition, natural language processing, and autonomous systems. In addition, deep neural networks outperform many traditional statistical models, significantly impacting fields such as healthcare and finance.

Various activation functions are studied to understand their role in introducing non-linearity, affecting gradient propagation, and influencing computational efficiency. Polynomial activation functions have gained interest for their ability to introduce higher-order interactions between inputs, allowing networks to model complex, nonlinear phenomena more effectively [OPP03]. Although feedforward neural networks with polynomial activation functions are well-known to be non-universal approximators [HSW89], there are many practical tasks where architectures with polynomial activations have outperformed other ones, especially in environments where data relationships are polynomial in nature. In particular, polynomial neural networks have led to state-of-the-art results in engineering tasks like face detection from cluttered images [HSHK03], image generation [CMB<sup>+</sup>20], 3D shape recognition [YHN<sup>+</sup>21], as well as financial applications like forecasting trading signals [GHL11], uncertain natural frequency quantification [DNM<sup>+</sup>16], and estimating stock closing indices [NM18]. However, the choice of degree and the potential for overfitting are critical considerations in applying polynomial neural networks to practical tasks.

Compared to other low-degree activation functions, polynomial functions capture intricate patterns within data without the need for additional layers, potentially reducing model complexity and computational costs. In addition, common neural network activation functions, including sigmoid and ReLU, can be effectively approximated using ratios of polynomials. Recent work also shows that fully-connected feedforward neural networks using ratios of polynomials as activation functions approximate smooth functions more efficiently than ReLU networks [BNT20]. Our exploration in the realm of polynomial networks lays the groundwork for further investigation of rational activation functions.

In this chapter, we perform an algebro-geometric study of neuromanifolds and their Zariski closures, neurovarieties, following the previous work by [KTB19]. In Section 4.1, we introduce polynomial neural networks, as well as their neuromanifolds and neurovarieties. Neuromanifolds provide a natural generalisation of the set of symmetric tensors of bounded symmetric rank. This perspective is taken in Subsection 4.2.1 where we characterise neuromanifolds for some shallow polynomial neural networks. In Subsection 4.2.2, we describe different approaches for studying neurovarieties. These can be seen as good approximations to neuromanifolds offering a more accessible framework for their algebro-geometric study. The dimension of the neurovariety provides a measure for the expressivity of the network. These dimensions are studied in Subsection 4.2.3. For a shallow polynomial neural network with a single output unit, this corresponds to the Alexander-Hirschowitz Theorem. In the deep case, we present conjectures supported by empirical evidence. Finally, in Section 4.3, we study the optimisation process of a polynomial neural network. We describe the complexity of the optimisation landscape by the *learning degree* of the neurovariety. It provides an upper bound on the number of functions a network can learn after a training process with different initialisations. We relate this number to the Euclidean distance degree and compute it for a family of architectures. We provide code for computations and experiments at the MathRepo code repository at [KLW24b].

Recall the definition and notation of feedforward neural networks from Section 2.5. In this chapter, we focus on *monomial* activations  $x \mapsto x^r$ . Then the network  $F_\theta$  is a polynomial map and the study of these networks is amenable to techniques from algebraic geometry.

The affine-linear map  $f_l$  can be written as  $f_l(\mathbf{x}) = W_l \mathbf{x} + \mathbf{b}_l$ , where  $W_l \in \mathbb{R}^{d_l \times d_{l-1}}$  is a linear map and  $\mathbf{b}_l \in \mathbb{R}^{d_l}$ . We are considering networks without biases to make the polynomial map  $F_\theta$  homogeneous. Let us summarise the setup in the following.

**Definition 4.1.1.** A *polynomial neural network (PNN)*  $p_{\mathbf{w}}$  with architecture

$$(\mathbf{d} = (d_0, d_1, \dots, d_L), r)$$

is a fully connected feedforward neural network

$$p_{\mathbf{w}} = W_L \circ \sigma_{L-1} \circ W_{L-1} \circ \sigma_{L-2} \circ \dots \circ \sigma_1 \circ W_1 : \mathbb{R}^{d_0} \rightarrow \mathbb{R}^{d_L},$$

where  $W_i \in \mathbb{R}^{d_i \times d_{i-1}}$  are linear maps and the activation functions

$$\sigma_i(\mathbf{x}) = \rho_r(\mathbf{x}) := (x_1^r, \dots, x_n^r)$$

are monomial. The number  $r$  is called the *activation degree* of  $p_{\mathbf{w}}$ . The parameters  $\mathbf{w}$  are given by the entries of the matrices  $W_i$ , i.e.  $\mathbf{w} = (W_1, W_2, \dots, W_L)$ .

A PNN  $p_{\mathbf{w}}$  with architecture  $(\mathbf{d}, r)$  is a homogeneous polynomial map of degree  $r^{L-1}$ . Hence, the associated parameter map  $\Psi_{\mathbf{d}, r}$  maps an  $L$ -tuple of matrices  $(W_1, W_2, \dots, W_L)$  to a  $d_L$ -tuple of homogeneous polynomials of degree  $r^{L-1}$  in  $d_0$  variables, i.e.

$$\Psi_{\mathbf{d}, r} : \mathbb{R}^{d_1 \times d_0} \times \dots \times \mathbb{R}^{d_L \times d_{L-1}} \rightarrow (\text{Sym}_{r^{L-1}}(\mathbb{R}^{d_0}))^{d_L}, \quad \mathbf{w} \mapsto p_{\mathbf{w}} = \left( p_{\mathbf{w}}^{(1)}, \dots, p_{\mathbf{w}}^{(d_L)} \right)^T.$$

We can identify elements in the image with their vectors of coefficients in  $\mathbb{R}^n$  with  $n = d_L \binom{r^{L-1} + d_0 - 1}{d_0 - 1}$ . Coordinates of this space will be denoted by  $c_I^{(j)}$  so that  $c_I^{(j)}$  is the coefficient of the monomial  $\mathbf{x}^I$  in the polynomial  $p_{\mathbf{w}}^{(j)}(\mathbf{x})$ ; here,  $I$  is a multiindex in

$$\binom{[d_0]}{r^{L-1}} = \left\{ r^{L-1} \text{ element subsets of } \{1, 2, \dots, d_0\} \right\}.$$

**Definition 4.1.2.** The image of  $\Psi_{\mathbf{d}, r}$  is the *neuromanifold*  $\mathcal{M}_{\mathbf{d}, r}$ . Its Zariski closure is the *neurovariety*  $\mathcal{V}_{\mathbf{d}, r}$ , an affine variety inside  $(\text{Sym}_{r^{L-1}}(\mathbb{R}^{d_0}))^{d_L}$ .

The neuromanifold  $\Psi_{\mathbf{d}, r}$  is a semialgebraic set inside  $(\text{Sym}_{r^{L-1}}(\mathbb{R}^{d_0}))^{d_L}$ , which means that it is a finite union of sets that can be defined by polynomial equations and inequalities. The observation that the neuromanifold  $\Psi_{\mathbf{d}, r}$  is a semialgebraic set follows from the Tarski–Seidenberg Theorem (see e.g. [BCR13, Thm. 1.4.2]) which states that the image of a semialgebraic set under a polynomial map is again semialgebraic. The neurovariety can be seen as an approximation to the neuromanifold which is much easier to describe. In general, Zariski closures can be “far off” from the original space (viewed in the Euclidean topology). For example, the dimension of the Zariski closure can increase. This is, however, not the case for semialgebraic sets: the dimension of the neuromanifold and the dimension of its neurovariety agree. Since the dimension of the neuromanifold is the primary measure for the network expressivity in Section 4.2, it is legitimate to study the neurovariety instead. Moreover, the degree theory developed in Subsection 4.3.1 only works for Zariski closed sets. Again, the neurovariety can be seen as a good approximation to the neuromanifold in terms of the complexity of the optimisation process. We reiterate the warning that neuromanifolds are typically not smooth manifolds.

Increasing the widths of hidden layers gives a containment of neuromanifolds.

**Proposition 4.1.3.** Let  $\mathbf{d} = (d_0, \dots, d_i, \dots, d_L)$  and let  $\mathbf{d}' = (d_0, \dots, d'_i, \dots, d_L)$  be a tuple which differs from  $\mathbf{d}$  precisely in the  $i^{\text{th}}$  entry for  $0 < i < L$  and assume  $d'_i \geq d_i$ . Then there is a containment  $\mathcal{M}_{\mathbf{d},r} \subseteq \mathcal{M}_{\mathbf{d}',r}$ , and consequently also  $\mathcal{V}_{\mathbf{d},r} \subseteq \mathcal{V}_{\mathbf{d}',r}$ .

*Proof.* Let  $\mathbf{w} = (W_1, \dots, W_L)$  be a parameter vector for the architecture  $\mathbf{d}$  and  $p_{\mathbf{w}} \in \mathcal{M}_{\mathbf{d},r}$  the corresponding polynomial network. Let  $\mathbf{w}' = (W'_1, \dots, W'_L)$  be the parameter vector for the architecture  $\mathbf{d}'$  such that each  $W'_i$  has  $W_i$  as left-top submatrix and zeros elsewhere. Then  $p_{\mathbf{w}} = p_{\mathbf{w}'} \in \mathcal{M}_{\mathbf{d}',r}$ .  $\square$

## 4.2. Expressivity

As the neuromanifold  $\mathcal{M}_{\mathbf{d},r}$  is semialgebraic, it has the same dimension as its Zariski closure,  $\mathcal{V}_{\mathbf{d},r}$ . In [KTB19], the dimension of  $\mathcal{V}_{\mathbf{d},r}$  was proposed as a measure for the *expressivity* of the network architecture  $(\mathbf{d}, r)$ .

**Definition 4.2.1.** An architecture  $(\mathbf{d}, r)$  is *filling* if  $\mathcal{V}_{\mathbf{d},r} = (\text{Sym}_{r,L-1}(\mathbb{R}^{d_0}))^{d_L}$ . In this case, we say that  $\mathcal{M}_{\mathbf{d},r}$  is *thick*, i.e. it has positive Lebesgue measure.

The case of filling architectures is particularly interesting from a machine learning perspective as filling networks have the most expressive power: if  $\mathcal{M}_{\mathbf{d},r} = (\text{Sym}_{r,L-1}(\mathbb{R}^{d_0}))^{d_L}$ , any target function in  $(\text{Sym}_{r,L-1}(\mathbb{R}^{d_0}))^{d_L}$  can be represented *exactly* by the network. In the case of non-filling architectures, a general target function can only be approximated by the network. For more details on the learning process, see Section 2.5. Moreover, from an optimisation perspective it is advantageous to work with thick neuromanifolds as non-thick neuromanifolds are non-convex, see [KTB19, Prop. 7].

**Example 4.2.2.** For architecture  $\mathbf{d} = (2, 1, 1)$ ,  $r = 2$  with input  $\mathbf{x} = (x_1, x_2)^T$  and weights

$$W_1 = (w_{111} \quad w_{112}), \quad W_2 = (w_{211}),$$

the network is

$$p_{\mathbf{w}}(\mathbf{x}) = W_2 \rho_2 W_1 \mathbf{x} = (w_{211}(w_{111}x_1 + w_{112}x_2)^2).$$

There are three parameters in  $W_1$  and  $W_2$  which  $\Psi_{(2,1,1),2}$  maps to the quadric above with  $\dim(\text{Sym}_2(\mathbb{R}^2)) = 3$  coefficients. Let  $c_{11}, c_{12}$  and  $c_{22}$  be coordinates for this space, representing the coefficients of  $x_1^2, x_1x_2$  and  $x_2^2$ , respectively. The neuromanifold is a hyper-surface defined by the quadratic equation  $c_{12}^2 - c_{11}c_{22} = 0$ , which is shown in Figure 4.1. The neurovariety is equal to the neuromanifold, and hence it is not filling. See Lemma 4.2.7 for more details and generalisations.  $\diamond$

**Example 4.2.3** ([KTB19], Example 3). For architecture  $\mathbf{d} = (2, 2, 3)$ ,  $r = 2$  with input  $\mathbf{x} = (x_1, x_2)^T$  and parameters

$$W_1 = \begin{pmatrix} w_{111} & w_{112} \\ w_{121} & w_{122} \end{pmatrix}, \quad W_2 = \begin{pmatrix} w_{211} & w_{212} \\ w_{221} & w_{222} \\ w_{231} & w_{232} \end{pmatrix},$$

the network is

$$p_{\mathbf{w}}(\mathbf{x}) = W_2 \rho_2 W_1 \mathbf{x} = \begin{pmatrix} w_{211}(w_{111}x_1 + w_{112}x_2)^2 + w_{212}(w_{121}x_1 + w_{122}x_2)^2 \\ w_{221}(w_{111}x_1 + w_{112}x_2)^2 + w_{222}(w_{121}x_1 + w_{122}x_2)^2 \\ w_{231}(w_{111}x_1 + w_{112}x_2)^2 + w_{232}(w_{121}x_1 + w_{122}x_2)^2 \end{pmatrix}.$$

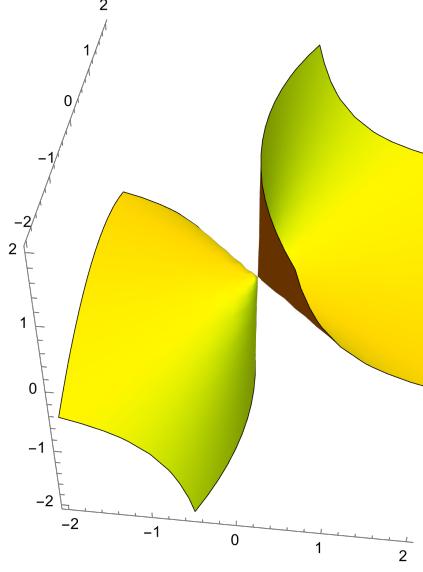


Figure 4.1: The neuromanifold  $\mathcal{M}_{(2,1,1),2}$  in  $\text{Sym}_2(\mathbb{R}^2) \cong \mathbb{R}^3$ .

There are ten parameters in  $W_1$  and  $W_2$  which  $\Psi_{(2,2,3),2}$  maps to the triple of quadrics above with  $\dim(\text{Sym}_2(\mathbb{R}^2)^3) = 9$  coefficients. The neurovariety is an eight-dimensional hypersurface defined by the single cubic equation

$$\det \begin{pmatrix} c_{11}^{(1)} & c_{12}^{(1)} & c_{22}^{(1)} \\ c_{11}^{(2)} & c_{12}^{(2)} & c_{22}^{(2)} \\ c_{11}^{(3)} & c_{12}^{(3)} & c_{22}^{(3)} \end{pmatrix} = 0.$$

This implies that the architecture  $\mathbf{d} = (2, 2, 3), r = 2$  is not filling. For more details, see Proposition 4.2.18 below. Note that in this case  $\mathcal{M}_{\mathbf{d},r} \subsetneq \mathcal{V}_{\mathbf{d},r}$ , see Proposition 4.2.19.  $\diamond$

There is a naïve expectation for the dimension of the neurovariety, namely the number of parameters. However, one immediately observes a symmetry in the parameters for any network architecture, called *multi-homogeneity*.

**Lemma 4.2.4** ([KTB19, Lem. 13]). *For all diagonal matrices  $D_i \in \mathbb{R}^{d_i \times d_i}$  and permutation matrices  $P_i \in \mathbb{Z}^{d_i \times d_i}$ , where  $i = 1, \dots, L-1$ , the parameter map  $\Psi_{\mathbf{d},r}$  returns the same network under the transformation*

$$\begin{aligned} W_1 &\leftarrow P_1 D_1 W_1 \\ W_2 &\leftarrow P_2 D_2 W_2 D_1^{-r} P_1^T \\ W_3 &\leftarrow P_3 D_3 W_3 D_2^{-r} P_2^T \\ &\vdots \\ W_L &\leftarrow W_L D_{L-1}^{-r} P_{L-1}^T. \end{aligned}$$

Hence, a generic fibre of  $\Psi_{\mathbf{d},r}$  has dimension at least  $\sum_{i=1}^{L-1} d_i$ . We call the number of parameters subtracted by this dimension the expected dimension of the neurovariety.

**Definition 4.2.5.** We define the *expected dimension* of the neurovariety  $\mathcal{V}_{\mathbf{d},r}$  to be

$$\text{edim}(\mathcal{V}_{\mathbf{d},r}) := \min \left\{ d_L + \sum_{i=0}^{L-1} (d_i d_{i+1} - d_{i+1}), d_L \binom{d_0 + r^{L-1} - 1}{r^{L-1}} \right\}.$$

| $\mathbf{d}$ | dim | edim | amb dim | ideal                             | $\mathcal{M}_{\mathbf{d},2} = \mathcal{V}_{\mathbf{d},2}?$ |
|--------------|-----|------|---------|-----------------------------------|------------------------------------------------------------|
| (1,1,1)      | 1   | 1    | 1       | $\langle 0 \rangle$               | yes                                                        |
| (1,1,2)      | 2   | 2    | 2       | $\langle 0 \rangle$               | yes                                                        |
| (1,1,3)      | 3   | 3    | 3       | $\langle 0 \rangle$               | yes                                                        |
| (1,2,1)      | 1   | 1    | 1       | $\langle 0 \rangle$               | yes                                                        |
| (1,2,2)      | 2   | 2    | 2       | $\langle 0 \rangle$               | yes                                                        |
| (1,2,3)      | 3   | 3    | 3       | $\langle 0 \rangle$               | yes                                                        |
| (1,3,1)      | 1   | 1    | 1       | $\langle 0 \rangle$               | yes                                                        |
| (1,3,2)      | 2   | 2    | 2       | $\langle 0 \rangle$               | yes                                                        |
| (1,3,3)      | 3   | 3    | 3       | $\langle 0 \rangle$               | yes                                                        |
| (2,1,1)      | 2   | 2    | 3       | determinantal, Lemma 4.2.8        | yes, Lemma 4.2.8                                           |
| (2,1,2)      | 3   | 3    | 6       | determinantal, Lemma 4.2.8        | yes, Lemma 4.2.8                                           |
| (2,1,3)      | 4   | 4    | 9       | determinantal, Lemma 4.2.8        | yes, Lemma 4.2.8                                           |
| (2,2,1)      | 3   | 3    | 3       | $\langle 0 \rangle$               | yes, Lemma 4.2.7                                           |
| (2,2,2)      | 6   | 6    | 6       | $\langle 0 \rangle$               | no, Proposition 4.2.19                                     |
| (2,2,3)      | 8   | 8    | 9       | determinantal, Proposition 4.2.18 | no, Proposition 4.2.19                                     |
| (2,3,1)      | 3   | 3    | 3       | $\langle 0 \rangle$               | yes, Lemma 4.2.7                                           |
| (2,3,2)      | 6   | 6    | 6       | $\langle 0 \rangle$               | yes, Proposition 4.2.12                                    |
| (2,3,3)      | 9   | 9    | 9       | $\langle 0 \rangle$               | yes, Proposition 4.2.12                                    |
| (3,1,1)      | 3   | 3    | 6       | determinantal, Lemma 4.2.8        | yes, Lemma 4.2.8                                           |
| (3,1,2)      | 4   | 4    | 12      | determinantal, Lemma 4.2.8        | yes, Lemma 4.2.8                                           |
| (3,1,3)      | 5   | 5    | 18      | determinantal, Lemma 4.2.8        | yes, Lemma 4.2.8                                           |
| (3,2,1)      | 5   | 6    | 6       | determinantal, Lemma 4.2.7        | yes, Lemma 4.2.7                                           |
| (3,2,2)      | 8   | 8    | 12      | determinantal, Example 4.2.6      | no, Remark 4.2.20                                          |
| (3,2,3)      | 10  | 10   | 18      | Example 4.2.6                     | no, Remark 4.2.20                                          |
| (3,3,1)      | 6   | 6    | 6       | $\langle 0 \rangle$               | yes, Lemma 4.2.8                                           |
| (3,3,2)      | 12  | 12   | 12      | $\langle 0 \rangle$               | no, Lemma 4.2.9                                            |
| (3,3,3)      | 15  | 15   | 18      | Example 4.2.21                    | no, Lemma 4.2.9                                            |

Table 4.1: Properties of neurovarieties for shallow polynomial neural networks with widths  $d_i \in [3]$  and activation degree  $r = 2$ .

By Lemma 4.2.4,  $\dim(\mathcal{V}_{\mathbf{d},r}) \leq \text{edim}(\mathcal{V}_{\mathbf{d},r})$ . The difference  $\text{edim}(\mathcal{V}_{\mathbf{d},r}) - \dim(\mathcal{V}_{\mathbf{d},r})$  is the *defect* of  $\mathcal{V}_{\mathbf{d},r}$ . If the defect is nonzero  $\mathcal{V}_{\mathbf{d},r}$  is called *defective*. We refer to

$$\dim((\text{Sym}_{r^{L-1}}(\mathbb{R}^{d_0}))^{d_L}) = d_L \binom{d_0 + r^{L-1} - 1}{r^{L-1}}$$

as the *ambient dimension* of  $\mathcal{V}_{\mathbf{d},r}$ .

In Table 4.1 we compute the ideal of the neurovariety and its dimension for shallow polynomial neural networks with widths  $d_i \in [3]$  and activation degree  $r = 2$ . We also compare the neurovariety with the neuromanifold.

Most ideals of neurovarieties in Table 4.1 are described by results in the upcoming subsections. The ideals for all but two cases ( $\mathbf{d} = (3,2,3)$  and  $\mathbf{d} = (3,3,3)$ ) in the table are determinantal. In general, we expect most neurovarieties to be non-determinantal. Among the table, only the neurovariety for the architecture  $\mathbf{d} = (3,2,1)$  does not have expected dimension. The following example describes the cases that are not explained by any results in the rest of the chapter.

| $\mathbf{d}$      | $r$                | assumptions                   | result             |
|-------------------|--------------------|-------------------------------|--------------------|
| $(d_0, d_1, 1)$   | 2                  |                               | Lemma 4.2.7        |
| $(d_0, 1, d_2)$   | 2                  |                               | Lemma 4.2.8        |
| $(d_0, d_0, d_2)$ | 2                  | $d_0, d_2 \geq 2$             | Lemma 4.2.9        |
| $(d_0, 1, d_2)$   | $r \in \mathbb{N}$ |                               | Lemma 4.2.11       |
| $(d_0, d_1, d_2)$ | $r \in \mathbb{N}$ | $d_1 \geq \binom{r+d_0-1}{r}$ | Proposition 4.2.12 |
| $(2, d_1, 1)$     | 3                  | $d_1 \geq 2$                  | Corollary 4.2.16   |
| $(2, d_1, 1)$     | 4                  | $d_1 \geq 3$                  | Corollary 4.2.16   |
| $(2, d_1, 1)$     | 5                  | $d_1 \geq 3$                  | Corollary 4.2.16   |
| $(3, d_1, 1)$     | 4                  | $d_1 \geq 6$                  | Corollary 4.2.17   |
| $(3, d_1, 1)$     | 5                  | $d_1 \geq 7$                  | Corollary 4.2.17   |
| $(4, d_1, 1)$     | 3                  | $d_1 \geq 5$                  | Corollary 4.2.17   |

Table 4.2: An overview of architectures for which we give results about their neuromanifolds in Subsection 4.2.1.

**Example 4.2.6.** The ideal of the neurovariety for the architecture  $\mathbf{d} = (3, 2, 2), r = 2$  is determinantal, and it is generated by the  $3 \times 3$  minors of

$$\begin{pmatrix} c_{11}^{(1)} & \frac{1}{2}c_{12}^{(1)} & \frac{1}{2}c_{13}^{(1)} & c_{11}^{(2)} & \frac{1}{2}c_{12}^{(2)} & \frac{1}{2}c_{13}^{(2)} \\ \frac{1}{2}c_{12}^{(1)} & c_{22}^{(1)} & \frac{1}{2}c_{23}^{(1)} & \frac{1}{2}c_{12}^{(2)} & c_{22}^{(2)} & \frac{1}{2}c_{23}^{(2)} \\ \frac{1}{2}c_{13}^{(1)} & \frac{1}{2}c_{23}^{(1)} & c_{33}^{(1)} & \frac{1}{2}c_{13}^{(2)} & \frac{1}{2}c_{23}^{(2)} & c_{33}^{(2)} \end{pmatrix}.$$

The ideal of the neurovariety for the architecture  $\mathbf{d} = (3, 2, 3), r = 2$  contains the ideal  $I_{\det}$  generated by the  $3 \times 3$  minors of the following  $3 \times 9$  matrix:

$$\begin{pmatrix} c_{11}^{(1)} & \frac{1}{2}c_{12}^{(1)} & \frac{1}{2}c_{13}^{(1)} & c_{11}^{(2)} & \frac{1}{2}c_{12}^{(2)} & \frac{1}{2}c_{13}^{(2)} & c_{11}^{(3)} & \frac{1}{2}c_{12}^{(3)} & \frac{1}{2}c_{13}^{(3)} \\ \frac{1}{2}c_{12}^{(1)} & c_{22}^{(1)} & \frac{1}{2}c_{23}^{(1)} & \frac{1}{2}c_{12}^{(2)} & c_{22}^{(2)} & \frac{1}{2}c_{23}^{(2)} & \frac{1}{2}c_{12}^{(3)} & c_{22}^{(3)} & \frac{1}{2}c_{23}^{(3)} \\ \frac{1}{2}c_{13}^{(1)} & \frac{1}{2}c_{23}^{(1)} & c_{33}^{(1)} & \frac{1}{2}c_{13}^{(2)} & \frac{1}{2}c_{23}^{(2)} & c_{33}^{(2)} & \frac{1}{2}c_{13}^{(3)} & \frac{1}{2}c_{23}^{(3)} & c_{33}^{(3)} \end{pmatrix}.$$

However, the ideal of  $\mathcal{V}_{(3,2,2),2}$  not equal to  $I_{\det}$ . The former is minimally generated by 94 polynomials. The dimension of  $\mathcal{V}(I_{\det})$  is eleven while  $\mathcal{V}_{(3,2,2),2}$  has dimension ten.  $\diamond$

## 4.2.1. Neuromanifolds and symmetric tensor decomposition

In this subsection, we investigate the relationship between shallow PNNs and symmetric tensor decompositions. Recall the definitions of symmetric tensors and symmetric tensor rank from Subsection 2.3.1. Building on this connection, we derive results for neuromanifolds for some shallow PNNs. This subsection has two parts: firstly, we focus on neuromanifolds with activation degree  $r = 2$ . Afterwards we present results on neuromanifolds for general  $r \in \mathbb{N}$ . An overview of the results in this subsection can be found in Table 4.2.

Consider a PNN with architecture  $\mathbf{d} = (d_0, d_1, 1)$ . Then  $\mathcal{M}_{\mathbf{d}, r}$  consists of homogeneous polynomials of degree  $r$  in  $d_0$  many variables that can be represented as a linear combination of  $d_1$  many linear forms raised to the  $r^{\text{th}}$  power. We should emphasise here that the linear combinations are taken over the *real* numbers. There is a bijection between the set of homogeneous polynomials of degree  $r$  in  $d_0$  many variables and the set of order- $r$  symmetric tensors of format  $d_0 \times d_0 \times \cdots \times d_0$  as explained in Subsection 2.3.1. Formulated in the language of tensors, the neuromanifold  $\mathcal{M}_{\mathbf{d}, r}$  consists of order- $r$  symmetric tensors of format  $d_0 \times d_0 \times \cdots \times d_0$  with *real symmetric rank*  $\leq d_1$ .

## Neuromanifolds for $r = 2$

**Lemma 4.2.7.** *The neuromanifold for the architecture  $\mathbf{d} = (d_0, d_1, 1), r = 2$  consists of  $(d_0 \times d_0)$  symmetric matrices*

$$\begin{pmatrix} c_{11} & \frac{1}{2}c_{12} & \cdots & \frac{1}{2}c_{1d_0} \\ \frac{1}{2}c_{12} & c_{22} & \cdots & \frac{1}{2}c_{2d_0} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{2}c_{1d_0} & \frac{1}{2}c_{2d_0} & \cdots & c_{d_0 d_0} \end{pmatrix} \quad (4.2.1)$$

of rank at most  $d_1$ . It is equal to its neurovariety and the ideal of the neurovariety is generated by all  $d_1 + 1$  minors of the symmetric matrix. The architecture is filling if and only if  $d_1 \geq d_0$ .

*Proof.* The neural network with architecture  $\mathbf{d} = (d_0, d_1, 1), r = 2$  and weights  $\mathbf{w}$  is

$$p_{\mathbf{w}}(x) = w_{211}(w_{111}x_1 + \dots + w_{11d_0}x_{d_0})^2 + \dots + w_{21d_1}(w_{1d_11}x_1 + \dots + w_{1d_1d_0}x_{d_0})^2.$$

Each of the summands of  $p_{\mathbf{w}}(x)$  corresponds to a rank-one symmetric matrix. Since there are  $d_1$  summands, their sum gives a symmetric matrix (4.2.1) of rank at most  $d_1$ .  $\square$

The next lemma is a special case of Lemma 4.2.11, and we state it here separately as the special case is used in Table 4.1.

**Lemma 4.2.8.** *The neuromanifold for the architecture  $\mathbf{d} = (d_0, 1, d_2), r = 2$  consists of  $d_2$  tuples of  $d_0 \times d_0$  symmetric matrices of rank at most one such that all the rank-one matrices are multiples of each other. The neuromanifold is equal to the neurovariety and the ideal of the neurovariety is generated by the  $2 \times 2$  minors of*

$$\begin{pmatrix} c_{11}^{(1)} & \frac{1}{2}c_{12}^{(1)} & \cdots & \frac{1}{2}c_{1d_0}^{(1)} & \cdots & c_{11}^{(d_2)} & \frac{1}{2}c_{12}^{(d_2)} & \cdots & \frac{1}{2}c_{1d_0}^{(d_2)} \\ \frac{1}{2}c_{12}^{(1)} & c_{22}^{(1)} & \cdots & \frac{1}{2}c_{2d_0}^{(1)} & \cdots & \frac{1}{2}c_{12}^{(d_2)} & c_{22}^{(d_2)} & \cdots & \frac{1}{2}c_{2d_0}^{(d_2)} \\ \vdots & \vdots & \ddots & \vdots & \cdots & \vdots & \vdots & \ddots & \vdots \\ \frac{1}{2}c_{1d_0}^{(1)} & \frac{1}{2}c_{2d_0}^{(1)} & \cdots & c_{d_0 d_0}^{(1)} & \cdots & \frac{1}{2}c_{1d_0}^{(d_2)} & \frac{1}{2}c_{2d_0}^{(d_2)} & \cdots & c_{d_0 d_0}^{(d_2)} \end{pmatrix}.$$

The next result was communicated to us by Maksym Zubkov and Leonie Kayser.

**Lemma 4.2.9.** *Let  $d_0, d_2 \geq 2$ , then for architectures  $\mathbf{d} = (d_0, d_0, d_2), r = 2$ , the parameter map  $\Psi_{\mathbf{d}, r}: \mathbb{R}^{d_0 \times d_0} \times \mathbb{R}^{d_2 \times d_0} \rightarrow (\text{Sym}_2(\mathbb{R}^{d_0}))^{d_2}$  is not surjective.*

*Proof.* We will first prove the statement for  $\mathbf{d} = (d_0, d_0, 2), r = 2$ . Assume to the contrary that the map  $\Psi_{\mathbf{d}, r}$  is surjective. For  $i \in [d_0]$ , we write  $\ell_i = (W_1 \mathbf{x})_i = w_{1i1}x_1 + \dots + w_{1id_0}x_{d_0}$ . For any pair of quadratic forms  $(q_1, q_2) \in \text{Sym}_2(\mathbb{R}^{d_0}) \times \text{Sym}_2(\mathbb{R}^{d_0})$ , there exists  $\mathbf{w}$  with

$$\Psi_{\mathbf{d}, r}(\mathbf{w}) = \begin{pmatrix} w_{211}\ell_1^2 + \dots + w_{21d_0}\ell_{d_0}^2 \\ w_{221}\ell_1^2 + \dots + w_{22d_0}\ell_{d_0}^2 \end{pmatrix} = \begin{pmatrix} \ell_1 w_{211}\ell_1 + \dots + \ell_{d_0} w_{21d_0}\ell_{d_0} \\ \ell_1 w_{221}\ell_1 + \dots + \ell_{d_0} w_{22d_0}\ell_{d_0} \end{pmatrix} = \begin{pmatrix} \mathbf{x}^T W_1^T D_1 W_1 \mathbf{x} \\ \mathbf{x}^T W_1^T D_2 W_1 \mathbf{x} \end{pmatrix},$$

where  $D_i$  is the diagonal matrix with  $w_{2ij}$  for  $i = 1, 2$  and  $j = 1, \dots, d_0$ . So, if  $A_1$  and  $A_2$  are the corresponding matrices for quadratic forms  $q_1$  and  $q_2$ , then we would have

$$A_1 = W_1^T D_1 W_1, \quad A_2 = W_1^T D_2 W_1. \quad (4.2.2)$$

Hence, the two matrices  $A_1$  and  $A_2$  are simultaneously congruent to diagonal matrices. Suppose this were true. Then the generalised characteristic polynomial (confer [HJ85, 2.4.P15])  $p_{A_1, A_2}(t_1, t_2) := \det(t_1 A_1 + t_2 A_2)$  can be written as

$$p_{A_1, A_2}(t_1, t_2) = \det(W_1)^2 \prod_{j=1}^{d_0} (w_{21j}t_1 + w_{22j}t_2),$$

i.e. it decomposes into linear factors. Over  $\mathbb{R}$ , one can easily find examples where this is violated. Therefore,  $\Psi_{\mathbf{d},r}$  is not surjective. If  $d_2 \geq 2$ , this corresponds to simultaneous diagonalisation by congruence of  $d_2$  many matrices which is again not always possible.  $\square$

**Remark 4.2.10.** If one of the matrices  $A_1$  or  $A_2$  from (4.2.2) above is definite, then they are in fact simultaneously congruent to diagonal matrices, see [HJ85, Thm. 7.6.1], so such pairs lie in the neuromanifold  $\mathcal{M}_{\mathbf{d},r}$  with  $\mathbf{d}$  and  $r$  as in Lemma 4.2.9.

### Neuromanifolds for $r \in \mathbb{N}$

**Lemma 4.2.11.** *The neuromanifold for the architecture  $\mathbf{d} = (d_0, 1, d_2), r \in \mathbb{N}$  consists of  $d_2$  order- $r$  ( $d_0 \times d_0 \times \dots \times d_0$ ) symmetric tensors of rank at most one such that all the rank-one tensors are multiples of each other. The neuromanifold equals the neurovariety and its ideal is generated by the  $2 \times 2$  minors of the flattenings of the  $(d_0 \times \dots \times d_0 \times d_0 d_2)$  tensor that is obtained from combining the  $d_2$  ( $d_0 \times d_0 \times \dots \times d_0$ ) symmetric tensors along the last index.*

*Proof.* The neural network with architecture  $\mathbf{d} = (d_0, 1, d_2), r \in \mathbb{N}$  and weights  $\mathbf{w}$  is

$$p_{\mathbf{w}}(x) = \begin{pmatrix} w_{211}(w_{111}x_1 + \dots + w_{11d_0}x_{d_0})^r \\ w_{221}(w_{111}x_1 + \dots + w_{11d_0}x_{d_0})^r \\ \vdots \\ w_{2d_21}(w_{111}x_1 + \dots + w_{11d_0}x_{d_0})^r \end{pmatrix}.$$

The form  $(w_{111}x_1 + \dots + w_{11d_0}x_{d_0})^r$  corresponds to an order- $r$  rank-one symmetric tensor. The  $i^{\text{th}}$  component of  $p_{\mathbf{w}}(x)$  is the symmetric tensor multiplied by the constant  $w_{2i1}$ . This proves the first statement. For the second statement, we note that combining the rank-one tensors to a  $d_0 \times \dots \times d_0 \times d_0 d_2$  tensor along the last index gives another rank-one tensor. Moreover, any  $d_0 \times \dots \times d_0 \times d_0 d_2$  tensor of rank one such that each of the  $d_2$  ( $d_0 \times d_0 \times \dots \times d_0$ )-subtensors is symmetric can be obtained from symmetric ( $d_0 \times d_0 \times \dots \times d_0$ ) rank-one tensors that are multiples of each other. The second statement follows from the result that the ideal of the tensors of rank at most one is generated by the  $2 \times 2$  minors of flattenings, see [Lan11, §3.4].  $\square$

**Proposition 4.2.12.** *Let  $\mathbf{d} = (d_0, d_1, d_2)$  with  $d_1 \geq \binom{r+d_0-1}{r}$ . Then the neuromanifold itself is filling, i.e.  $\mathcal{M}_{\mathbf{d},r}$  equals its ambient space  $(\text{Sym}_r(\mathbb{R}^{d_0}))^{d_2}$ .*

*Proof.* Let  $N := \binom{r+d_0-1}{r}$ . Write the coefficients of  $p_{\mathbf{w}} = \Psi_{\mathbf{d},r}(\mathbf{w})$  as

$$C = \begin{pmatrix} (c_I^{(1)})_I \\ \vdots \\ (c_I^{(d_2)})_I \end{pmatrix} = \begin{pmatrix} w_{211} & w_{212} & \dots & w_{21d_1} \\ \vdots & \vdots & \ddots & \vdots \\ w_{2d_21} & w_{2d_22} & \dots & w_{2d_2d_1} \end{pmatrix} \begin{pmatrix} (W_{1,1})^I \\ \vdots \\ (W_{1,d_1})^I \end{pmatrix},$$

where  $I$  ranges over all multiindices in  $\binom{[d_0]}{r}$  and  $(W_{1,i})^I$  denotes the row consisting of all degree- $r$  monomials formed by entries of the  $i^{\text{th}}$  row of  $W_1$  with the corresponding multinomial coefficients. We want to show that any real  $d_2 \times N$  matrix  $C$  can be written in this form. If  $d_1 \geq N$  then for a general choice of parameters  $W_1$  this matrix has a left-inverse with real entries and hence the system above can be solved over the reals.  $\square$

**Definition 4.2.13.** Let  $\mathbb{K}$  be  $\mathbb{R}$  or  $\mathbb{C}$  and let

$$S_{d,k}^n(\mathbb{K}) = \{T \in \text{Sym}_d(\mathbb{K}^n) : \text{symmetric rank of } T \text{ is } k\}.$$

If  $S_{d,k}^n(\mathbb{K}) \subseteq \text{Sym}_d(\mathbb{K}^n)$  has non-empty interior with respect to the Euclidean topology,  $k$  is called a *typical symmetric rank* for order- $d$  tensors of dimension  $n$ .

Over the complex numbers, there is a unique typical symmetric rank which is called the generic symmetric rank. However, over the real numbers there can be multiple typical symmetric ranks. In the following we will write  $S_{d,k}^n$  for  $S_{d,k}^n(\mathbb{R})$ . A neuromanifold  $\mathcal{M}_{\mathbf{d},r}$  having non-empty interior with respect to the Euclidean topology implies that the architecture  $(\mathbf{d}, r)$  is filling. The importance of typical symmetric ranks for PNNs comes from the following fact which is a consequence of the discussions above and the inclusion

$$S_{d,k}^n \subseteq \mathcal{M}_{(n,k,1),d}.$$

**Theorem 4.2.14.** *If  $k$  is a typical symmetric rank for order- $d$  tensors of dimension  $n$  then the architecture  $\mathbf{d} = (n, k, 1)$ ,  $r = d$  is filling, i.e.  $\mathcal{V}_{\mathbf{d},r} = \text{Sym}_d(\mathbb{R}^n)$ . Moreover, if there are two typical symmetric ranks  $k < k'$  then  $\text{cl}_{\text{Eucl}}(\mathcal{M}_{(n,k,1),d}) \subsetneq \text{Sym}_d(\mathbb{R}^n)$  where  $\text{cl}_{\text{Eucl}}$  denotes the closure with respect to the Euclidean topology.*

Let us consider the case  $n = 2$  first; this is the setting of the paper [CO12]. We summarise their findings below.

**Theorem 4.2.15** ([CO12, Prop. 2.2 & Main Theorem]).

1.  $S_{3,k}^2$  has non-empty interior only for  $k \in \{2, 3\}$ .
2.  $S_{4,k}^2$  has non-empty interior only for  $k \in \{3, 4\}$ .
3.  $S_{5,k}^2$  has non-empty interior only for  $k \in \{3, 4, 5\}$ .

Using these results we obtain the following as a consequence of Theorem 4.2.14.

**Corollary 4.2.16.**

1.  $\mathbf{d} = (2, d_1, 1)$ ,  $r = 3$  is filling for  $d_1 \geq 2$  and

$$\text{cl}_{\text{Eucl}}(\mathcal{M}_{(2,2,1),3}) \subsetneq \text{cl}_{\text{Eucl}}(\mathcal{M}_{(2,3,1),3}) = \text{Sym}_3(\mathbb{R}^2).$$

2.  $\mathbf{d} = (2, d_1, 1)$ ,  $r = 4$  is filling for  $d_1 \geq 3$  and

$$\text{cl}_{\text{Eucl}}(\mathcal{M}_{(2,3,1),4}) \subsetneq \text{cl}_{\text{Eucl}}(\mathcal{M}_{(2,4,1),4}) = \text{Sym}_4(\mathbb{R}^2).$$

3.  $\mathbf{d} = (2, d_1, 1)$ ,  $r = 5$  is filling for  $d_1 \geq 3$  and

$$\text{cl}_{\text{Eucl}}(\mathcal{M}_{(2,3,1),5}) \subsetneq \text{cl}_{\text{Eucl}}(\mathcal{M}_{(2,4,1),5}) \subsetneq \text{cl}_{\text{Eucl}}(\mathcal{M}_{(2,5,1),5}) = \text{Sym}_5(\mathbb{R}^2).$$

Some similar results on typical symmetric ranks are also known for higher dimensional tensors though the difficulty increases quickly. We summarise some consequences below.

**Corollary 4.2.17.**

1.  $\mathbf{d} = (3, d_1, 1)$ ,  $r = 4$  is filling for  $d_1 \geq 6$  and

$$\text{cl}_{\text{Eucl}}(\mathcal{M}_{(3,6,1),4}) \subsetneq \text{cl}_{\text{Eucl}}(\mathcal{M}_{(3,7,1),4}) \subseteq \text{cl}_{\text{Eucl}}(\mathcal{M}_{(3,8,1),4}) = \text{Sym}_4(\mathbb{R}^3).$$

2.  $\mathbf{d} = (3, d_1, 1)$ ,  $r = 5$  is filling for  $d_1 \geq 7$  and

$$\text{cl}_{\text{Eucl}}(\mathcal{M}_{(3,7,1),5}) \subsetneq \text{cl}_{\text{Eucl}}(\mathcal{M}_{(3,8,1),5}) \subseteq \cdots \subseteq \text{cl}_{\text{Eucl}}(\mathcal{M}_{(3,13,1),5}) = \text{Sym}_5(\mathbb{R}^3).$$

3.  $\mathbf{d} = (4, d_1, 1)$ ,  $r = 3$  is filling for  $d_1 \geq 5$  and

$$\text{cl}_{\text{Eucl}}(\mathcal{M}_{(4,5,1),3}) \subsetneq \text{cl}_{\text{Eucl}}(\mathcal{M}_{(4,6,1),3}) = \text{Sym}_3(\mathbb{R}^4)$$

*Proof.* The statements follow from Theorem 4.2.14 and [BBO18, Theorem 1.2].  $\square$

The algebraic boundary  $\partial_{\text{alg}}(\mathcal{M})$  of a set  $\mathcal{M}$  is the Zariski closure of its topological boundary  $\partial(\mathcal{M})$ . Little is known about the algebraic boundaries in the cases considered above. Theorem 4.1 in [MMSV16] implies that the algebraic boundary  $\partial_{\text{alg}}(\mathcal{M}_{(3,6,1),4})$  has an irreducible component of degree 51.

### 4.2.2. Neurovarieties

In this subsection, we describe different approaches to studying neurovarieties. We start by studying the fibres of a parameter map to characterise the neurovariety  $\mathcal{V}_{(2,2,d_2),2}$  (Proposition 4.2.18) and partially the neuromanifold  $\mathcal{M}_{(2,2,d_2),2}$  (Proposition 4.2.19). We use Grassmannians to describe the neurovariety  $\mathcal{V}_{(3,3,3),2}$  (Example 4.2.21) and apply the Hilbert–Burch Theorem to study  $\mathcal{V}_{(2,2,2,2),2}$  (Example 4.2.22).

**Proposition 4.2.18.** *The neurovariety  $\mathcal{V}_{(2,2,d_2),2}$  is the vanishing locus of all  $3 \times 3$  minors of*

$$C_{d_2} = \begin{pmatrix} c_{11}^{(1)} & c_{12}^{(1)} & c_{22}^{(1)} \\ \vdots & \vdots & \vdots \\ c_{11}^{(d_2)} & c_{12}^{(d_2)} & c_{22}^{(d_2)} \end{pmatrix} = \begin{pmatrix} w_{211} & w_{212} \\ \vdots & \vdots \\ w_{2d_21} & w_{2d_22} \end{pmatrix} \begin{pmatrix} w_{111}^2 & 2w_{111}w_{112} & w_{112}^2 \\ w_{121}^2 & 2w_{121}w_{122} & w_{122}^2 \end{pmatrix}.$$

*Proof.* Clearly  $\mathcal{V}_{(2,2,d_2),2} \subseteq \mathcal{V}(\langle 3 \times 3 \text{ minors of } C_{d_2} \rangle)$  as  $\text{rk}(C_{d_2}) \leq 2$ . The variety  $\mathcal{V}(\langle 3 \times 3 \text{ minors of } C_{d_2} \rangle)$  is irreducible and  $\dim(\mathcal{V}(\langle 3 \times 3 \text{ minors of } C_{d_2} \rangle)) = 2d_2 + 2$ , see e.g. [BV06, Proposition 1.1]. Consider the parameter map and the corresponding neurovariety

$$\Psi_{(2,2,d_2),2}: \underbrace{\mathbb{R}^{2 \times 2} \times \mathbb{R}^{d_2 \times 2}}_{\dim 2d_2 + 4} \rightarrow (\text{Sym}_2(\mathbb{R}^2))^{d_2}, \quad \mathcal{V}_{(2,2,d_2),2} = \overline{\text{im}(\Psi_{(2,2,d_2),2})}.$$

The fibres of  $\Psi_{(2,2,d_2),2}$  over  $\mathcal{V}_{(2,2,d_2),2}$  are at least two-dimensional by Lemma 4.2.4. For  $d_2 = 2$ , one can computationally check that over a generic matrix  $C_{d_2}$  the fibre is 2-dimensional. Since generic fibres are at least 2-dimensional, they are exactly 2-dimensional. For the case  $d_2 > 2$ , note that by setting the parameters  $w_{211}, w_{212}, \dots, w_{2(d_2-1)1}, w_{2(d_2-1)2}$  to zero and varying  $w_{2d_21}, w_{2d_22}$ , we obtain a two-dimensional family not contained in the embedding of  $\mathcal{M}_{(2,2,d_2-1),2}$  into  $\mathcal{M}_{(2,2,d_2),2}$ . As the dimension of the parameter space increases by two when passing from  $d_2 - 1$  to  $d_2$ , the generic fibre is two-dimensional by induction. Thus,  $\dim(\mathcal{V}_{(2,2,d_2),2}) = 2d_2 + 2$  and moreover  $\mathcal{V}_{(2,2,d_2),2}$  is irreducible, hence

$$\mathcal{V}_{(2,2,d_2),2} = \mathcal{V}(\langle 3 \times 3 \text{ minors of } C_{d_2} \rangle).$$

$\square$

**Proposition 4.2.19.** *For  $\mathbf{d} = (2, 2, d_2)$  ( $d_2 \geq 2$ ) and  $r = 2$ , the Euclidean closure of  $\mathcal{M}_{\mathbf{d},r}$  is strictly included in  $\mathcal{V}_{\mathbf{d},r}$ . In the case  $\mathbf{d} = (2, 2, 2)$ , we have*

$$C = \begin{pmatrix} c_{11}^{(1)} & c_{12}^{(1)} & c_{22}^{(1)} \\ c_{11}^{(2)} & c_{12}^{(2)} & c_{22}^{(2)} \end{pmatrix} \in \mathcal{M}_{(2,2,2),2}$$

if and only if

$$M_{1,3}(C)^2 \geq M_{1,2}(C)M_{2,3}(C),$$

where  $M_{i,j}(C)$  is the  $2 \times 2$  minor of  $C$  obtained by taking the determinant of the  $i^{\text{th}}$  and  $j^{\text{th}}$  column. In particular, the algebraic boundary of  $\mathcal{M}_{(2,2,2),2}$  is given by

$$\partial_{\text{alg}} \mathcal{M}_{(2,2,2),2} = \mathcal{V}(M_{1,3}(C)^2 - M_{1,2}(C)M_{2,3}(C)).$$

*Proof.* First consider the case  $\mathbf{d} = (2, 2, 2)$ . We are looking for conditions on a matrix  $A \in \mathbb{R}^{2 \times 3}$  such that there exists  $G \in GL(2, \mathbb{R})$  with

$$\begin{pmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{pmatrix} \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{pmatrix} = \begin{pmatrix} w_{111}^2 & 2w_{111}w_{112} & w_{112}^2 \\ w_{121}^2 & 2w_{121}w_{122} & w_{122}^2 \end{pmatrix}. \quad (4.2.3)$$

We consider two cases: when both  $w_{111}, w_{121}$  are nonzero and when at least one of them is zero. If  $w_{111}$  and  $w_{121}$  are nonzero, then after rescaling, without loss of generality we can assume  $w_{111} = w_{121} = 1$ . Each matrix entry in (4.2.3) yields a polynomial equation; these generate an ideal  $I$  in the polynomial ring  $\mathbb{R}[g_{11}, \dots, g_{22}, w_{112}, w_{122}, a_{11}, \dots, a_{23}]$ . Two generators in a Gröbner basis of  $I$  with respect to lexicographic ordering are

$$w_{1i2}^2 a_{11}a_{22} - w_{1i2}^2 a_{12}a_{21} - 2w_{1i2}a_{11}a_{23} + 2w_{1i2}a_{13}a_{21} + a_{12}a_{23} - a_{13}a_{22}, \quad (4.2.4)$$

where  $i \in \{1, 2\}$ . These quadratic equations in  $w_{1i2}$  have discriminant

$$\Delta = M_{1,3}(A)^2 - M_{1,2}(A)M_{2,3}(A),$$

where  $A$  is the  $2 \times 3$  matrix with entries  $a_{ij}$  from (4.2.3). The denominator of the quadratic formula for (4.2.4) is  $M_{1,2}(A)$ . Therefore, real solutions to (4.2.3) with  $w_{111}, w_{121}$  nonzero exist if and only if  $M_{1,3}(A)^2 \geq M_{1,2}(A)M_{2,3}(A)$  and  $M_{1,2}(A) \neq 0$ .

When at least one  $w_{1i1} = 0$ , then, possibly after rescaling, the Gröbner basis of  $I$  with respect to the lexicographic ordering contains  $w_{1i2}^2 a_{11}a_{22} - w_{1i2}^2 a_{12}a_{21}$ . This implies  $w_{1i2} = 0$  or  $M_{1,2}(A) = 0$ . The former condition implies  $M_{1,2}(A) = 0$ , so we do not have to consider it separately. When  $M_{1,2}(A) = 0$ , the inequality  $M_{1,3}(A)^2 \geq M_{1,2}(A)M_{2,3}(A)$  is automatically satisfied. Therefore, real solutions to (4.2.3) with at least one of  $w_{111}, w_{121}$  zero exist if and only if  $M_{1,3}(A)^2 \geq M_{1,2}(A)M_{2,3}(A)$  and  $M_{1,2}(A) = 0$ . Combining the two cases gives that real solutions to (4.2.3) exist if and only if  $M_{1,3}(A)^2 \geq M_{1,2}(A)M_{2,3}(A)$ .

As this inequality remains unchanged after multiplying  $A$  with a  $2 \times 2$  matrix (both sides get multiplied by the squared determinant of the matrix which is positive), we conclude  $C \in \mathcal{M}_{(2,2,2),2}$  if and only if  $M_{1,3}(C)^2 \geq M_{1,2}(C)M_{2,3}(C)$ .

In the case  $d_2 > 2$ , it is clear that at least for any two rows  $i, j$  of  $C$ , the inequality

$$M_{1,3}(C_{i,j})^2 \geq M_{1,2}(C_{i,j})M_{2,3}(C_{i,j})$$

needs to hold. Thus, we get a full-dimensional set of points in  $\mathcal{V}_{\mathbf{d},r} \setminus \mathcal{M}_{\mathbf{d},r}$ , hence

$$\text{cl}_{\text{Eucl}}(\mathcal{M}_{\mathbf{d},r}) \subsetneq \mathcal{V}_{\mathbf{d},r}.$$

□

**Remark 4.2.20.** For example, if  $A = \begin{pmatrix} a & * & -a \\ b & * & -b \end{pmatrix}$  in the proof of Proposition 4.2.19, then the equation

$$\begin{pmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{pmatrix} \begin{pmatrix} a & * & -a \\ b & * & -b \end{pmatrix} = \begin{pmatrix} w_{111}^2 & 2w_{111}w_{112} & w_{112}^2 \\ w_{121}^2 & 2w_{121}w_{122} & w_{122}^2 \end{pmatrix}$$

does not have a solution. This example can be extended to any architecture  $\mathbf{d} = (d_0, 2, d_2)$  with  $d_0, d_2 \geq 2$  and activation degree  $r = 2$  to show that  $\mathcal{M}_{\mathbf{d},r} \neq \mathcal{V}_{\mathbf{d},r}$ .

In the rest of the subsection, we use more advanced tools from algebraic geometry to compute neurovarieties. A naïve elimination approach fails to compute these varieties.

**Example 4.2.21.** Consider the architecture  $\mathbf{d} = (3, 3, 3)$ ,  $r = 2$ ; the image of  $\Psi_{\mathbf{d}, r}$  consists of three linear combinations of three squares of linear forms in three variables. Let us take a  $3 \times 6$  matrix representing the embedding of  $\mathbb{P}_x^2 \times \mathbb{P}_y^2 \times \mathbb{P}_z^2$  via the degree two Veronese map (with corresponding scaling) into  $\mathbb{P}^5 \times \mathbb{P}^5 \times \mathbb{P}^5$ :

$$\begin{pmatrix} x_0^2 & 2x_0x_1 & 2x_0x_2 & x_1^2 & 2x_1x_2 & x_2^2 \\ y_0^2 & 2y_0y_1 & 2y_0y_2 & y_1^2 & 2y_1y_2 & y_2^2 \\ z_0^2 & 2z_0z_1 & 2z_0z_2 & z_1^2 & 2z_1z_2 & z_2^2 \end{pmatrix}$$

By taking all the  $3 \times 3$  minors (20 in total) of this matrix, we get an embedding into the Grassmannian  $\text{Gr}(3, 6)$  in its Plücker coordinates. The ideal of this variety in Plücker coordinates can be found using elimination. Its dimension is six and its degree is 57. Now for each of the 20 Plücker coordinates we substitute a  $3 \times 3$  minor of a  $3 \times 6$  matrix of unknowns, where each unknown represents a coefficient of a degree two monomial (we view  $\mathbb{P}^5$  as the space of all homogeneous quadrics in three variables):

$$C = \begin{pmatrix} c_{11}^{(1)} & c_{12}^{(1)} & \dots & c_{33}^{(1)} \\ c_{11}^{(2)} & c_{12}^{(2)} & \dots & c_{33}^{(2)} \\ c_{11}^{(3)} & c_{12}^{(3)} & \dots & c_{33}^{(3)} \end{pmatrix}$$

The resulting ideal  $I$  gives rise to a variety  $X = \mathcal{V}(I)$ . This variety has two irreducible components: one is given by the vanishing of all  $3 \times 3$  minors of  $C$ , the other is the neurovariety  $\mathcal{V}_{(3,3,3),2}$  embedded into  $\mathbb{P}_{c^{(1)}}^5 \times \mathbb{P}_{c^{(2)}}^5 \times \mathbb{P}_{c^{(3)}}^5$ . The codimension of  $\mathcal{V}_{(3,3,3),2} \subseteq \mathbb{P}_{c^{(1)}}^5 \times \mathbb{P}_{c^{(2)}}^5 \times \mathbb{P}_{c^{(3)}}^5$  is three, hence the affine dimension is  $\dim(\mathcal{V}_{\mathbf{d},r}) = 18 - 3 = 15$ .

The first part of the computation can be found in [BS11]; we implemented the whole procedure and made it available at [KLW24b].  $\diamond$

Now consider the architecture  $(\mathbf{d} = (2, 2, 2, 2), r = 2)$ . We expect that  $\mathcal{V}_{\mathbf{d},r}$  has dimension eight and thus has codimension two in  $(\text{Sym}_4(\mathbb{R}^2))^2$ . Therefore, we can hope that by the Hilbert–Burch Theorem (Theorem 2.3.5) we can find a matrix whose minors generate the ideal of  $\mathcal{V}_{\mathbf{d},r}$ . Again, we work in projective space, i.e. we consider  $\mathcal{V}_{\mathbf{d},r} \subseteq \mathbb{P}_{c^{(1)}}^4 \times \mathbb{P}_{c^{(2)}}^4$ . The following example was communicated to us by Bernd Sturmfels.

**Example 4.2.22.** One finds that the coordinate ring of  $\mathcal{V}_{(2,2,2,2),2} \subseteq \mathbb{P}_{c^{(1)}}^4 \times \mathbb{P}_{c^{(2)}}^4$  admits a Hilbert–Burch resolution. Therefore, the ideal defining  $\mathcal{V}_{(2,2,2,2),2}$  is generated by the  $5 \times 5$  minors of the matrix

$$\begin{pmatrix} c_{1112}^{(2)} & -c_{1111}^{(2)} & 0 & c_{1112}^{(1)} & -c_{1111}^{(1)} & 0 \\ 4c_{1122}^{(2)} & -c_{1112}^{(2)} & 8c_{1111}^{(2)} & 4c_{1122}^{(1)} & -c_{1112}^{(1)} & 8c_{1111}^{(1)} \\ 8c_{1222}^{(2)} & 0 & 8c_{1112}^{(2)} & 8c_{1222}^{(1)} & 0 & 8c_{1112}^{(1)} \\ 8c_{2222}^{(2)} & c_{1222}^{(2)} & 4c_{1122}^{(2)} & 8c_{2222}^{(1)} & c_{1222}^{(1)} & 4c_{1122}^{(1)} \\ 0 & c_{2222}^{(2)} & c_{1222}^{(2)} & 0 & c_{2222}^{(1)} & c_{1222}^{(1)} \end{pmatrix}. \quad \diamond$$

### 4.2.3. Dimension

Determining dimensions of neurovarieties is a difficult task and there exists a wide range of open conjectures in this direction. A classical result is the Alexander–Hirschowitz Theorem; already conjectured in the late 19<sup>th</sup> century, the proof was completed about a century later in 1995. Here is a formulation in the language of PNNs.

**Theorem 4.2.23** (Alexander–Hirschowitz [AH95]). *If  $\mathbf{d} = (d_0, d_1, 1)$ ,  $\mathcal{V}_{\mathbf{d},r}$  attains the expected dimension  $\min\{d_0 d_1, \binom{d_0+r-1}{r}\}$ , except for the following cases:*

1.  $r = 2$ ,  $2 \leq d_1 < d_0$ ,
2.  $r = 3$ ,  $d_0 = 5$ ,  $d_1 = 7$  where  $\dim(\mathcal{V}_{(5,7,1),3}) = 34$  (defect 1),
3.  $r = 4$ ,  $d_0 = 3$ ,  $d_1 = 5$  where  $\dim(\mathcal{V}_{(3,5,1),4}) = 14$  (defect 1),
4.  $r = 4$ ,  $d_0 = 4$ ,  $d_1 = 9$  where  $\dim(\mathcal{V}_{(4,9,1),4}) = 34$  (defect 1),
5.  $r = 4$ ,  $d_0 = 5$ ,  $d_1 = 14$  where  $\dim(\mathcal{V}_{(5,14,1),4}) = 69$  (defect 1).

It was conjectured that the neurovariety attains the expected dimension for large activation degree. Compare this for example with the Alexander–Hirschowitz Theorem which tells us there exist no defective single-output two-layer neurovarieties with activation degree  $r \geq 5$ . The conjecture appeared as a theorem in [KTB19], but a mistake in the proof was pointed out by Theo Elenius in his Bachelor thesis. After the article [KLW24a] (on which this chapter is based) first appeared as a preprint, the conjecture has been proven in [FRWY24, Thm. 12] and is stated precisely as follows.

**Theorem 4.2.24.** *For any fixed widths  $\mathbf{d} = (d_0, \dots, d_L)$  with  $d_i > 1$  for  $i = 1, \dots, L-1$ , there exists  $\tilde{r} = \tilde{r}(\mathbf{d})$  such that whenever  $r > \tilde{r}$ , the neurovariety  $\mathcal{V}_{\mathbf{d},r}$  attains the expected dimension.*

It is, however, not true that  $\mathcal{V}_{\mathbf{d},r}$  being defective implies that  $\mathcal{V}_{\mathbf{d},r'}$  is defective for all  $r' < r$ . For example,  $\mathcal{V}_{(5,7,1),3}$  has defect one whereas  $\mathcal{V}_{(5,7,1),2}$  is non-defective. The assumption  $d_i > 1$  for  $i = 1, \dots, L-1$  in the theorem is necessary as is shown in the following example.

**Example 4.2.25.** Consider the widths  $\mathbf{d} = (2, 1, 2, 1)$ . We claim that for any  $r > 1$ ,  $\mathcal{V}_{\mathbf{d},r}$  has defect one, i.e.  $\dim(\mathcal{V}_{\mathbf{d},r}) = 2$ . Indeed, observe that the parameter map  $\Psi_{\mathbf{d},r}$  is given by

$$\begin{aligned} \mathbf{x} \mapsto (w_{111}x_1 + w_{112}x_2)^r &\mapsto \begin{pmatrix} w_{211}^r (w_{111}x_1 + w_{112}x_2)^{r^2} \\ w_{221}^r (w_{111}x_1 + w_{112}x_2)^{r^2} \end{pmatrix} \\ &\mapsto (w_{311}w_{211}^r + w_{312}w_{221}^r)(w_{111}x_1 + w_{112}x_2)^{r^2}. \end{aligned}$$

As all weights coming from the last two layers can be factored out, we get that  $\mathcal{V}_{\mathbf{d},r} \cong \mathcal{V}_{(2,1,1,1),r}$ . The latter neurovariety is immediately seen to have dimension two. This argument generalises to the following statement.  $\diamond$

**Proposition 4.2.26.** *Let  $\mathbf{d}_0 \in \mathbb{N}^{n_0}$  and  $\mathbf{d}_2 \in \mathbb{N}^{n_2}$ . For  $\mathbf{d} = (\mathbf{d}_0, 1, \mathbf{d}_2, d_3)$ , the neuromanifold  $\mathcal{M}_{\mathbf{d},r}$  is equal to  $\mathcal{M}_{(\mathbf{d}_0, 1, 1, d_3),r}$  for any  $r$ . In particular, for  $d_3 = 1$ ,  $\dim(\mathcal{M}_{\mathbf{d},r}) = \dim(\mathcal{M}_{(\mathbf{d}_0, 1),r})$ .*

*Proof.* Fix parameters  $\mathbf{w} = (W_1, \dots, W_L)$ . Let  $p = \Psi_{(\mathbf{d}_0, 1),r}(\pi_1(\mathbf{w}))$  and  $q = \Psi_{(1, \mathbf{d}_2, 1)}(\pi_2(\mathbf{w}))$ , where  $\pi_1$  denotes the projection to the parameter space corresponding to the first  $n_0 + 1$  layers, and  $\pi_2$  denotes the projection to the remaining parameters. Then  $p_{\mathbf{w}}(\mathbf{x})$  factors as a composition  $p_{\mathbf{w}}(\mathbf{x}) = (q \circ p)(\mathbf{x})$ . The polynomial  $q$  is a single-input neural network, hence we can write  $(q \circ p)_i(\mathbf{x})$  as a product  $\alpha_i(\pi_2(\mathbf{w})) \cdot p(\mathbf{x})^{r^{n_2}}$  where  $\alpha_i(\pi_2(\mathbf{w}))$  only depends on weights corresponding to the last  $n_2 + 2$  layers and thus

$$\mathcal{M}_{\mathbf{d},r} = \{(\alpha_i \cdot p^{r^{n_2}})_i : \alpha_i \in \mathbb{R}, p \in \mathcal{M}_{(\mathbf{d}_0, 1),r}\}.$$

The latter set is by definition equal to  $\mathcal{M}_{(\mathbf{d}_0, 1, 1, d_3),r}$ . In the case  $d_3 = 1$ , its dimension is equal to  $\dim(\mathcal{M}_{(\mathbf{d}_0, 1),r})$ .  $\square$

**Remark 4.2.27.** Proposition 4.2.26 allows to give a strong bound on the dimension of  $\mathcal{M}_{\mathbf{d},r}$  for widths  $\mathbf{d} = (\mathbf{d}_0, 1, \mathbf{d}_2, d_3)$ :

$$\dim(\mathcal{M}_{\mathbf{d},r}) = \dim(\mathcal{M}_{(\mathbf{d}_0, 1, \mathbf{d}_2, d_3),r}) \leq \text{edim}(\mathcal{M}_{(\mathbf{d}_0, 1, \mathbf{d}_2, d_3),r}) = \dim(\mathcal{M}_{\mathbf{d}_0,r}) + d_3 - 1.$$

If  $d_0 > 1$ , the RHS is strictly less than the ambient dimension of  $\mathcal{M}_{\mathbf{d},r}$  which is  $d_3 \binom{r^{L-1} + d_0 - 1}{r^{L-1}}$ , where  $d_0$  is the width of the input layer. Hence, the width one at the  $(n_0 + 1)^{\text{st}}$  layer is an *asymptotic bottleneck*, i.e. the architecture is not filling regardless of adding more layers or changing the widths of all layers except the input and  $(n_0 + 1)^{\text{st}}$  layer [KTB19, Def. 18].

**Remark 4.2.28.** For  $\mathbf{d} = (d_0, 1, d_2, 1)$ , the neuromanifold  $\mathcal{M}_{\mathbf{d},r}$  is equal to  $\mathcal{M}_{(d_0, 1, 1, 1),r}$  for any  $r$ . In particular,  $\dim(\mathcal{M}_{\mathbf{d},r}) = d_0$ . This allows to compute the dimension of many defective neurovarieties. For example,  $\dim(\mathcal{V}_{(3,1,5,1),3}) = 3$  and therefore  $\mathcal{V}_{(3,1,5,1),3}$  has defect 4.

In contrast to the asymptotic statement for large activation degree in Theorem 4.2.24, we also conjecture the following.

**Conjecture 4.2.29.** Let  $\mathbf{d} = (d_0, d_1, \dots, d_L)$  be a non-increasing sequence with  $d_L > 1$ . Then for any  $r$ , the neurovariety  $\mathcal{V}_{\mathbf{d},r}$  attains the expected dimension.

We verified this conjecture using Algorithm 1 described in Section 4.1 and the paragraph below Conjecture 4.2.33 for four- and five-layer PNNs with widths up to three and activation degree up to five, see [KLW24b]. In the context of feedforward neural networks with ReLU activation the analogous statement has been proven [GLMW22, Thm. 8.11].

Using Lemma 4.2.4, one can deduce the following recursive dimension bound.

**Proposition 4.2.30** ([KTB19, Prop. 17]). For any  $r \in \mathbb{N}$ ,  $\mathbf{d} = (d_0, \dots, d_i, \dots, d_L)$  and  $i \in \{1, \dots, L-1\}$ , we have

$$\dim(\mathcal{V}_{\mathbf{d},r}) \leq \dim(\mathcal{V}_{(d_0, \dots, d_i),r}) + \dim(\mathcal{V}_{(d_i, \dots, d_L),r}) - d_i.$$

**Corollary 4.2.31.** Let  $\tilde{\mathbf{d}} = (d_0, \dots, d_{i-1}, \mathbf{d}, d_{j+1}, \dots, d_L)$ , where  $\mathbf{d} = (d_i, d_{i+1}, \dots, d_j)$  is an architecture such that  $\mathcal{V}_{\mathbf{d},r}$  is defective for some  $r > 0$ . Here, we allow  $i = 0$  or  $j = L$ . Suppose  $\mathcal{V}_{\tilde{\mathbf{d}},r}$  is not filling, then  $\mathcal{V}_{\tilde{\mathbf{d}},r}$  is defective.

*Proof.* It follows from Proposition 4.2.30 that

$$\dim(\mathcal{V}_{\tilde{\mathbf{d}},r}) \leq \dim(\mathcal{V}_{(d_0, \dots, d_i),r}) + \dim(\mathcal{V}_{\mathbf{d},r}) + \dim(\mathcal{V}_{(d_j, \dots, d_L),r}) - d_i - d_j \quad (4.2.5)$$

where  $\dim(\mathcal{V}_{(d_0, \dots, d_i),r})$  or  $\dim(\mathcal{V}_{(d_j, \dots, d_L),r})$  might not appear in the sum above. By Definition 4.2.5, the expected dimension is defined as

$$\text{edim}(\mathcal{V}_{\mathbf{d},r}) = \min \left\{ d_L + \sum_{i=0}^{L-1} (d_i d_{i+1} - d_{i+1}), d_L \binom{d_0 + r^{L-1} - 1}{r^{L-1}} \right\}.$$

Since  $\dim(\mathcal{V}_{\mathbf{d},r}) < \text{edim}(\mathcal{V}_{\mathbf{d},r})$ , it follows from equation (4.2.5) that

$$\begin{aligned} \dim(\mathcal{V}_{\tilde{\mathbf{d}},r}) &< d_L + \sum_{\alpha=0}^{i-1} (d_\alpha d_{\alpha+1} - d_{\alpha+1}) + \sum_{\beta=i}^{j-1} (d_\beta d_{\beta+1} - d_{\beta+1}) + \sum_{\gamma=j}^L (d_\gamma d_{\gamma+1} - d_{\gamma+1}) \\ &= d_L + \sum_{\alpha=0}^{L-1} (d_\alpha d_{\alpha+1} - d_{\alpha+1}), \end{aligned}$$

where  $\sum_{\alpha=0}^{i-1} (d_\alpha d_{\alpha+1} - d_{\alpha+1})$  or  $\sum_{\gamma=j}^L (d_\gamma d_{\gamma+1} - d_{\gamma+1})$  might not appear in the sum above. Since  $\mathcal{V}_{\tilde{\mathbf{d}},r}$  is non-filling, it follows directly that  $\mathcal{V}_{\tilde{\mathbf{d}},r}$  is defective.  $\square$

**Remark 4.2.32.** The converse to this statement is not true. For example, consider  $\mathbf{d} = (2, 2, 1, 2)$ ,  $r = 2$ . Then  $\mathcal{V}_{\mathbf{d},r}$  has defect one, but both  $\mathcal{V}_{(2,2,1),2}$  and  $\mathcal{V}_{(2,1,2),2}$  are non-defective.

Let  $\preccurlyeq$  be the partial order on the set of widths induced by coordinatewise comparison. The following conjecture suggests a unimodal distribution of layer widths within a neural network is efficient. In the realm of machine learning theory, this architectural pattern is commonly believed to allow the network to initially expand its capacity for feature representation, enabling the network to capture intricate patterns within the data. As the layers narrow, heuristically, the network would refine these features into more sophisticated representations suitable for making predictions. The conjecture agrees with this heuristic.

**Conjecture 4.2.33** ([KTB19], Conj. 12). *Fix  $L, d_0, d_L$  and  $r$ ; any minimal (w.r.t.  $\preccurlyeq$ ) vector of widths  $\mathbf{d} = (d_0, d_1, \dots, d_L)$  such that the architecture  $\mathcal{V}_{\mathbf{d},r}$  is filling, is unimodal, i.e. there exists  $i \in \{0, 1, \dots, L\}$  such that  $(d_0, \dots, d_i)$  is weakly increasing and  $(d_i, \dots, d_L)$  is weakly decreasing.*

To compute the dimension of a neurovariety for a fixed architecture it is advantageous to employ the backpropagation algorithm (Algorithm 1). This algorithm is commonly used in machine learning applications to compute the gradient in the update rule (2.5.3) of the gradient descent algorithm. One can also use it to compute the dimension of a neurovariety. This approach provides an exponential speed-up compared to a direct computation of the rank of the Jacobian of the parametrisation  $\Psi_{\mathbf{d},r}$ . Since many varieties in the context of tensors can be recast as neurovarieties (see Subsection 4.2.1), the approach described below might also prove useful for other dimension computations in algebraic statistics.

Algorithm 1 computes the gradient of a loss function  $\ell$  of a neural network evaluated at a specific sample. If we choose the loss function  $\ell$  to be the network  $F_\theta$  itself, backpropagation returns the gradient  $\nabla_{\mathbf{w}} F_\theta(\hat{\mathbf{x}})$ . However, when we are interested in computing the Jacobian of  $\Psi_{\mathbf{d},r}$ , we need to compute the partial derivatives  $\partial_{w_{j,k,l}} c_I^{(i)}$ , whereas

$$\frac{\partial F_\theta^{(i)}}{\partial w_{j,k,l}}(\hat{\mathbf{x}}) = \sum_I \frac{\partial c_I^{(i)}}{\partial w_{j,k,l}} \hat{\mathbf{x}}^I. \quad (4.2.6)$$

By choosing  $N := \binom{r^{L-1} + d_0 - 1}{d_0 - 1}$  many samples  $\hat{\mathbf{x}}_1, \hat{\mathbf{x}}_2, \dots, \hat{\mathbf{x}}_N$  and applying backpropagation to each of them, we obtain a linear system from (4.2.6) in the unknowns  $\partial_{w_{j,k,l}} c_I^{(i)}$ . For generic samples, this system has a unique solution, and we obtain the Jacobian  $J_{\Psi_{\mathbf{d},r}}$ .

This routine allows us to compute Jacobians and in particular the dimension of the neurovariety exponentially faster than via the naïve method of computing all derivatives of the parametrisation  $\Psi_{\mathbf{d},r}$ . It can be easily adapted for computations over finite fields which makes dimension computations even faster. In [KTB19] this routine has already been implemented, however (as of Jan 22 2025), their implementation at [KTB] contains a mistake in the set-up of the linear system described above. We provide a corrected implementation at [KLW24b].

#### 4.2.4. The symmetries of an exceptional shallow network

We explicitly describe a family of symmetries beyond multi-homogeneity (Lemma 4.2.4) for the architecture  $\mathbf{d} = (5, 7, 1)$ ,  $r = 3$ , one of the exceptional cases of the Alexander–Hirschowitz Theorem. The expected dimension of  $\mathcal{V}_{\mathbf{d},r}$  is 35, however, its actual dimension is 34. Therefore, there exists a one-dimensional family of symmetries in the parameter space not of the form as in Lemma 4.2.4.

In the following we will use an equivalent formulation of the Alexander–Hirschowitz Theorem as can be found in [BO08, Theorem 1.1].

**Theorem 4.2.34** (Alexander–Hirschowitz, geometric version). *Let  $X$  be a general collection of  $k$  double points in  $\mathbb{P}^n$  and let  $I_X(d) \subseteq \text{Sym}_d(\mathbb{C}^n)$  be the subspace of polynomials through  $X$ , i.e. with all first partial derivatives vanishing at the points of  $X$ . Then the subspace  $I_X(d)$  has the expected codimension  $\min\{(n+1)k, \binom{n+d}{n}\}$  except in the cases as in Theorem 4.2.23 (with the notation  $d = r$ ,  $n = d_0 - 1$ ,  $k = d_1$ ).*

Geometrically the situation depicts as follows: consider seven points  $p_1, \dots, p_7$  in  $\mathbb{P}^4$ . We are interested in cubics singular at these seven points, i.e. we are looking for  $f \in \text{Sym}_3(\mathbb{C}^5)$  such that  $f(p_i) = df_{p_i} = 0$  for  $i = 1, \dots, 7$ . One would expect that no such cubic exists: the space of cubics in five variables has dimension 35 and the seven points are expected to impose 35 independent conditions. But through seven points there exists a rational normal curve  $C$  which after a suitable coordinate transformation is given as

$$C = \left\{ (x_0 : \dots : x_4) \in \mathbb{P}^4 : \text{rk} \begin{pmatrix} x_0 & x_1 & x_2 \\ x_1 & x_2 & x_3 \\ x_2 & x_3 & x_4 \end{pmatrix} \leq 1 \right\}.$$

A different way to write the curve  $C$  is as follows. Assume the five points  $p_1, \dots, p_5$  are  $(1 : 0 : 0 : 0 : 0), \dots, (0 : 0 : 0 : 0 : 1)$ ; if  $p_1, \dots, p_7$  are in general position then  $p_6$  and  $p_7$  have nonzero coordinates. Consider the birational Cremona transformation  $\tau: (x_0 : \dots : x_4) \mapsto (x_0^{-1} : \dots : x_4^{-1})$ . Then  $C$  is the preimage under  $\tau$  of the line  $\overline{\tau(p_6)\tau(p_7)}$ .

It is well-known that the secant variety of such a determinantal variety is given by

$$\sigma(C) = \left\{ (x_0 : \dots : x_4) \in \mathbb{P}^4 : \det \begin{pmatrix} x_0 & x_1 & x_2 \\ x_1 & x_2 & x_3 \\ x_2 & x_3 & x_4 \end{pmatrix} = 0 \right\}.$$

This is a cubic hypersurface with singular locus  $C$ ; in particular,  $\sigma(C)$  is singular at  $p_1, \dots, p_7$ . See [BO08, §3] for more details.

Now it is possible to vary  $p_7$  along  $C$  such that the construction of  $\sigma(C)$  remains unchanged. Consider the PNN

$$W_2 \circ \rho_3 \circ W_1 \mathbf{x}$$

where  $\mathbf{x} \in \mathbb{R}^5$ ,  $W_1 \in \mathbb{R}^{7 \times 5}$ ,  $W_2 \in \mathbb{R}^{1 \times 7}$ . Assume similar to above that the curve  $\tau(C)$  is given by the  $2 \times 2$  minors of

$$\begin{pmatrix} w_{161}^{-1} & \dots & w_{165}^{-1} \\ w_{171}^{-1} & \dots & w_{175}^{-1} \end{pmatrix}. \quad (4.2.7)$$

We are looking for a symmetric matrix  $G \in \text{GL}(7, \mathbb{R})$  such that  $GW_1$  gives rise to the same curve  $\tau(C)$ . Again, w.l.o.g., we assume that the first five rows and columns of  $G$  are the identity matrix. Thus, restricting to the last two rows and columns of  $G$ , we require that

$$\begin{pmatrix} g_1 & g_2 \\ g_2 & g_3 \end{pmatrix} \begin{pmatrix} w_{161}^{-1} & \dots & w_{165}^{-1} \\ w_{171}^{-1} & \dots & w_{175}^{-1} \end{pmatrix}$$

have the same  $2 \times 2$  minors as (4.2.7). Normalising  $G$  to have determinant one we can solve this to obtain

$$G = \begin{pmatrix} \text{Id}_5 & 0 & 0 \\ 0 & g & \phi \\ 0 & \phi & \frac{1+\phi^2}{g} \end{pmatrix}$$

where  $\phi = (\frac{1}{2}(-1 + \sqrt{3}))^{1/2}$  and  $g \in \mathbb{R}$  is a free parameter. Thus, we have obtained a one-dimensional family of symmetries on the parameter space leaving the image of the parameter map  $\Psi_{d,r}$  unchanged.

## 4.3. Optimisation landscape

When training a neural network one needs to solve an empirical minimisation problem, as was explained in Section 2.5, and in particular (2.5.2). It turns out that, analogously to maximum likelihood estimation (see Section 2.2) or Euclidean distance minimisation, there exists a degree of this optimisation problem constituting a measure for the algebraic complexity of solving this problem. In Subsection 4.3.1, we study the (static) optimisation properties of a PNN by introducing the *learning degree* of the corresponding neurovariety and relate it to the generic Euclidean distance degree. This provides an upper bound on the number of nonsingular functions that can be learned by a network after a training process. We compute this number for a specific family in Theorem 4.3.5 and relate this theoretical result and its practical implications to a machine learning experiment in Subsection 4.3.2. In the expository Subsection 4.3.3, we review the trajectory-based convergence results of gradient descent on linear neural networks and phrase open questions for polynomial neural networks with activation degree  $r \geq 2$ .

### 4.3.1. The learning degree of a PNN

Often one tries to solve the empirical risk optimisation problem (2.5.1) using gradient based methods, see Section 2.5. Therefore, one is interested in computing the critical points on the neurovariety of the loss function  $\ell_{p_w}$  as defined in (2.5.2) for a PNN  $p_w$ . In particular, the number of such points provides an estimate on the number of functions the optimisation process can converge to, see the experiment in Subsection 4.3.2.

**Definition 4.3.1.** If there is a finite number of critical points of  $\ell_{p_w}$  on the regular locus of  $\mathcal{V}_{d,r}$  which is constant for a general choice of training data  $(\mathbf{x}_i, \mathbf{y}_i)_i$ , we call this number the *learning degree* of the network  $p_w$  with respect to loss  $\ell$ , and denote it by  $\text{Ldeg}_\ell(p_w)$ .

In the following we study the case of the Euclidean loss function. We show that in this case the learning degree is independent of the choice of training data  $(\mathbf{x}_i, \mathbf{y}_i)_i$ . An upper bound for the learning degree is provided by the generic Euclidean distance (ED) degree.

**Definition 4.3.2** ([BKS24, Definition 2.8]). Let  $X$  be a variety, let  $\Lambda = (\lambda_1, \dots, \lambda_n) \in \mathbb{R}_+^n$  and let  $\mathbf{u} \in \mathbb{R}^n$  be a general point. The  $\Lambda$ -weighted Euclidean distance degree of  $X$  is the number of (complex) critical points on  $X$  of the function

$$\|\mathbf{x} - \mathbf{u}\|_\Lambda^2 := \sum_{i=1}^n \lambda_i (x_i - u_i)^2.$$

For general weights  $\Lambda$ , this number is independent of  $\Lambda$ , and we call it the *generic Euclidean distance (ED) degree* of  $X$ , which we denote as  $\text{EDdeg}_{\text{gen}}(X)$ .

Let  $(p_w^{(i)}(\mathbf{x}))_i \in (\text{Sym}_d(\mathbb{R}^n))^m$  be a tuple of polynomials in  $n$  variables  $\mathbf{x} = (x_1, \dots, x_n)$  occurring as the output of a fixed PNN with weights  $w$ . Suppose we want the PNN to learn the tuple of polynomials  $(f_i(\mathbf{x}))_i \in (\text{Sym}_d(\mathbb{R}^n))^m$ . To this end, we sample  $N$  input vectors  $\hat{\mathbf{x}}_1, \dots, \hat{\mathbf{x}}_N \in \mathbb{R}^n$  according to some distribution. Evaluating  $(f_i(\mathbf{x}))_i$  at these points yields  $\hat{\mathbf{y}}_j := (f_i(\hat{\mathbf{x}}_j))_i \in \mathbb{R}^m$  for  $j = 1, \dots, N$ ; the pairs  $(\hat{\mathbf{x}}_j, \hat{\mathbf{y}}_j)_j$  constitute the training data.

The learning process is defined by minimising the average Euclidean distance between  $\hat{\mathbf{y}}_j$  and the points  $(p_w^{(i)}(\hat{\mathbf{x}}_j))_i$  obtained from evaluating the network at the samples  $\hat{\mathbf{x}}_j$ . This corresponds to empirical risk minimisation with  $\ell_2$  loss. Concretely, we need to

$$\text{minimise } \frac{1}{N} \sum_{j=1}^N \| (p_w^{(i)}(\hat{\mathbf{x}}_j))_i - \hat{\mathbf{y}}_j \|^2 \quad \text{over weights } w. \quad (4.3.1)$$

Note that  $(p_{\mathbf{w}}^{(i)}(\hat{\mathbf{x}}_j))_i, \hat{\mathbf{y}}_j \in \mathbb{R}^m$  for all  $j = 1, \dots, N$ , and we have

$$\frac{1}{N} \sum_{j=1}^N \|(p_{\mathbf{w}}^{(i)}(\hat{\mathbf{x}}_j))_i - \hat{\mathbf{y}}_j\|^2 = \frac{1}{N} \sum_{i=1}^m \sum_{j=1}^N \|p_{\mathbf{w}}^{(i)}(\hat{\mathbf{x}}_j) - (\hat{\mathbf{y}}_j)_i\|^2.$$

Our goal is to show that for each  $i = 1, 2, \dots, m$ , this is a quadratic form in the coefficients of  $p_{\mathbf{w}}^{(i)}$  and  $f_i$  over training data. Let us write

$$p_{\mathbf{w}}^{(i)}(\mathbf{x}) = \sum_{I \in \binom{[n]}{d}} \rho_I^i \mathbf{x}^I \quad \text{and} \quad f_i(\mathbf{x}) = \sum_{J \in \binom{[n]}{d}} \phi_J^i \mathbf{x}^J.$$

We define a matrix  $E$  as follows:  $E$  is a block-diagonal matrix with blocks  $E^i$  for  $i = 1, 2, \dots, m$ , and each block  $E^i$  has as row and column labels the multiindices  $\alpha, \beta \in \binom{[n]}{d}$ . Then the entry  $E_{\alpha, \beta}^i$  is defined as

$$E_{\alpha, \beta}^i := \frac{1}{N} \sum_{j=1}^N \hat{\mathbf{x}}_j^{\alpha+\beta}.$$

This allows us to re-express the empirical risk as

$$\frac{1}{N} \sum_{j=1}^N \|p_{\mathbf{w}}^{(i)}(\hat{\mathbf{x}}_j) - f_i(\hat{\mathbf{x}}_j)\|^2 = \sum_{\alpha, \beta} (\rho_{\alpha}^i - \phi_{\alpha}^i) E_{\alpha, \beta}^i (\rho_{\beta}^i - \phi_{\beta}^i).$$

Therefore, the loss function (4.3.1) is an  $E_{\alpha, \beta}^i$ -weighted Euclidean distance in the space  $(\text{Sym}_d(\mathbb{R}^n))^m$ , the ambient space of the neurovariety  $\mathcal{V}_{\mathbf{d}, r}$ . For a general choice of training data  $(\hat{\mathbf{x}}_i, \hat{\mathbf{y}}_i)_i$ , each block  $E^i$  of the linear transformation is generic. Note, however, that all blocks  $E^i$ , for  $i = 1, \dots, m$ , will be equal, hence, the transformation  $E$  might be non-generic. We summarise the discussion in the theorem below.

**Theorem 4.3.3.** *The learning degree  $\text{Ldeg}_{\ell_2}(p_{\mathbf{w}})$  of a PNN  $p_{\mathbf{w}}$  with architecture  $(\mathbf{d}, r)$  with respect to the Euclidean loss function exists. It is at most the generic Euclidean distance degree of its neurovariety, i.e.*

$$\text{Ldeg}_{\ell_2}(p_{\mathbf{w}}) \leq \text{EDdeg}_{\text{gen}}(\mathcal{V}_{\mathbf{d}, r}). \quad (4.3.2)$$

If  $p_{\mathbf{w}}$  has only a single output neuron, one has equality in (4.3.2).

For large samples, the quadratic form defined by  $E$  is independent of the samples.

**Remark 4.3.4.** Suppose  $\mathbf{x}_1, \dots, \mathbf{x}_N$  are drawn independently from a fixed distribution  $\mathcal{D}$  with known moments  $\mu_k$ ,  $k = 1, 2, \dots$  By the law of large numbers, for any  $\epsilon > 0$ ,

$$\mathbb{P}(|E_{\alpha, \beta}^i - \mu_{\alpha+\beta}| > \epsilon) \rightarrow 0 \text{ as } N \rightarrow \infty.$$

Then for any  $\epsilon > 0$ ,

$$\mathbb{P} \left( \left| \frac{1}{N} \sum_{j=1}^N \|p_{\mathbf{w}}^{(i)}(\hat{\mathbf{x}}_j) - f_i(\hat{\mathbf{x}}_j)\|^2 - (\boldsymbol{\rho}^i - \boldsymbol{\phi}^i)^T A (\boldsymbol{\rho}^i - \boldsymbol{\phi}^i) \right| > \epsilon \right) \rightarrow 0 \text{ as } N \rightarrow \infty,$$

where  $A$  is a fixed  $N \times N$  matrix whose entries are determined by the distribution  $\mathcal{D}$ .

For the architecture  $\mathbf{d} = (2, 2, k)$ ,  $r = 2$ , we compute the bound in (4.3.2) explicitly.

**Theorem 4.3.5.** For  $k \geq 2$ , the generic ED degree of  $\mathcal{V}_{(2,2,k),2}$  is  $8k^2 - 12k + 3$ . Hence, the learning degree of the PNN with architecture  $(\mathbf{d} = (2, 2, k), r = 2)$  is at most  $8k^2 - 12k + 3$ .

The generic ED degree of a variety can be computed as the sum of its polar degrees [BKS24, Cor. 2.14]. For a smooth variety  $X$ , this can be reformulated in terms of degrees of Chern classes of  $X$ , see [BKS24, Thm. 4.20]. However,  $V := \mathcal{V}_{(2,2,k),2}$  is not smooth: recall that  $V$  consists of  $k \times 3$  matrices with rank  $\leq 2$ ; this determinantal variety has a non-empty singular locus consisting of all matrices with rank  $\leq 1$ . Therefore, we have to make use of the machinery of *Chern–Mather classes*. The reader wishing to learn more about Chern classes in the context of optimisation is referred to [BKS24, §4.3]. We give a brief definition of Chern–Mather classes below, more details can be found for example in [Yok86].

Let  $X \hookrightarrow \mathbb{P}^N$  be a projective variety of pure dimension  $d$ . Let  $X_{\text{reg}}$  denote the open subvariety of  $X$  of nonsingular points and consider the embedding

$$g: X_{\text{reg}} \hookrightarrow \text{Gr}(d, T\mathbb{P}^N), \quad x \mapsto T_x X_{\text{reg}}.$$

Then  $\hat{X} := \overline{g(X_{\text{reg}})}$  is called the *Nash blow-up* of  $X$ . The restriction of the projection  $\text{Gr}(d, T\mathbb{P}^N) \rightarrow \mathbb{P}^N$  to  $\hat{X}$  gives the *Nash blow-up map*  $v: \hat{X} \rightarrow X$ . Restricting the tautological subbundle of  $\text{Gr}(d, T\mathbb{P}^N)$  to  $\hat{X}$  gives the *Nash tangent bundle*  $\widehat{T\mathbb{X}}$  of  $\hat{X}$ .

**Definition 4.3.6.** The  $i^{\text{th}}$  Chern–Mather class  $c_i^M(X)$  of  $X$  is the pushforward

$$c_i^M(X) := v_*(c_i(\widehat{T\mathbb{X}} \cap [\hat{X}])).$$

The total Chern–Mather class  $c^M(X)$  is the sum of all  $c_i^M(X)$ .

The following result by Zhang expresses the generic ED degree of a determinantal variety in terms of degrees of Chern–Mather classes.

**Lemma 4.3.7** ([Zha18, Prop. 5.5]). Let  $V_{m,n,r} \hookrightarrow \mathbb{P}^{mn-1}$  be the determinantal variety of  $m \times n$  matrices with rank  $\leq r$ . Then

$$\text{EDdeg}_{\text{gen}}(V_{m,n,r}) = \sum_{l=0}^{(m+k)(n-k)-1} \sum_{i=0}^l (-1)^i \binom{(m+k)(n-k)-i}{(m+k)(n-k)-l} \beta_{(m+k)(n-k)-1-i},$$

where  $k = m - r$  and  $c^M(V_{m,n,r}) = \sum_{l=0}^{mn-1} \beta_l H^l \in \mathbb{Z}[H]/\langle H^{mn} \rangle = A(\mathbb{P}^{mn-1})$  with  $H = c_1(\mathcal{O}_{\mathbb{P}^{mn-1}}(1))$  being the generator of the Chow ring of projective space  $A(\mathbb{P}^{mn-1})$ .

The Chern–Mather class of a determinantal variety has a quite involved, yet very explicit description due to Zhang.

**Theorem 4.3.8** ([Zha18, Thm. 4.3]). With the notation as in Lemma 4.3.7,

$$c^M(V_{m,n,r}) = \text{Tr}(A(m, n, r) \cdot \mathcal{H}(m, n, r) \cdot B(m, n, r)).$$

Here,  $A, B$  and  $\mathcal{H}$  are the following  $(m(n-k)+1) \times (m(n-k)+1)$  matrices ( $k = m - r$ ):

$$\begin{aligned} A(m, n, r)_{i,j} &= \int_{\text{Gr}(k, n)} c(\mathcal{T}_{\text{Gr}(k, n)}) c_i(\mathcal{Q}^{\vee m}) c_{j-i}(\mathcal{S}^{\vee m}) \cap [\text{Gr}(k, n)], \\ B(m, n, r)_{i,j} &= \binom{m(n-k)-j}{i-j}, \\ \mathcal{H}(m, n, r)_{i,j} &= H^{mk+j-i}, \end{aligned}$$

where  $\mathcal{S}$  and  $\mathcal{Q}$  are the universal sub- and quotient bundle of the Grassmannian, respectively, and  $i, j = 0, 1, \dots, m(n-k)$ .

*Proof of Theorem 4.3.5.* We first need to compute the Chern–Mather class  $c^M(V_{k,3,2})$ . Note that in this case  $k = 1$ , so  $\mathrm{Gr}(k, n) = \mathrm{Gr}(1, 3) \cong \mathbb{P}^2$ . Let  $A(\mathbb{P}^2) = \mathbb{Z}[h]/\langle h^3 \rangle$  be the Chow ring of  $\mathbb{P}^2$  where  $h = c_1(\mathcal{O}_{\mathbb{P}^2}(1))$ . The universal subbundle  $\mathcal{S}$  is  $\mathcal{O}_{\mathbb{P}^2}(-1)$ , so we have

$$c(\mathcal{T}_{\mathbb{P}^2}) = 1 + 3h + 3h^2, \quad c(\mathcal{S}^{\vee k}) = 1 + kh + \frac{1}{2}k(k-1)h^2, \quad c(\mathcal{Q}^{\vee k}) = 1 - kh + \frac{1}{2}k(k+1)h^2.$$

Using these, we compute, for example,

$$A(k, 3, 2)_{1,2} = \int_{\mathbb{P}^2} (1 + 3h + 3h^2)(-kh)kh \cap [\mathbb{P}^2] = -k^2.$$

Similarly, we obtain that  $A(k, 3, 2)$  is the  $(2k+1) \times (2k+1)$  matrix

$$A(k, 3, 2) = \begin{pmatrix} 3 & 3k & \frac{1}{2}k(k-1) & 0 & \dots & 0 \\ 0 & -3k & -k^2 & 0 & \dots & 0 \\ 0 & 0 & \frac{1}{2}k(k+1) & 0 & \dots & 0 \\ 0 & 0 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \dots & 0 \end{pmatrix}.$$

The matrix  $B(k, 3, 2)$  is given by

$$B(k, 3, 2) = \begin{pmatrix} \binom{2k}{0} & 0 & 0 & \dots & 0 \\ \binom{2k}{1} & \binom{2k-1}{0} & 0 & \dots & 0 \\ \binom{2k}{2} & \binom{2k}{1} & \binom{2k}{0} & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \binom{2k}{2k} & \binom{2k-1}{2k-1} & \binom{2k-2}{2k-2} & \dots & \binom{0}{0} \end{pmatrix}$$

and the matrix  $\mathcal{H}(k, 3, 2)$  is

$$\mathcal{H}(k, 3, 2) = \begin{pmatrix} H^k & H^{k+1} & H^{k+2} & \dots & 0 \\ H^{k-1} & H^k & H^{k+1} & \dots & H^{3k-1} \\ H^{k-1} & H^{k-1} & H^k & \dots & H^{3k-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & H^k \end{pmatrix},$$

where  $H = c_1(\mathcal{O}_{\mathbb{P}^{3k-1}}(1))$  is the generator of  $A(\mathbb{P}^{3k-1})$ . Carrying out the multiplication  $B(k, 3, 2)A(k, 3, 2)\mathcal{H}(k, 3, 2)$ , one finds that the diagonal has entries

$$\begin{pmatrix} \binom{2k}{0} (3H^k + 3kH^{k-1} + \frac{1}{2}k(k-1)H^{k-2}) \\ \binom{2k}{1} (3H^{k+1} + 3kH^k + \frac{1}{2}k(k-1)H^{k-1}) + \binom{2k-1}{0} (-3kH^k - k^2H^{k-1}) \\ \binom{2k}{2} (3H^{k+2} + 3kH^{k+1} + \frac{1}{2}k(k-1)H^k) + \binom{2k-1}{1} (-3kH^{k+1} - k^2H^k) + \binom{2k}{0} \frac{1}{2}k(k+1)H^{k-1} \\ \vdots \\ \binom{2k}{2k} \cdot 0 + \binom{2k-1}{2k-1} (-3kH^{3k-1} - k^2H^{3k-2}) + \binom{2k-2}{2k-2} \frac{1}{2}k(k+1)H^{3k-3} \end{pmatrix}.$$

Computing the trace we find that

$$\begin{aligned} c^M(V_{k,3,2}) &= \sum_{j=-2}^{2k-1} \left[ 3\binom{2k}{j} + 3k \left( \binom{2k}{j+1} - \binom{2k-1}{j} \right) + \frac{1}{2}k(k-1)\binom{2k}{j+2} \right. \\ &\quad \left. + \frac{1}{2}k(k+1)\binom{2k-2}{j} - k^2\binom{2k-1}{j+1} \right] \end{aligned}$$

and hence, by Lemma 4.3.7,

$$\text{EDdeg}_{\text{gen}}(\mathcal{V}_{(2,2,k),2}) = \sum_{l=0}^{2(k+1)-1} \sum_{i=0}^l (-1)^i \binom{2(k+1)-i}{2(k+1)-l} \left[ 3 \binom{2k}{i-2} + 3k \binom{2k}{i-1} - \binom{2k-1}{i-2} \right] + \frac{1}{2} k(k-1) \binom{2k}{i} + \frac{1}{2} k(k+1) \binom{2k-2}{i-2} - k^2 \binom{2k-1}{i-1}.$$

Rearranging the double sum to

$$\sum_{i=0}^{2(k+1)-1} \left( \sum_{l=i}^{2(k+1)-1} \binom{2(k+1)-i}{2(k+1)-l} \right) (-1)^i \left[ 3 \binom{2k}{i-2} + 3k \binom{2k}{i-1} - \binom{2k-1}{i-2} \right] + \frac{1}{2} k(k-1) \binom{2k}{i} + \frac{1}{2} k(k+1) \binom{2k-2}{i-2} - k^2 \binom{2k-1}{i-1}$$

and noting that

$$\sum_{l=i}^{2(k+1)-1} \binom{2(k+1)-i}{2(k+1)-l} = 2^{2(k+1)-i} - 1,$$

one verifies that the above expression indeed equals  $8k^2 - 12k + 3$ .  $\square$

**Remark 4.3.9.** For the architecture  $(\mathbf{d} = (2, 2, 3), r = 2)$ , Theorem 4.3.8 yields the bound  $\text{Ldeg}_{\ell_2}(p_{\mathbf{w}}) \leq 39$ . A numerical computation for a random choice of block  $E^i$  in the linear transformation  $E$  reveals that the actual learning degree is  $\text{Ldeg}_{\ell_2}(p_{\mathbf{w}}) \leq 3$ , which is also the learning degree for the network with architecture  $(\mathbf{d} = (2, 2, 1), r = 2)$ . Moreover, we observe that all critical points of the loss function are in fact real. This interesting phenomenon should be further studied in future work.

### 4.3.2. Machine learning experiment

In this subsection we connect our theoretical study of the learning degree with a machine learning experiment that is close to how neural networks are used in practice. This means that we (approximately) solve the optimisation problem (2.5.1) in the training process via stochastic gradient descent, as is explained in Section 2.5. The update rule for gradient descent (2.5.3) reads as follows in the case of PNNs. For a PNN  $p_{\mathbf{w}}$  with weights  $\mathbf{w} = (W_1, W_2, \dots, W_L)$ , at time step  $t$ , each  $W_i$  is updated by

$$W_i^{(t+1)} = W_i^{(t)} - \eta \nabla_{W_i} \ell_{p_{\mathbf{w}^{(t)}}} \quad (4.3.3)$$

where  $\nabla_{W_i} \ell_{p_{\mathbf{w}^{(t)}}}$  is computed efficiently by backpropagation (see Algorithm 1).

We generate synthetic data consisting of  $m = 5000$  datasets, each containing  $N = 50$  data points. For each dataset, input pairs  $\hat{\mathbf{x}} = (\hat{x}_1, \hat{x}_2)^\top$  are sampled uniformly from the range  $[-1, 1]$ . The corresponding outputs  $\hat{\mathbf{y}} = (\hat{y}_1, \hat{y}_2, \hat{y}_3)^\top$  are generated using quadratic functions with randomly sampled coefficients

$$\begin{aligned} \hat{y}_1 &= c_{1,1} \hat{x}_1^2 + c_{1,2} \hat{x}_1 \hat{x}_2 + c_{1,3} \hat{x}_2^2, \\ \hat{y}_2 &= c_{2,1} \hat{x}_1^2 + c_{2,2} \hat{x}_1 \hat{x}_2 + c_{2,3} \hat{x}_2^2, \\ \hat{y}_3 &= c_{3,1} \hat{x}_1^2 + c_{3,2} \hat{x}_1 \hat{x}_2 + c_{3,3} \hat{x}_2^2, \end{aligned} \quad (4.3.4)$$

where the coefficients  $c_{i,j}$  were sampled from a normal distribution  $\mathcal{N}(0, 1)$ .

We employ a polynomial neural network  $p_{\mathbf{w}}$  with architecture  $\mathbf{d} = (2, 2, 3)$  and quadratic activation function ( $r = 2$ ), hence satisfying the assumptions in Theorem 4.3.5. The parameters  $\mathbf{w}$  consist of two matrices  $\mathbf{w} = (W_1, W_2)$  with  $W_1 \in \mathbb{R}^{2 \times 2}$  and  $W_2 \in \mathbb{R}^{3 \times 2}$ .

Each network is trained on its corresponding dataset using stochastic gradient descent (SGD) with an initial learning rate of  $\eta = 0.1$ . The mean squared error loss function

$$L(W_1, W_2) = \frac{1}{N} \sum_{i=1}^N \|p_{\mathbf{w}}(\hat{\mathbf{x}}^{(i)}) - \hat{\mathbf{y}}^{(i)}\|_2^2$$

is minimised. A learning rate scheduler reduces the learning rate by half every 1000 epochs. Training proceeds for a maximum of 15000 epochs or until the maximum gradient magnitude falls below a threshold of  $1 \times 10^{-4}$ , indicating convergence. We refer to this hyperparameter as the gradient norm threshold.

After training, we extract the quadratic coefficients learned by each network to represent the functions it had approximated. The extraction is performed by expressing the network's output as a quadratic function of the inputs. For each output unit  $j$ , the coefficients  $(a_{1j}, a_{2j}, a_{3j})$  are computed as

$$\begin{aligned} a_{1j} &= v_{j1} w_{11}^2 + v_{j2} w_{21}^2, \\ a_{2j} &= 2(v_{j1} w_{11} w_{12} + v_{j2} w_{21} w_{22}), \\ a_{3j} &= v_{j1} w_{12}^2 + v_{j2} w_{22}^2, \end{aligned} \tag{4.3.5}$$

where  $v_{jk}$  are entries of  $W_2$  and  $w_{kl}$  are entries of  $W_1$ .

To identify distinct functions learned by the network, we compare the extracted coefficients using a coefficient comparison method with a specified tolerance  $\epsilon$ . Two functions given by coefficients  $a_{ij}^{(1)}$  and  $a_{ij}^{(2)}$  as in (4.3.5) are considered the same if

$$|a_{ij}^{(1)} - a_{ij}^{(2)}| < \epsilon \quad \forall i, j.$$

We systematically adjust the tolerance  $\epsilon$  to explore its impact on the number of distinct functions identified. By varying  $\epsilon$ , we can control the granularity of function distinction, allowing minor variations in coefficients to be considered functionally equivalent.

In addition, we perform perturbations on the function's coefficients to assess whether each learned function corresponds to a local minimum in the neuromanifold. Small perturbations  $\delta$  are added to the coefficients

$$\tilde{c}_{ij} = c_{ij} + \delta_{ij}, \quad \delta_{ij} \in [-\epsilon, \epsilon],$$

where  $\epsilon$  is a small value (e.g.  $1 \times 10^{-4}$ ). The perturbed coefficients also give rise to a perturbed neural network which we denote by  $\tilde{p}_{\mathbf{w}}$ . The perturbed loss then becomes

$$L_\delta(W_1, W_2) = \frac{1}{N} \sum_{i=1}^N \left\| \tilde{p}_{\mathbf{w}}(\hat{\mathbf{x}}^{(i)}) - \tilde{\mathbf{y}}^{(i)} \right\|_2^2,$$

where  $\tilde{\mathbf{y}}^{(i)}$  are defined as in (4.3.4) but using the perturbed coefficients  $\tilde{c}_{ij}$ . If the value of the original loss function is less than or equal to the losses of all perturbed functions, we consider it to be a local minimum in the neuromanifold.

Through our experimental setup, we are able to

- train the network  $p_{\mathbf{w}}$  with 5000 datasets via gradient descent over 15000 epochs and gradient norm threshold  $10^{-4}$ ;

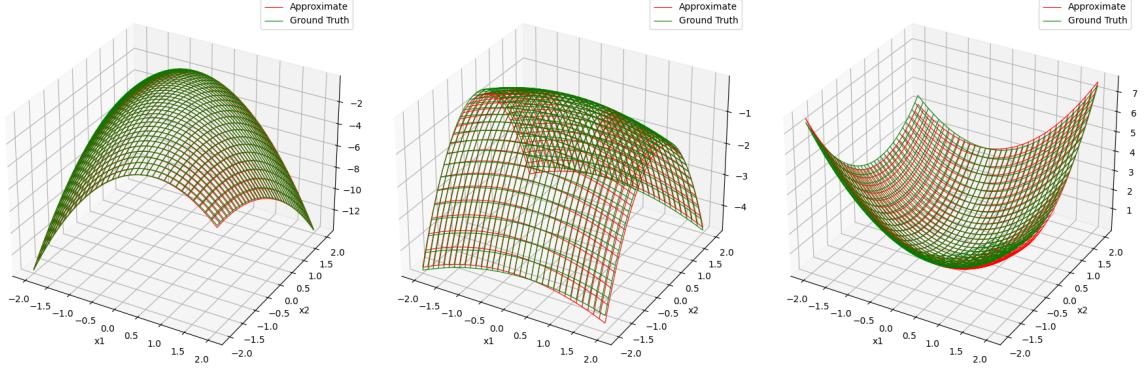


Figure 4.2: Coordinatewise comparison of the function most frequently learned by a PNN with  $\mathbf{d} = (2, 2, 3)$  and  $r = 2$  (red graph) with the ground truth (green graph).

- identify 17 distinct functions whose coefficients differ at least by tolerance threshold  $\epsilon = 1/10$  and that were learned with frequency no less than 10 out of 5000 datasets;
- check that for 16 out of 17 functions, the coefficient matrices have rank one;
- verify the local minimality of the one learned function whose coefficient matrix has rank two in the neuromanifold by evaluating the loss landscape around the function.

The functions with rank-one coefficient matrices are singular points of the neuromanifold. At those points the gradient descent algorithm “gets stuck” and does not converge to a local minimum (as verified by the fourth bullet point above). Therefore, we should only consider the single learned function that is a regular point of the neuromanifold. It is also the most frequently learned function and is depicted in Figure 4.2. The fact that the learning degree is three (see Remark 4.3.9) gives us *a priori* knowledge that most of the 17 learned functions in the training must be singular points and hence are not necessarily local minima. This experiment illustrates how the learning degree informs the training process. The experiment implementation is available at [KLW24b].

### 4.3.3. Case study: linear neural networks

In addition to static properties of neural network optimisation, we are also interested in the trajectory of practical optimisation algorithms. This problem presents a more complex challenge, particularly in its general form. Limited results are known for the special subfamily of polynomial neural networks with activation degree  $r = 1$ . In this subsection, we survey prior results on the trajectory of the gradient descent algorithms applied to linear neural networks from literature in the algebraic statistics and machine learning communities separately. An interesting open question is to extend the analysis to polynomial neural networks with general activation degrees.

Linear neural networks are a special class of polynomial neural networks with the activation function  $\rho_1(\mathbf{x}) = \mathbf{x}$ . The associated map from the set of parameters to vectors of homogeneous polynomials of degree one in  $d_0$  many variables is given by

$$\Psi_{\mathbf{d},1} : \mathbb{R}^{d_0 \times d_1} \times \cdots \times \mathbb{R}^{d_{L-1} \times d_L} \rightarrow (\text{Sym}_1(\mathbb{R}^{d_0}))^{d_L} \cong \mathbb{R}^{d_0 \times d_L}, \quad \mathbf{w} \mapsto p_{\mathbf{w}} = \left( p_{\mathbf{w}}^{(1)}, \dots, p_{\mathbf{w}}^{(d_L)} \right)^T.$$

The neuromanifold for the architecture  $\mathbf{d}$  is

$$\mathcal{M}_{\mathbf{d},1} = \{M \in \mathbb{R}^{d_0 \times d_L} : \text{rk}(M) \leq \min\{d_0, d_1, \dots, d_L\}\}.$$

In the linear case the neuromanifold is always equal to the neurovariety. By  $\mu_{\mathbf{d}}: \mathbb{R}^{d_1 \times d_0} \times \dots \times \mathbb{R}^{d_L \times d_{L-1}} \rightarrow \mathbb{R}^{d_0 \times d_L}$  we denote the multiplication  $\mu_{\mathbf{d}}(W_1, W_2, \dots, W_L) = W_L \cdots W_2 W_1$ .

**Proposition 4.3.10** ([TKB20, Thm. 4]). *Let  $\tilde{d} = \min(\mathbf{d})$ ,  $\mathbf{w} = (W_1, \dots, W_L)$  and  $\bar{W} = \mu_{\mathbf{d}}(\mathbf{w})$ .*

- (*Filling case*) *If  $\tilde{d} = \min\{d_0, d_L\}$ , the differential  $d\mu_d(\mathbf{w})$  has maximal rank equal to  $\dim(\mathcal{M}_{\mathbf{d},1}) = d_0 d_L$  if and only if for any  $i \in [L-1]$ , either  $\text{rk}(W_j) = d_L$  for all  $j > i$  or  $\text{rk}(W_j) = d_0$  for all  $j < i+1$ .*
- (*Non-filling case*) *If  $\tilde{d} < \min\{d_0, d_L\}$ , the differential  $d\mu_d(\mathbf{w})$  has maximal rank equal to  $\dim(\mathcal{M}_{\mathbf{d},1}) = \tilde{d}(d_0 + d_L - \tilde{d})$  if and only if  $\text{rk}(\bar{W}) = \tilde{d}$ .*

Besides static properties of the loss landscape, in the realm of neural networks, optimisation often focuses on devising and refining algorithms to efficiently find the best model parameters that minimise a loss function. Analysing the convergence trajectory is one of the most challenging problems in machine learning theory. We recall results on the convergence trajectory of deep linear neural networks with respect to  $\ell_2$  loss in Theorem 4.3.11.

**Theorem 4.3.11** ([ACGH19, Thm. 1], informal). *Consider training deep linear neural networks with  $\ell_2$  loss using the gradient descent algorithm. Assume that at initialisation, the weight matrices are well-conditioned. Moreover, assume the existence of  $\delta > 0$  such that at any time step  $T$ ,*

$$\|W_{j+1}^T W_{j+1} - W_j W_j^T\|_F \leq \delta, \quad (4.3.6)$$

*for all  $j = 1, 2, \dots, L-1$ . Then for any  $\epsilon > 0$ , with a sufficiently small learning rate, there exists  $T_0$  such that the loss is no greater than  $\epsilon$  for any  $T \geq T_0$ .*

Theorem 4.3.11 shows that gradient descent converges to the global minimum at a linear rate, given mild assumptions. Similar results have been shown for deep linear neural networks with losses beyond the  $\ell_2$  loss. In [BPAM23], the authors consider deep linear neural networks trained with the Bures–Wasserstein distance, a loss function commonly used in generative models. They analyse not only the critical points of the loss landscape but also the convergence of both gradient flow and gradient descent for the Bures–Wasserstein loss. The convergence trajectory of polynomial neural networks with activation degree  $r \geq 2$  remains an open problem left for future work.

## 4.4. Conclusion

We have studied polynomial neural networks from an algebro-geometric perspective, aiming at a better understanding of the learning capabilities and dynamics of these networks. This framework provides an important step towards understanding networks with rational activation functions which can efficiently model common activations like ReLU or sigmoid, and testing our theoretical understandings of deep learning successes in practice.

We have characterised neuromanifolds and neurovarieties for some specific architectures, and studied their dimensions. The geometry and dimension of the neuromanifold relates directly to the expressive power of the neuromanifold. Beyond expressivity, we also studied the optimisation process of polynomial neural networks, and bounded the number of nonsingular critical points of a loss function using our new notion of learning degree. Besides our conjectures from Subsection 4.2.3 concerning the dimension of neurovarieties, there are further questions to be studied, for example: What are the learning degrees of more general architectures? How many of the critical points of the loss functions are real?



# Chapter 5

# Quantum physics

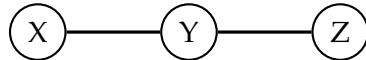
In this chapter we study connections between nonlinear algebra and quantum physics. These connections work in two ways: in Section 5.1 we apply techniques from nonlinear algebra to study the structure of quantum states satisfying certain conditions on their mutual information. To this end, we associate several algebraic varieties to quantum graphical models. These varieties are interesting in their own right and provide computational challenges; moreover, our new perspective provides insights into problems like finding the quantum information projection to a quantum exponential family, the quantum analogue of maximum likelihood estimation (see Subsection 5.1.7). When we restrict to quantum states that are represented by diagonal matrices, these quantum states can simply be considered as discrete random variables. In this case, the algebraic varieties we associate to the quantum graphical model boil down to the classical undirected graphical model from statistics as discussed in Section 2.2. Therefore, we can regard Section 5.1 as a noncommutative generalisation of the algebraic statistics of classical graphical models.

In Section 5.2 we take inspiration from quantum field theory. More precisely, we relate the asymptotic expansion of an exponential integral to the generating function of certain edge-coloured graphs. In the context of perturbative quantum field theory, this amounts to an expansion of a path integral in a zero-dimensional quantum field theory in terms of Feynman graphs. This integral representation then helps us find an expression for the asymptotics of the number of edge-bicoloured graphs with a fixed vertex incidence structure (see Theorem 5.2.17). Here, the connection to nonlinear algebra becomes apparent as the asymptotics is governed by certain critical points of a polynomial encoding the allowed vertex incidence structure. Varying a coupling parameter between the two colours can lead to phase transitions in the asymptotics resembling phenomena known from statistical physics. We connect our combinatorial graph enumeration problem to the partition function of a critical Ising model in Remark 5.2.12.

## 5.1. Quantum graphical models

The goal of this section is to consider *quantum graphical models* [LP08] from the point of view of algebraic geometry with the aim of offering a new perspective on open problems in quantum information theory. Roughly speaking, when passing from the classical to the quantum setting, we replace probability distributions with *density matrices*, with the classical case being recovered when these matrices are diagonal. The graph describes a physical quantum system with nodes representing subsystems. Such models have also been coined *quantum Markov networks* in the quantum information theory literature [BP12, DGMM20, PH11] and have applications to quantum many-body systems, quantum error correction and the study of entanglement. This section describes different approaches to obtain an algebraic variety associated to a quantum graphical model. See Section 2.6 for the necessary background on quantum information theory.

In Section 2.2 we already explained how *graphical models* play a prominent role in algebraic statistics. The key insight is that (undirected) graphical models give rise to nice (toric) varieties. For example [PS05, Ex. 1.29], consider the chain graph  $G$  on three vertices



with binary random variables  $X, Y$  and  $Z$ . Then  $G$  encodes the conditional independence statement  $X \perp\!\!\!\perp Z \mid Y$  giving rise to a statistical model described by the algebraic variety

$$\mathcal{M}_G = \mathcal{V}(p_{001}p_{100} - p_{000}p_{101}, p_{011}p_{110} - p_{010}p_{111}) \subseteq \mathbb{P}^7 = \text{Proj}(\mathbb{C}[p_{000}, \dots, p_{111}]).$$

This algebraic perspective advances both the theoretical foundations for these statistical models and the development of new computational methods for the practical use. At the core of these advances lies the understanding of the implicit and parametric descriptions of the model and its likelihood geometry. This proved to be useful in model selection, causal discovery and maximum likelihood estimation [Eva20, GMS06, LUSB14, URBY13].

Motivated by this, we find ways to associate algebraic varieties to quantum graphical models and make progress towards understanding their implicit and parametric descriptions. This leads to a number of interesting varieties and new computational challenges. The role of the maximum likelihood estimator is taken by the *quantum information projection*. We study its geometry for *quantum exponential families* of commuting Hamiltonians, e.g. Hamiltonians arising in the context of *graph states*.

A main motivation for us from quantum information theory is the problem of finding a description of the set of compatible density matrices on subsystems of a composite system. This is known as the *Quantum Marginal Problem* and is mostly open.

**Problem 5.1.1** (Quantum Marginal Problem). Let  $S = \{1, \dots, N\}$  be a composite system on  $N$  qudits and suppose we are given density matrices  $\rho_{S_1}, \dots, \rho_{S_n}$  of  $n$  subsystems  $S_1, \dots, S_n \subseteq S$ . What conditions do  $\rho_{S_1}, \dots, \rho_{S_n}$  have to satisfy to arise from  $\rho_S$  as

$$\rho_{S_i} = \text{Tr}_{S \setminus S_i} \rho_S ?$$

For general graphs, this problem has only been solved in the case of disjoint subsystems  $S_i$ . See [TV15] for a survey. However, for trees it is possible to reconstruct a quantum state from its two-body marginals [DGM21, DGMM20]. This can be done using algebraic methods and motivates the algebro-geometric notions of quantum graphical models we introduce in this section. Associating algebraic varieties to quantum graphical models and studying their defining equations might open a new way of attacking this problem.

The section is organised as follows. In Subsection 5.1.1, we introduce the *quantum conditional mutual information* (QCMI) *variety* and in Subsection 5.1.2 the *Petz variety*. The former is obtained from studying the structure of quantum states satisfying *strong subadditivity* with equality [HJPW04]. The latter is related to the Petz recovery map [Pet86] and the solution of the Quantum Marginal Problem for the 3-chain [TV15]. In both instances, the graph imposes quantum conditional independence statements, in direct analogy to the classical case. In Section 5.1.3, we suggest a notion of a quantum graphical model as the *Gibbs manifold* [PST23] of a certain family of Hamiltonians and consider the smallest variety that contains it (known as the *Gibbs variety*). Here, the graph encodes a locality structure imposed on the Hamiltonians [BP12]. This subsection also includes results on Gibbs varieties. Finally, we present results on quantum exponential families coming from *stabiliser codes* [NC02, §10]; one particular example are families of Hamiltonians associated to graph states [HEB04]. We study the quantum information projection [NGKG13] and relate it to maximum likelihood estimation, proving a generalisation of Birch's Theorem.

Throughout the sections we provide algorithms to compute the varieties appearing in our study and present computational examples. Those are implemented in Julia making use of the computer algebra package `Oscar.jl` [OSC24] and the numerical algebraic geometry tool `HomotopyContinuation.jl` [BT18]; the code is available at [DPW23b].

The ambient space of the algebraic varieties we consider is  $\mathbb{S}^n \cong \mathbb{C}^{n(n+1)/2}$ , each point being a complex symmetric matrix. To recover a specific quantum model, we intersect the variety with the PSD cone and the hyperplane of trace one matrices.

### 5.1.1. Quantum conditional mutual information varieties

In this subsection we give the definition of quantum conditional mutual information (QCMI) and collect some of its properties. The vanishing of QCMI should be thought of as a quantum analogue to conditional independence and gives rise to an algebraic variety that we call the *QCMI variety*.

The *von Neumann entropy*  $S(\rho)$  of a quantum state  $\rho$  is  $S(\rho) = -\text{Tr}(\rho \log \rho)$  and is a straightforward generalisation of the classical Shannon entropy; here, the logarithm has base two. Let  $\rho_{ABC}$  be a tripartite state; then the *quantum conditional mutual information* between  $A$  and  $C$  given  $B$  is defined as

$$I(A : C | B) := S(AB) + S(BC) - S(ABC) - S(B),$$

where  $S(ABC) = S(\rho_{ABC})$ ,  $S(AB) = S(\text{Tr}_C \rho_{ABC})$  etc. See [Wil13, §11.7] for more details. If one replaces the von Neumann entropy with the Shannon entropy in the above definition, one obtains the classical conditional mutual information  $I_{\text{cl}}(A : C | B)$  between random variables  $A$  and  $C$  given  $B$ . Its vanishing  $I_{\text{cl}}(A : C | B) = 0$  is well-known to be equivalent to the conditional independence  $A \perp\!\!\!\perp C | B$  and leads to two possible different factorisations of the joint probability distribution  $p(a, b, c)$  [HJPW04]. The vanishing of QCMI of a tripartite system behaves similarly, implying a more involved factorisation of the density matrix of the tripartite system (see Construction 5.1.4).

The following constitutes a quantum analogue to the conditional independence axioms for probability distributions, see [LP08, Thm. 4.5].

**Proposition 5.1.2.** *Let  $S$  be a composite quantum system with disjoint subsystems  $A, B, C, D \subseteq S$ . Then the following implications hold:*

1.  $I(A : C | B) = 0 \Rightarrow I(C : A | B) = 0$  (*Symmetry*),
2.  $I(A : CD | B) = 0 \Rightarrow I(A : C | B) = 0$  (*Decomposition*),

3.  $I(A : CD | B) = 0 \Rightarrow I(A : C | BD) = 0$  (*Weak Union*),
4.  $I(A : B | CD) = 0$  and  $I(A : D | C) = 0 \Rightarrow I(A : BD | C) = 0$  (*Contraction*).

The QCMI is closely related to the *strong subadditivity* (SSA) inequality [LR73]

$$S(ABC) + S(B) \leq S(AB) + S(BC).$$

The case of equality in SSA, i.e.  $I(A : C | B) = 0$ , has been intensively studied; the main result is the following statement from [HJPW04].

**Theorem 5.1.3.** *A quantum state  $\rho_{ABC}$  on  $\mathcal{H}_A \otimes \mathcal{H}_B \otimes \mathcal{H}_C$  satisfies SSA with equality if and only if there exists a decomposition of  $\mathcal{H}_B$  as*

$$\mathcal{H}_B = \bigoplus_j \mathcal{H}_{b_j^L} \otimes \mathcal{H}_{b_j^R}$$

such that  $\rho_{ABC}$  decomposes as

$$\rho_{ABC} = \bigoplus_j q_j \rho_{Ab_j^L} \otimes \rho_{b_j^R C},$$

where  $\{q_j\}_j$  is a discrete probability distribution and  $\rho_{Ab_j^L}$ ,  $\rho_{b_j^R C}$  are states on  $\mathcal{H}_A \otimes \mathcal{H}_{b_j^L}$  and  $\mathcal{H}_{b_j^R} \otimes \mathcal{H}_C$ , respectively.

**Construction 5.1.4** (QCMI variety of the 3-chain graph). The following reformulation of Theorem 5.1.3 plays a central role in the construction of the QCMI variety. Setting  $\Lambda_{AB} := \bigoplus_j q_j \rho_{Ab_j^L} \otimes \text{Id}_{b_j^R C}$  and  $\Lambda_{BC} := \bigoplus_j \text{Id}_{Ab_j^L} \otimes \rho_{b_j^R C}$ , we arrive at

$$I(A : C | B) = 0 \quad \text{if and only if} \quad \rho_{ABC} = \Lambda_{AB} \Lambda_{BC} \text{ with } [\Lambda_{AB}, \Lambda_{BC}] = 0, \quad (5.1.1)$$

where  $\Lambda_{AB}$ ,  $\Lambda_{BC}$  are symmetric matrices acting on  $\mathcal{H}_A \otimes \mathcal{H}_B \otimes \mathcal{H}_C$  and  $\Lambda_{AB}$  acts as identity on  $\mathcal{H}_C$ , and, likewise,  $\Lambda_{BC}$  acts as identity on  $\mathcal{H}_A$  [BP12]. Then the right-hand side of (5.1.1) gives rise to a parametrisation of a variety we denote by  $\mathcal{Q}_{I(A:C|B)}$ .

**Example 5.1.5** ( $\mathcal{Q}_{I(A:C|B)}$  in the qubit case). Let  $\mathcal{H}_A \cong \mathcal{H}_B \cong \mathcal{H}_C \cong \mathbb{C}^2$  and write  $\Lambda_{AB} = M \otimes \text{Id}_2$ ,  $\Lambda_{BC} = \text{Id}_2 \otimes N$  for  $M, N \in \mathbb{S}^4$ . In this case, the parametrisation of  $\mathcal{Q}_{I(A:C|B)}$  induced by the right-hand side of (5.1.1) sends

$$M = \begin{pmatrix} x_1 & x_2 & x_3 & x_4 \\ x_2 & x_5 & x_6 & x_7 \\ x_3 & x_6 & x_8 & x_9 \\ x_4 & x_7 & x_9 & x_{10} \end{pmatrix} \text{ and } N = \begin{pmatrix} y_1 & y_2 & y_3 & y_4 \\ y_2 & y_5 & y_6 & y_7 \\ y_3 & y_6 & y_8 & y_9 \\ y_4 & y_7 & y_9 & y_{10} \end{pmatrix}$$

to the matrix

$$\begin{pmatrix} x_1y_1 + x_2y_4 & x_1y_2 + x_2y_5 & x_1y_4 + x_2y_6 & x_1y_7 + x_2y_9 & x_4y_1 + x_7y_4 & x_4y_2 + x_7y_5 & x_4y_4 + x_7y_6 & x_4y_7 + x_7y_9 \\ x_1y_2 + x_2y_7 & x_1y_3 + x_2y_8 & x_1y_5 + x_2y_9 & x_1y_8 + x_2y_{10} & x_4y_2 + x_7y_7 & x_4y_3 + x_7y_8 & x_4y_5 + x_7y_9 & x_4y_8 + x_7y_{10} \\ x_2y_1 + x_3y_4 & x_2y_2 + x_3y_5 & x_2y_4 + x_3y_6 & x_2y_7 + x_3y_9 & x_5y_1 + x_8y_4 & x_5y_2 + x_8y_5 & x_5y_4 + x_8y_6 & x_5y_7 + x_8y_9 \\ x_2y_2 + x_3y_7 & x_2y_3 + x_3y_8 & x_2y_5 + x_3y_9 & x_2y_8 + x_3y_{10} & x_5y_2 + x_8y_7 & x_5y_3 + x_8y_8 & x_5y_5 + x_8y_9 & x_5y_8 + x_8y_{10} \\ x_4y_1 + x_5y_4 & x_4y_2 + x_5y_5 & x_4y_4 + x_5y_6 & x_4y_7 + x_5y_9 & x_6y_1 + x_9y_4 & x_6y_2 + x_9y_5 & x_6y_4 + x_9y_6 & x_6y_7 + x_9y_9 \\ x_4y_2 + x_5y_7 & x_4y_3 + x_5y_8 & x_4y_5 + x_5y_9 & x_4y_8 + x_5y_{10} & x_6y_2 + x_9y_7 & x_6y_3 + x_9y_8 & x_6y_5 + x_9y_9 & x_6y_8 + x_9y_{10} \\ x_7y_1 + x_8y_4 & x_7y_2 + x_8y_5 & x_7y_4 + x_8y_6 & x_7y_7 + x_8y_9 & x_9y_1 + x_{10}y_4 & x_9y_2 + x_{10}y_5 & x_9y_4 + x_{10}y_6 & x_9y_7 + x_{10}y_9 \\ x_7y_2 + x_8y_7 & x_7y_3 + x_8y_8 & x_7y_5 + x_8y_9 & x_7y_8 + x_8y_{10} & x_9y_2 + x_{10}y_7 & x_9y_3 + x_{10}y_8 & x_9y_5 + x_{10}y_9 & x_9y_8 + x_{10}y_{10} \end{pmatrix}.$$

This results in a twelve-dimensional variety inside  $\mathbb{S}^8$  cut out by 735 equations in degrees one to five; the degree of  $\mathcal{Q}_{I(A:C|B)}$  is 110. As all of these equations are homogeneous,

$\mathcal{Q}_{I(A:C|B)}$  can be considered as a subvariety of  $\mathbb{P}^{35} = \text{Proj}(\mathbb{C}[z_1, \dots, z_{36}])$ . Among these equations only two are linear:

$$z_{14} - z_{18} + z_{23} - z_{29} = 0, \quad z_{12} - z_{16} - z_{25} + z_{31} = 0,$$

and just one has degree five:

$$\begin{aligned} & -z_{13}z_{22}z_{29}z_{31}z_{33} - z_{13}z_{22}z_{31}^2z_{35} + z_{13}z_{24}z_{29}^2z_{33} + z_{13}z_{24}z_{29}z_{31}z_{35} + z_{22}^2z_{29}z_{31}z_{33} \\ & + z_{22}^2z_{31}^2z_{35} + z_{22}z_{24}z_{25}z_{29}z_{35} - z_{22}z_{24}z_{25}z_{31}z_{33} - z_{22}z_{24}z_{29}^2z_{33} - 2z_{22}z_{24}z_{29}z_{31}z_{35} \\ & + z_{22}z_{24}z_{31}^2z_{33} - z_{23}z_{24}^2z_{29}z_{35} + z_{23}z_{24}^2z_{31}z_{33} + z_{24}^2z_{29}^2z_{35} - z_{24}^2z_{29}z_{31}z_{33} = 0. \end{aligned}$$

Here, the variables  $z_1, \dots, z_{36}$  denote the entries of a symmetric  $8 \times 8$  matrix written in order starting from left to right and continuing from top to bottom. Note that if you set all non-diagonal entries in  $M$  and  $N$  to zero, this results in a monomial parametrisation of the classical graphical model of the 3-chain as presented in the introduction.  $\diamond$

**Proposition 5.1.6.** *The variety  $\mathcal{Q}_{I(A:C|B)}$  is irreducible.*

*Proof.* Under the composition of morphisms

$$\begin{aligned} \mathrm{U}(8) \times \mathbb{R}^4 \times \mathbb{R}^4 & \rightarrow \mathbb{S}^8 \times \mathbb{S}^8 \xrightarrow{\text{mult.}} \mathbb{S}^8 \\ (U, \lambda, \mu) & \mapsto \left( U \text{diag}(\lambda_1, \lambda_1, \dots, \lambda_4, \lambda_4) U^{-1}, U \text{diag}(\lambda_1, \dots, \lambda_4, \lambda_1, \dots, \lambda_4) U^{-1} \right) \\ (M, N) & \mapsto M \cdot N, \end{aligned}$$

where  $\lambda = (\lambda_1, \dots, \lambda_4)$  and  $\mu = (\mu_1, \dots, \mu_4)$ ,  $\mathcal{Q}_{I(A:C|B)}$  is the image of an irreducible variety and hence irreducible itself.  $\square$

In analogy to the classical theory of graphical models, we associate QCMI statements to separations in an undirected graph. For classical graphical models on undirected graphs, the Hammersley–Clifford Theorem (Theorem 2.2.7) states that a positive probability distribution satisfies the conditional independence statements associated to separations in a graph if and only if it factorises according to the graph. One might attempt to achieve a similar factorisation theorem for quantum graphical models, however, such description is not available for arbitrary graphs. Nevertheless, there is a “quantum Hammersley–Clifford Theorem” for trees which is given by the following statement.

**Theorem 5.1.7** ([PH11, Thm. 1]). *Let  $G = (V, E)$  with  $V = \{v_1, \dots, v_N\}$  be a tree and let  $\rho$  be a positive definite quantum state on a Hilbert space  $\mathcal{H} = \mathcal{H}_1 \otimes \dots \otimes \mathcal{H}_N$  satisfying all QCMI statements imposed by  $G$ . Then  $\rho$  can be written as the exponential of a sum of local commuting Hamiltonians, i.e.  $\rho = \exp(H)$  with*

$$H = \sum_{C \in \mathcal{C}(G)} h_C, \quad [h_C, h_{C'}] = 0 \quad \text{for all } C, C' \in \mathcal{C}(G),$$

where  $\mathcal{C}(G)$  is the set of cliques of  $G$  and  $h_C$  is only nontrivial on the clique  $C$ , i.e.  $h_C$  is an endomorphism on  $\mathcal{H}$  acting as identity on each  $\mathcal{H}_i$  where  $v_i \notin C$ .

This quantum Hammersley–Clifford Theorem is a generalisation of equation (5.1.1) to trees. Along with Example 5.1.5, this suggests the following construction of the QCMI variety of a tree and an associated quantum graphical model.

**Construction 5.1.8** (QCMI variety of a tree). Let  $G = (V, E)$  be an undirected tree with vertices labelled  $S_1, \dots, S_N$ . Let  $\rho_V = \rho_{S_1 \dots S_N}$  be a quantum state on  $\mathcal{H}_1 \otimes \dots \otimes \mathcal{H}_N$ . For each triple of vertices  $S_i, S_j, S_k$  such that  $S_j$  separates  $S_i$  from  $S_k$  in  $G$ , we impose the QCMI statement  $I(S_i : S_k | S_j) = 0$ , i.e. we require

$$\mathrm{Tr}_{V \setminus \{S_i, S_j, S_k\}} \rho_V = \Lambda_{S_i S_j} \Lambda_{S_j S_k} \text{ with } [\Lambda_{S_i S_j}, \Lambda_{S_j S_k}] = 0 \quad (5.1.2)$$

as in (5.1.1). Moreover, for any two QCMI statements  $I(S_i : S_k | S_j) = 0 = I(S_{i'} : S_{k'} | S_{j'})$  we impose the *compatibility constraints*

$$\mathrm{Tr}_{\mathcal{T} \setminus (\mathcal{T} \cap \mathcal{T}')} \rho_{\mathcal{T}} = \mathrm{Tr}_{\mathcal{T}' \setminus (\mathcal{T} \cap \mathcal{T}')} \rho_{\mathcal{T}'} \quad \text{where } \mathcal{T} = (S_i, S_j, S_k), \mathcal{T}' = (S_{i'}, S_{j'}, S_{k'}). \quad (5.1.3)$$

In the qubit case, this construction gives rise to Algorithm 7 whose output is a variety inside  $\mathbb{S}^{2^N}$ . We call this variety the *QCMI variety* associated to  $G$  and denote it by  $\mathcal{Q}_G$ . Algorithm 7 constructs the QCMI variety by considering the conditions (5.1.2), (5.1.3) as polynomial constraints in the entries of an arbitrary density matrix  $\rho$  and of the matrices  $\Lambda_{S_i S_j}, \Lambda_{S_j S_k}$ , then it eliminates the  $\Lambda$  parameters. The QCMI variety  $\mathcal{Q}_G$  defines a *quantum graphical model*  $\mathcal{M}_G$  by restricting to positive semidefinite matrices with trace one inside  $\mathcal{Q}_G$ . Note that Algorithm 7 and the notion of the QCMI variety generalise straightforwardly to arbitrary qudit systems.

---

**Algorithm 7** Computing the QCMI variety  $\mathcal{Q}_G$

---

**Input:** A tree  $G = (V, E)$

**Output:** Polynomials defining  $\mathcal{Q}_G \subseteq \mathbb{S}^{2^N}$

- 1:  $N \leftarrow \#V$
  - 2:  $\rho_V \leftarrow$  symmetric  $(2^N \times 2^N)$ -matrix consisting of variables  $\rho_{11}, \rho_{12}, \dots, \rho_{2^N 2^N}$
  - 3:  $\mathcal{E} \leftarrow \emptyset$  initialise list of equations
  - 4: **for** every triple of vertices  $\mathcal{T} = (S_i, S_j, S_k)$  such that  $S_j$  separates  $S_i$  from  $S_k$  in  $G$  **do**
  - 5:    $\Lambda_{S_i S_j} \leftarrow (\lambda_{lm}^{\mathcal{T}}) \otimes \mathrm{Id}_2$  where  $(\lambda_{lm}^{\mathcal{T}})$  is a symmetric  $4 \times 4$  matrix of variables
  - 6:    $\Lambda_{S_j S_k} \leftarrow \mathrm{Id}_2 \otimes (\mu_{lm}^{\mathcal{T}})$  where  $(\mu_{lm}^{\mathcal{T}})$  is a symmetric  $4 \times 4$  matrix of variables
  - 7:    $\mathcal{E}' \leftarrow$  entries of  $\mathrm{Tr}_{V \setminus \mathcal{T}} \rho_V - \Lambda_{S_i S_j} \Lambda_{S_j S_k}$
  - 8:    $\mathcal{E}'' \leftarrow$  entries of  $[\Lambda_{S_i S_j}, \Lambda_{S_j S_k}]$
  - 9:    $\mathcal{E} \leftarrow \mathcal{E} \cup \mathcal{E}' \cup \mathcal{E}''$
  - 10: **end for**
  - 11: **for** every pair of triples of vertices  $\mathcal{T} = (S_i, S_j, S_k)$  and  $\mathcal{T}' = (S_{i'}, S_{j'}, S_{k'})$  **do**
  - 12:    $\mathcal{E}''' \leftarrow$  entries of  $\mathrm{Tr}_{\mathcal{T} \setminus (\mathcal{T} \cap \mathcal{T}')} \rho_{\mathcal{T}} - \mathrm{Tr}_{\mathcal{T}' \setminus (\mathcal{T} \cap \mathcal{T}')} \rho_{\mathcal{T}'}$
  - 13:    $\mathcal{E} \leftarrow \mathcal{E} \cup \mathcal{E}'''$
  - 14: **end for**
  - 15:  $I \leftarrow$  ideal generated by  $\mathcal{E}$  in  $\mathbb{C}[\rho, \lambda, \mu]$
  - 16:  $J \leftarrow$  elimination ideal  $I \cap \mathbb{C}[\rho]$
  - 17: **return** a set of generators of  $J$
- 

**Remark 5.1.9.** Let  $G$  be the 3-chain graph with ordered vertex labels  $A, B$  and  $C$  and consider the qubit case  $\mathcal{H}_A \cong \mathcal{H}_B \cong \mathcal{H}_C \cong \mathbb{C}^2$ . Then  $\mathcal{Q}_G$  is the variety  $\mathcal{Q}_{I(A:C|B)}$  from Example 5.1.5. The computations in this example were carried out using Algorithm 7.

**Remark 5.1.10.** Note that as we consider trees, it is equivalent to impose QCMI statements on triples of vertices as in Construction 5.1.8 or to impose a *global quantum Markov property*,

in the sense that one imposes the QCMI statement  $I(A : C | B)$  for any triple of sets of vertices  $A, B, C \subseteq V$  such that  $B$  separates  $A$  from  $C$ , as one can derive the latter from the former by using the Weak Union and Contraction axioms from Proposition 5.1.2.

**Example 5.1.11.** Consider the claw tree  $G$  on four vertices with labels  $A, B, C, D$ , the set of edges  $\{\{A, D\}, \{B, D\}, \{C, D\}\}$ , and the corresponding Hilbert spaces  $\mathcal{H}_A \cong \mathcal{H}_B \cong \mathcal{H}_C \cong \mathcal{H}_D \cong \mathbb{C}^2$ . The Hilbert space of the full system is  $\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_B \otimes \mathcal{H}_C \otimes \mathcal{H}_D \cong \mathbb{C}^{16}$ . Every path in  $G$  with three vertices imposes a QCMI statement and any such path contains the node  $D$ . The model  $\mathcal{M}_G$  consists of all states  $\rho$  that satisfy the three QCMI statements

$$I(A : B | D) = I(A : C | D) = I(B : C | D) = 0.$$

These lead to factorisations of the marginal subsystems as

$$\begin{aligned} \text{Tr}_A \rho_{ABCD} &= \rho_{BCD} = \Lambda_{BD} \Lambda_{CD} \text{ with } [\Lambda_{BD}, \Lambda_{CD}] = 0, \\ \text{Tr}_B \rho_{ABCD} &= \rho_{ACD} = \Lambda_{AD} \Lambda_{CD} \text{ with } [\Lambda_{AD}, \Lambda_{CD}] = 0, \\ \text{Tr}_C \rho_{ABCD} &= \rho_{ABD} = \Lambda_{AD} \Lambda_{BD} \text{ with } [\Lambda_{AD}, \Lambda_{BD}] = 0. \end{aligned}$$

In addition, there are compatibility conditions on the marginals which lead to

$$\text{Tr}_B \rho_{ABD} = \text{Tr}_C \rho_{ACD}, \quad \text{Tr}_A \rho_{ABD} = \text{Tr}_C \rho_{BCD}, \quad \text{Tr}_A \rho_{ACD} = \text{Tr}_B \rho_{BCD}.$$

It is computationally very challenging to obtain defining equations of  $\mathcal{Q}_G$  as Algorithm 7 would involve eliminating 60 variables in a polynomial ring in 196 variables, which is infeasible with current computational resources.  $\diamond$

**Question 5.1.12.** Is the variety  $\mathcal{Q}_G$  irreducible for any tree  $G$ ?

It would be desirable to find a parametrisation of  $\mathcal{Q}_G$ . Quantum information theory provides a map that recovers a unique quantum state compatible with given two-body marginals on a tree, called the *Petz recovery map* [HJPW04, Pet86]. However, our algebraic version of this map does not parametrise the QCMI variety  $\mathcal{Q}_G$  as we are working with complex symmetric matrices instead of Hermitian matrices. The Petz recovery map therefore gives rise to a different variety, which we introduce in the next subsection.

## 5.1.2. Petz varieties

The Quantum Marginal Problem asks about how to reconstruct a quantum state of a composite system from the states of its subsystems. In the case of the 3-chain graph with ordered vertices  $A, B$  and  $C$ , one can ask for a quantum state  $\rho_{ABC}$  with given two-body marginals  $\rho_{AB}$  and  $\rho_{BC}$  and satisfying the quantum Markov condition  $I(A : C | B) = 0$ . The answer to this particular problem is given by the Petz recovery map. This map is of algebraic nature and gives rise to the *Petz variety* which we study in this subsection.

We start by introducing the Petz recovery map for the 3-chain graph with the associated Hilbert space  $\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_B \otimes \mathcal{H}_C$ . Let  $\mathcal{C}$  be the set of pairs of compatible invertible density operators on  $\mathcal{H}_A \otimes \mathcal{H}_B$  and  $\mathcal{H}_B \otimes \mathcal{H}_C$ , respectively, i.e. an element in  $\mathcal{C}$  is of the form  $(\rho_{AB}, \rho_{BC})$ , where  $\rho_{AB}$  and  $\rho_{BC}$  are invertible density operators satisfying the compatibility condition  $\text{Tr}_A \rho_{AB} = \text{Tr}_C \rho_{BC}$ . The Petz recovery map  $\mathcal{R}_G$  for the 3-chain graph  $G$  is

$$\mathcal{R}_G: \mathcal{C} \rightarrow \mathcal{D}(\mathcal{H}_A \otimes \mathcal{H}_B \otimes \mathcal{H}_C) \tag{5.1.4}$$

$$\mathcal{R}_G(\rho_{AB}, \rho_{BC}) = (\rho_{AB}^{1/2} \otimes \text{Id}_C)(\text{Id}_A \otimes \rho_B^{-1/2} \otimes \text{Id}_C)(\text{Id}_A \otimes \rho_{BC})(\text{Id}_A \otimes \rho_B^{-1/2} \otimes \text{Id}_C)(\rho_{AB}^{1/2} \otimes \text{Id}_C),$$

where  $\text{Id}_A$ ,  $\text{Id}_B$  and  $\text{Id}_C$  are the identity operators on  $\mathcal{H}_A$ ,  $\mathcal{H}_B$  and  $\mathcal{H}_C$ , respectively. The recovered state is compatible with the marginals and satisfies  $I(A: C | B) = 0$ . Moreover, it is the unique maximiser of the von Neumann entropy among all states on  $\mathcal{H}$  [DGMM20, Thm. 1]. This, in particular, shows that the map  $\mathcal{R}'_G$  defined by

$$\mathcal{R}'_G(\rho_{AB}, \rho_{BC}) = (\text{Id}_A \otimes \rho_{BC}^{1/2})(\text{Id}_A \otimes \rho_B^{-1/2} \otimes \text{Id}_C)(\rho_{AB} \otimes \text{Id}_C)(\text{Id}_A \otimes \rho_B^{-1/2} \otimes \text{Id}_C)(\text{Id}_A \otimes \rho_{BC}^{1/2})$$

recovers the same state [DGMM20, Rem. 2].

The Petz recovery map (5.1.4) gives rise to a rational map  $R_G$ . From this point forward we restrict to the qubit case  $\mathcal{H}_A \cong \mathcal{H}_B \cong \mathcal{H}_C \cong \mathbb{C}^2$ ; the general case is a straightforward generalisation. Let  $\rho_{AB}^{1/2} = X = (x_{ij})$  be a  $4 \times 4$  symmetric matrix of variables  $x_{11}, x_{12}, \dots, x_{44}$ . In the same way, let  $\rho_{BC}^{1/2} = Y = (y_{ij})$ . Finally, let  $\rho_B^{1/2} = Z = (z_{ij})$  be a  $2 \times 2$  symmetric matrix of variables. To reflect the required marginal compatibilities between  $\rho_{AB}, \rho_{BC}$  and  $\rho_B$ , we impose the conditions  $\text{Tr}_C(Y^2) = \text{Tr}_A(X^2) = Z^2$ . These conditions cut out a variety  $V$  in  $\mathbb{S}_{\mathbb{R}}^4 \times \mathbb{S}_{\mathbb{R}}^4 \times \mathbb{S}_{\mathbb{R}}^2$ . In analogy to (5.1.4), the map  $R_G : V \dashrightarrow \mathbb{S}_{\mathbb{R}}^8$  sends a point  $(x, y, z) \in V$  to

$$(x \otimes \text{Id}_C)(\text{Id}_A \otimes z^{-1} \otimes \text{Id}_C)(\text{Id}_A \otimes y)(\text{Id}_A \otimes z^{-1} \otimes \text{Id}_C)(x \otimes \text{Id}_C). \quad (5.1.5)$$

We call the Zariski closure of  $R_G(V)$  inside  $\mathbb{S}^8$ , the space of *complex* symmetric  $8 \times 8$  matrices, the *Petz variety* of  $G$  and denote it  $\mathcal{P}_G$ .

**Remark 5.1.13.** The expression (5.1.5) for  $R_G$  gives a concrete polynomial parametrisation of  $\mathcal{P}_G$ . The polynomials appearing in  $R_G$  have degree five and have a minimum of 20 and maximum of 32 terms. The number of parameter variables is 23 while the variety  $V$  has dimension 17. Algorithm 8 provides a symbolic routine to compute the ideal of the Petz variety for arbitrary trees. When restricting to the case where  $x, y$  and  $z$  are diagonal, (5.1.5) gives yet another parametrisation of the classical graphical model of the 3-chain from the introduction.

**Proposition 5.1.14.** *Let  $G$  be the 3-chain graph. The Petz variety  $\mathcal{P}_G$  is irreducible.*

*Proof.* Consider the subset  $S \subset \mathbb{S}_{\mathbb{R}}^8 \times \mathbb{S}_{\mathbb{R}}^8$  consisting of pairs of invertible matrices whose partial traces agree. The condition that their partial traces agree defines a linear subspace of  $\mathbb{S}_{\mathbb{R}}^8 \times \mathbb{S}_{\mathbb{R}}^8$ . Linear spaces are irreducible and taking out the locus of positive codimension where the matrices become singular preserves irreducibility. Therefore,  $S$  is irreducible. Note that  $\mathcal{R}_G$  can be considered as a map on  $S$  and the Zariski closure of its image coincides with  $\mathcal{P}_G$  since square roots and inverses of symmetric matrices are again symmetric. Therefore, since  $\mathcal{R}_G$  is continuous,  $\mathcal{P}_G$  is irreducible.  $\square$

The Petz map can be generalised to arbitrary trees by iteratively applying the procedure for 3-chains of a tree  $G$  [DGM21, DGMM20]. This is done by taking two leaves  $v_1$  and  $v_2$  of a tree  $G$ , and replacing  $G \setminus \{v_1, v_2\}$  by a single vertex representing a joint state on this subgraph. The joint state on  $G$  is then expressed in terms of states on  $G \setminus \{v_1\}$  and  $G \setminus \{v_2\}$  via (5.1.4); by applying this procedure iteratively to  $G \setminus \{v_1\}$  and  $G \setminus \{v_2\}$ , we reduce to the level of two-body marginals. This process leads to a map as in (5.1.4) involving only one- and two-body marginals; again, we denote the resulting map by  $\mathcal{R}_G$ . Note that the expression for  $\mathcal{R}_G$  depends on the choice of  $v_1$  and  $v_2$  in each iteration.

We now generalise the construction of the Petz variety to arbitrary trees.

**Construction 5.1.15** (Petz variety). Let  $G$  be a tree with  $N$  vertices and let us fix an expression for  $\mathcal{R}_G$  as obtained in the previous paragraph. Let  $\varrho_1$  and  $\varrho_2$  be the sets of

one- and two-body marginals occurring in  $\mathcal{R}_G$ . Moreover, let  $V$  be the variety inside  $S = (\mathbb{S}_{\mathbb{R}}^2)^{\#q_1} \times (\mathbb{S}_{\mathbb{R}}^4)^{\#q_2}$  consisting of tuples of symmetric matrices satisfying compatibility constraints according to  $G$ . In analogy to (5.1.4),  $\mathcal{R}_G$  gives rise to a rational map  $R_G: V \dashrightarrow \mathbb{S}_{\mathbb{R}}^{2^N}$ . The *Petz variety*  $\mathcal{P}_G$  of  $G$  is defined as  $\overline{R_G(V)} \subseteq \mathbb{S}^{2^N}$ .

Algorithm 8 makes this construction explicit and computes the ideal of  $\mathcal{P}_G$ .

---

**Algorithm 8** Computing the Petz variety  $\mathcal{P}_G$ 


---

**Input:** A tree  $G$  with  $N$  vertices

**Output:** An ideal defining the Petz variety  $\mathcal{P}_G \subseteq \mathbb{S}^{2^N}$

- 1:  $R_G \leftarrow$  expression for  $\mathcal{R}_G$  in terms of one- and two-body marginals
  - 2:  $q_1, q_2 \leftarrow$  sets of one- and two-body marginals, respectively, involved in  $R_G$
  - 3: **for all**  $\rho_v \in q_1$  **do**
  - 4:      $Z_v \leftarrow (z_{ij}^v)$  symmetric  $2 \times 2$  matrix of variables
  - 5: **end for**
  - 6: **for all**  $\rho_{v_1 v_2} \in q_2$  **do**
  - 7:      $X_{v_1 v_2} \leftarrow (x_{ij}^{v_1 v_2})$  symmetric  $4 \times 4$  matrix of variables
  - 8: **end for**
  - 9:  $S \leftarrow (\mathbb{S}_{\mathbb{R}}^2)^{\#q_1}_{\{Z_{v_i}\}} \times (\mathbb{S}_{\mathbb{R}}^4)^{\#q_2}_{\{X_{v_i v_j}\}}$
  - 10:  $\mathcal{E} \leftarrow \emptyset$
  - 11: **for all pairs**  $(\rho_{v_1 v_2}, \rho_{w_1 w_2}) \in q_2^2$  such that  $v_2 = w_1$  **do**
  - 12:      $\mathcal{E} \leftarrow \mathcal{E} \cup \{\text{entries of } \text{Tr}_{v_1}(X_{v_1 v_2}^2) - \text{Tr}_{w_2}(X_{w_1 w_2}^2)\} \cup \{\text{entries of } \text{Tr}_{v_1}(X_{v_1 v_2}^2) - Z_{v_2}^2\}$
  - 13: **end for**
  - 14:  $V \leftarrow$  variety defined by  $\mathcal{E}$  inside  $S$
  - 15:  $R_G \leftarrow R_G$  with every  $\rho_{v_1 v_2}$  replaced by  $X_{v_1 v_2}$  and every  $\rho_v$  replaced by  $Z_v$
  - 16: **return**  $\ker(R_G: \mathbb{C}[\mathbb{S}^{2^N}] \rightarrow \mathbb{C}[V])$
- 

**Proposition 5.1.16.** *The Petz variety  $\mathcal{P}_G$  does not depend on the particular choice of expression for the recovery map  $\mathcal{R}_G$ .*

*Proof.* By the same argument as for the 3-chain graph above, the map  $\mathcal{R}_G$  does not depend on the chosen expression. Let  $\mathcal{C}$  be the domain of  $\mathcal{R}_G$ ; each element of  $\mathcal{C}$  is a tuple consisting of  $\#E$ -many compatible invertible two-body marginals, where  $E$  is the set of edges of  $G$ . Consider the set  $(\mathbb{S}_{\mathbb{R}}^4)^{\#E}$  of  $\#E$ -tuples of real symmetric  $4 \times 4$  matrices; it is Zariski dense in the set of complex symmetric matrices. We have  $\mathcal{R}_G(\mathcal{C} \cap (\mathbb{S}_{\mathbb{R}}^4)^{\#E}) = R_G(V \cap U)$  where  $U \subseteq \mathbb{C}^{2^N \times 2^N}$  is the Zariski dense open set of invertible matrices. The set  $R_G(V \cap U)$  is Zariski dense in the Petz variety  $\mathcal{P}_G$ . Let  $\mathcal{R}'_G$  be another expression for the recovery map, then  $\mathcal{R}_G(\mathcal{C} \cap (\mathbb{S}_{\mathbb{R}}^4)^{\#E}) = R'_G(V \cap U)$  and denote by  $\mathcal{P}'_G$  the variety defined by  $R'_G$ . It follows that  $R'_G(V \cap U) = R_G(V \cap U)$  so  $\mathcal{P}_G$  and  $\mathcal{P}'_G$  agree on a dense open set, thus  $\mathcal{P}_G = \overline{R_G(V)} = \overline{R'_G(V)} = \mathcal{P}'_G$ .  $\square$

**Proposition 5.1.17.** *For any tree  $G$ , the Petz variety  $\mathcal{P}_G$  is irreducible.*

*Proof.* The proof is analogous to that of Proposition 5.1.14:  $\mathcal{P}_G$  can be represented as the Zariski closure of the image of a linear space under a continuous map.  $\square$

**Remark 5.1.18.** Computing the defining equations of the Petz variety is very challenging. Even in the case of the 3-chain graph, Algorithm 8 does not terminate as it involves symbolic computations in a polynomial ring with 59 variables. Applying numerical implicitisation techniques is also not straightforward for the same reason.

**Remark 5.1.19.** If we considered Hermitian matrices, the set of states recovered by the Petz map would coincide with the set of states satisfying SSA with equality [HJPW04]. However, since the ambient space of our varieties is that of complex symmetric matrices, the QCMI variety  $\mathcal{Q}_G$  and the Petz variety  $\mathcal{P}_G$  are not the same.

### 5.1.3. Quantum graphical models from Gibbs manifolds

In this subsection, we consider a new class of quantum graphical models, which arise as Gibbs manifolds of families of Hamiltonians. These serve as examples of *quantum exponential families* [Zho08]. Gibbs manifolds and Gibbs varieties have been introduced in [PST23] as tools to study convex optimisation and, in particular, quantum optimal transport from an algebro-geometric perspective; we recall the definition.

**Definition 5.1.20.** Let  $\mathcal{L}$  be a linear space of real symmetric matrices (LSSM) of size  $n \times n$ . The *Gibbs manifold*  $\text{GM}(\mathcal{L})$  of  $\mathcal{L}$  is  $\exp(\mathcal{L})$  where  $\exp$  denotes the matrix exponential. The *Gibbs variety*  $\text{GV}(\mathcal{L})$  of  $\mathcal{L}$  is the Zariski closure of  $\text{GM}(\mathcal{L}) \subseteq \mathbb{S}^n$ .

In physics, the Gibbs manifold parametrises thermal states of a family of Hamiltonians. Those states are crucial to quantum many-body systems theory and computation [Alh23]. The Gibbs manifold is not in general algebraic, while the Gibbs variety is. In special cases, their dimensions coincide. This happens, for instance, when  $\mathcal{L}$  consists of diagonal matrices with rational entries. For such  $\mathcal{L}$  the resulting Gibbs variety is toric [PST23, Thm. 2.7] and recovers the classical notion of exponential families [Efr22]. Moreover, the Gibbs manifold in this case is the intersection of the Gibbs variety with the cone of positive definite matrices. One advantage of considering the Gibbs variety instead of the Gibbs manifold is that it becomes possible to treat related problems in quantum information by using symbolic and numerical methods from algebraic geometry.

### 5.1.4. Gibbs varieties of linear systems of Hamiltonians

The quantum Hammersley–Clifford Theorem (Theorem 5.1.7) suggests considering exponentials of *local* Hamiltonians, i.e. those Hamiltonians that act nontrivially only on a small subsystem. However, we do not consider the class of local *commuting* Hamiltonians as they neither form an LSSM nor a unirational variety (see Subsection 5.1.5).

To a simple, undirected graph  $G = (V, E)$  we associate an LSSM of Hamiltonians as follows. For each clique  $C$  in  $G$ , let  $\mathcal{L}_C$  be the LSSM given by all Hamiltonians *supported* on  $C$ , i.e. those that act nontrivially only on the tensor factors  $\mathcal{H}_i$  such that  $v_i \in C$  and act as identity on all other subsystems. More precisely,  $\mathcal{L}_C = \otimes_i L_C^i$  where  $L_C^i = \mathbb{S}^{d_i}$  for  $v_i \in C$  and  $L_C^i = \text{Id}_{d_i}$  otherwise. The family of Hamiltonians associated to  $G$  is then  $\mathcal{L}_G = \sum_{C \in \mathcal{C}(G)} \mathcal{L}_C$  where the sum runs over all cliques of  $G$  [WG23, Equ. 17]. The quantum graphical model is  $\text{GM}(\mathcal{L}_G)$  intersected with the space of trace one matrices. The Gibbs variety  $\text{GV}(\mathcal{L}_G)$  gives an algebraic description of this model.

**Example 5.1.21.** Consider the 3-chain graph  $G$  and assume we are in the qubit case, i.e.  $\mathcal{H}_A \cong \mathcal{H}_B \cong \mathcal{H}_C \cong \mathbb{C}^2$ . Then we have  $\mathcal{L}_G = \mathbb{S}^2 \otimes \mathbb{S}^2 \otimes \text{Id}_2 + \text{Id}_2 \otimes \mathbb{S}^2 \otimes \mathbb{S}^2$ . Using numerical algebraic geometry techniques, we verify that no linear or quadratic equations vanish on  $\text{GV}(\mathcal{L}_G)$ . The higher degree equations are not amenable to our computational techniques. The dimension of  $\text{GM}(\mathcal{L}_G)$  is 15 while the dimension of  $\text{GV}(\mathcal{L}_G)$  is at most 22.  $\diamond$

Gibbs varieties have a number of nice theoretical properties, e.g. under mild assumptions they are unirational [PST23, Thm. 3.6]. However, as seen in Example 5.1.21, their

defining ideals are often difficult to compute. In view of this, one might hope to simplify the defining equations of the Gibbs variety by restricting the family of Hamiltonians to a subset inside  $\mathcal{L}_G$ . This approach is pursued in the next subsection.

### 5.1.5. Gibbs varieties of unirational varieties

A natural subset to consider inside  $\mathcal{L}_G$  is  $X_G := \sum_{C \in \mathcal{C}(G)} X_C$ , where  $X_C$  is the set of decomposable tensors supported on  $C$ . Note that  $X_G$  is not a linear space. However, it is still a unirational variety. This motivates the following extension of the notion of Gibbs varieties.

**Definition 5.1.22.** Let  $X$  be a unirational variety of symmetric matrices of size  $n \times n$ . The *Gibbs variety*  $GV(X)$  of  $X$  is the Zariski closure of  $\exp(X) \subseteq \mathbb{S}^n$ .

A number of concepts related to Gibbs varieties of linear spaces generalise to the case of unirational varieties of symmetric matrices. If  $X$  is unirational and has dimension  $d$ , then it can be parametrised by rational functions in  $d$  variables  $y_1, \dots, y_d$ . Therefore, one can think of  $X$  as a single matrix with entries in  $\mathbb{C}(y_1, \dots, y_d)$ . The eigenvalues of this matrix are elements of  $\mathbb{C}(y_1, \dots, y_d)$  and will be referred to as the eigenvalues of  $X$ . Let  $A \in \mathbb{S}^n$ , then the  $X$ -centraliser of  $A$  is the set  $\mathcal{Z}_X(A) = \{B \in X : AB - BA = 0\}$ . We collect properties of Gibbs varieties of unirational varieties in the following statement.

**Proposition 5.1.23.** Let  $X$  be a unirational variety of  $n \times n$  symmetric matrices of dimension  $d$ . Let  $m$  be the dimension of the  $\mathbb{Q}$ -linear space spanned by the eigenvalues of  $X$  and  $k$  be the dimension of the  $X$ -centraliser of a generic element of  $X$ . Then  $\dim(GV(X)) = m + d - k$ . In particular,  $\dim(GV(X)) \leq n + d$ . If  $X$  has distinct eigenvalues, then  $GV(X)$  is irreducible and unirational.

*Proof.* This proposition generalises [Pav24, Thm. 2.6] and [PST23, Thm. 3.6]. Proofs of these statements carry over to the case of unirational varieties, since they only use the fact that an LSSM can be parametrised by rational functions in  $y_1, \dots, y_d$  and do not depend on these functions being linear.  $\square$

Note that symbolic [PST23, Alg. 1] and numerical [Pav24, Alg. 1] implicitisation algorithms for Gibbs varieties generalise accordingly.

**Example 5.1.24.** Again, consider the 3-chain graph  $G$  in the qubit case. The associated unirational variety is  $X_G = \{K \otimes L \otimes \text{Id}_2 + \text{Id}_2 \otimes M \otimes N : K, L, M, N \in \mathbb{S}^2\} \subseteq \mathbb{S}^8$ . The dimension of  $X$  is equal to ten inside the 36-dimensional space of symmetric  $8 \times 8$  matrices. The Gibbs variety  $GV(X)$  is a 14-dimensional irreducible variety cut out by nine linear forms and 66 quadratic equations in  $\mathbb{S}^8$ . These results were obtained by using techniques of numerical algebraic geometry. More precisely, we create a sample of points on  $GM(X)$  and then interpolate with polynomials of a fixed degree by setting up a Vandermonde matrix and computing its kernel via QR-decomposition to obtain a sparse representation, see [BKSW18]; this procedure yields polynomials of degree one and two. As these equations cut out an irreducible variety of the correct dimension, we obtain a generating set of the prime ideal of  $GV(X)$ .

The equations we obtained exhibit a remarkably simple structure; e.g. all polynomials have coefficients  $\pm 1$  and consist of at most eight terms. It would be very interesting to obtain a theoretical explanation of this phenomenon.  $\diamond$

In the following two subsections we explore a completely different family of Hamiltonians  $\mathfrak{H}_G$  associated to a graph  $G$ . This results in quantum exponential families that

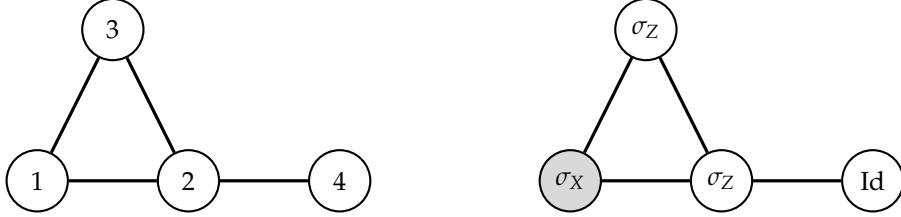


Figure 5.1: Left: the graph  $G$ . Right: illustration of the Hamiltonian  $H_1$ .

have a richer structure than the ones considered in Section 5.1.3. However, it should be emphasised that, unlike the previous constructions, this is not a generalisation of classical graphical models. On the other hand, all the results in this section hold for general undirected graphs, not only trees.

### 5.1.6. Quantum exponential families of commuting Hamiltonians

In quantum physics, to an undirected graph  $G$  one associates an LSSM  $\mathfrak{H}_G$  which gives rise to the definition of *graph states*, used in the study of entanglement, e.g. [HEB04]. More precisely, the graph state associated to  $G$  is the *stabiliser state* of  $\mathfrak{H}_G$ . Stabiliser states appear in the framework of the stabiliser formalism used in quantum error correction, see Section 2.6; in fact, all results in this section generalise to stabilisers. However, instead of studying the graph states associated to  $\mathfrak{H}_G$ , here we focus on its Gibbs variety; this latter perspective gives yet another example of quantum exponential families.

Let  $G = (V, E)$  be a graph with vertices  $V = \{v_1, \dots, v_N\}$ . To each vertex  $v_i$ , we associate a Hamiltonian  $H_i = \bigotimes_{j=1}^N H_{i,j}$  with

$$H_{i,j} = \begin{cases} \sigma_X & \text{if } i = j, \\ \sigma_Z & \text{if } (i, j) \in E, \\ \text{Id}_2 & \text{else.} \end{cases}$$

Here,  $\sigma_X$  and  $\sigma_Z$  are the Pauli matrices introduced in Section 2.6. Denote the linear span of this set of Hamiltonians by  $\mathfrak{H}_G := \langle H_i : i = 1, \dots, N \rangle$ . The Hamiltonians  $H_i$  are elements of the Pauli group  $\mathcal{P}_N$ .

**Example 5.1.25.** Consider the graph  $G$  on four vertices depicted in Figure 5.1.25. The Hamiltonian  $H_1$  is given by

$$H_1 = \sigma_X \otimes \sigma_Z \otimes \sigma_Z \otimes \text{Id}_2.$$

The linear space  $\mathfrak{H}_G$  is spanned by the four Hamiltonians

$$\sigma_X \otimes \sigma_Z \otimes \sigma_Z \otimes \text{Id}_2, \sigma_Z \otimes \sigma_X \otimes \sigma_Z \otimes \sigma_Z, \sigma_Z \otimes \sigma_Z \otimes \sigma_X \otimes \text{Id}_2, \text{Id}_2 \otimes \sigma_Z \otimes \text{Id}_2 \otimes \sigma_X. \quad \diamond$$

In the following we consider the Gibbs variety of  $\mathfrak{H}_G$ . We start by showing that  $\mathfrak{H}_G$  is a commuting family, implying that  $\text{GV}(\mathfrak{H}_G)$  is *toric* after a linear change of coordinates and the Gibbs manifold  $\text{GM}(\mathfrak{H}_G)$  is semialgebraic [PST23, Thm. 2.7].

**Lemma 5.1.26.** *Any  $H, H' \in \mathfrak{H}_G$  commute.*

*Proof.* Without loss of generality, assume  $H$  and  $H'$  are generators  $H_m$  and  $H_n$  of  $\mathfrak{H}_G$ . For  $P \in \mathcal{P}_N$ , let  $\text{Supp}_X(P) := \{j \in [N] : \sigma^{(j)} = \sigma_X\}$  and  $\text{Supp}_Z(P) := \{j \in [N] : \sigma^{(j)} = \sigma_Z\}$  denote the supports of  $\sigma_X$  and  $\sigma_Z$ , respectively. As a consequence of the commutation

relation (2.6.1), two Pauli product matrices  $P, Q \in \mathcal{P}_N$  containing only  $\text{Id}_2, \sigma_X$  or  $\sigma_Z$  as tensor factors commute if and only if

$$\#(\text{Supp}_X(P) \cap \text{Supp}_Z(Q)) + \#(\text{Supp}_Z(P) \cap \text{Supp}_X(Q)) \equiv 0 \pmod{2}. \quad (5.1.6)$$

Let  $\mathcal{N}(v)$  denote the set of neighbouring vertices of  $v$  in  $G$ . Assume  $v_n \in \mathcal{N}(v_m)$ ; then the left-hand side of (5.1.6) for  $P = H_m$  and  $Q = H_n$  becomes  $\#\{m\} + \#\{n\} = 2$ . Finally, if  $v_n \notin \mathcal{N}(v_m) \cup \{v_m\}$  the left-hand side of (5.1.6) is just zero.  $\square$

Let us recall from [PST23] how to obtain the toric variety and the coordinate change from  $\text{GV}(\mathfrak{H}_G)$ . The symmetric matrices  $H_1, \dots, H_N \in \mathfrak{H}_G$  are simultaneously diagonalisable, i.e. there exist an orthogonal matrix  $U$  and diagonal matrices  $D_1, \dots, D_N$  such that  $U^{-1}H_iU = D_i$  for  $i = 1, \dots, N$ . The exponential of an element in  $\mathfrak{H}_G$  is then

$$\exp(x_1H_1 + \dots + x_NH_N) = U \exp(x_1D_1 + \dots + x_ND_N)U^{-1}$$

and thus  $\text{GV}(\mathfrak{H}_G) = U \cdot \text{GV}(\mathfrak{D}) \cdot U^{-1}$ , where  $\mathfrak{D} = \langle D_1, \dots, D_N \rangle$ . Let  $D_i = \text{diag}(d_i), i = 1, \dots, N$  for  $d_i \in \mathbb{R}^{2^N}$  and let  $\mathcal{D} = \langle d_1, \dots, d_N \rangle \subseteq \mathbb{R}^{2^N}$  be the  $\mathbb{R}$ -vector space spanned by the diagonals. Consider the smallest vector subspace  $\mathcal{D}_{\mathbb{Q}} \subseteq \mathbb{R}^{2^N}$  containing  $\mathcal{D}$  that is spanned by elements of  $\mathbb{Q}^{2^N}$  and choose an integral basis  $a_1, \dots, a_N \in \mathbb{Z}^{2^N}$  of  $\mathcal{D}_{\mathbb{Q}}$ . If  $A$  denotes the  $N \times 2^N$  matrix with rows  $a_1, \dots, a_N$  then  $\text{GV}(\mathfrak{D})$  is the toric variety  $X_A$  associated to  $A$ .

We now connect quantum exponential families to toric varieties.

**Remark 5.1.27.** Note that any Pauli product matrix  $P \in \mathcal{P}_N$  has eigenvalues  $\pm 1$ , both with multiplicity  $2^{N-1}$  each. Lemma 2.6.2 can then be rephrased as follows: the  $(\pm 1)$ -eigenspaces of  $P_i$  intersect the eigenspaces of all  $P_1, \dots, P_{i-1}$  in half their dimension. This fact is essential to the next theorem establishing a strong connection between quantum exponential families and classical algebraic statistics.

**Theorem 5.1.28.** *For any graph  $G$  with  $N$  vertices,  $\text{GV}(\mathfrak{H}_G)$  is an independence model on  $N$  binary random variables after a linear change of coordinates.*

*Proof.* As shown above,  $\text{GV}(\mathfrak{H}_G) = U \cdot X_A \cdot U^{-1}$  where the rows of  $A$  are the diagonal entries of  $D_i = U^{-1}H_iU$  for  $i = 1, \dots, N$  and  $X_A$  is the (affine) toric variety associated to  $A$ . By Remark 5.1.27, we can assume  $A$  to be of the form

$$A = \begin{pmatrix} -1 & -1 & -1 & -1 & \cdots & -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 & 1 & \cdots & 1 & 1 & 1 & 1 \\ -1 & -1 & -1 & -1 & \cdots & 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 & -1 & \cdots & 1 & 1 & 1 & 1 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ -1 & -1 & 1 & 1 & \cdots & -1 & -1 & 1 & 1 & -1 & -1 & 1 & 1 & 1 & \cdots & -1 & -1 & 1 & 1 \\ -1 & 1 & -1 & 1 & \cdots & -1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 & \cdots & -1 & 1 & -1 & 1 \end{pmatrix} \quad (5.1.7)$$

i.e. the columns of  $A$  are the vertices of the  $N$ -dimensional hypercube  $[-1, 1]^N$ . Thus,  $X_A$  is an independence model on  $N$  binary random variables.  $\square$

**Remark 5.1.29.** The variety  $X_A$  above is not the independence model in its standard description (as e.g. in Example 2.2.3). For example, for  $N = 3$  we have

$$A = \begin{pmatrix} -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 \\ -1 & -1 & 1 & 1 & -1 & -1 & 1 & 1 \\ -1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 \end{pmatrix};$$

the prime ideal of  $X_A$  is

$$I_A = \langle x_1x_8 - 1, x_6x_7 - x_5x_8, x_4x_7 - x_3x_8, x_2x_7 - 1, x_4x_6 - x_2x_8, \\ x_3x_6 - 1, x_4x_5 - 1, x_3x_5 - x_1x_7, x_2x_5 - x_1x_6, x_2x_3 - x_1x_4 \rangle \subseteq \mathbb{C}[x_1, \dots, x_8].$$

The matrix  $A'$  of the independence model has as columns the vertices of the hypercube  $[0, 1]^N$ . Adding a row of ones to  $A$  and to  $A'$  yields the same variety. Thus,  $X_A$  is an affine patch of the independence model, i.e. of the Segre variety  $\sigma(\mathbb{P}^1 \times \mathbb{P}^1 \times \mathbb{P}^1) \subseteq \mathbb{P}^7$ .

In view of the stabiliser formalism laid out in Section 2.6 it becomes apparent that Theorem 5.1.28 can be generalised to arbitrary stabilisers.

**Theorem 5.1.30.** *Let  $S = \langle p_1, \dots, p_N \rangle \leq \mathcal{P}_N$  be generated by  $N$  independent and commuting Pauli product matrices such that  $-\text{Id}_{2^N} \notin S$ . Then  $\text{GV}(S)$  is an independence model on  $N$  binary random variables after a linear change of coordinates.*

*Proof.* The proof of Theorem 5.1.28 immediately generalises to this situation since we only used that the Hamiltonians from graph states  $\mathfrak{H}_G$  are independent and commuting Pauli product matrices not containing the negative identity.  $\square$

A priori, it is not obvious how to obtain defining equations for  $\text{GV}(\mathfrak{H}_G)$  computationally in an efficient manner. However, Theorem 5.1.28 gives rise to Algorithm 9 making computations of defining ideals for graphs with four or more vertices feasible.

---

**Algorithm 9** Computing defining equations of  $\text{GV}(\mathfrak{H}_G)$

---

**Input:** A graph  $G$

**Output:** Polynomials defining  $\text{GV}(\mathfrak{H}_G)$

- 1: Compute  $\mathfrak{H}_G = \langle H_1, \dots, H_N \rangle$
  - 2:  $U \leftarrow$  matrix simultaneously diagonalising  $H_1, \dots, H_N$
  - 3:  $I \leftarrow$  ideal of the independence model defined by  $A$  as in (5.1.7) in variables  $p_{i,i}$  for  $i = 1, \dots, 2^N$
  - 4:  $Y = (y_{ij}) \leftarrow$  linear coordinate change according to  $UPU^{-1}$  where  $P = (p_{ij})$
  - 5: **for** all generators  $g_k$  of  $I$  **do**
  - 6:      $h_k \leftarrow g_k$  changed to  $Y$ -coordinates
  - 7: **end for**
  - 8:  $J \leftarrow$  ideal generated by all  $h_k$  and  $y_{ij} = 0$  for all  $i \neq j \in \{1, \dots, 2^N\}$
  - 9: **return** a set of generators of  $J$
- 

Note that Step 3 in Algorithm 9 can be pre-computed. This allows to reduce finding the equations of  $\text{GV}(\mathfrak{H}_G)$  to a linear algebra problem, therefore reducing the computational complexity. An implementation of this algorithm is available at [DPW23b].

**Example 5.1.31.** Let  $G$  be the graph from Example 5.1.25. Using Algorithm 9, we compute that  $\text{GV}(\mathfrak{H}_G) \subseteq \mathbb{S}^{16}$  is defined by 296 quadratic equations in 136 variables. The average number of terms of each generator is about 1982. This highlights the fact that the equations defining the Gibbs variety can be quite involved. It would be impossible to compute these equations without using the additional structure of  $\mathfrak{H}_G$ : both [PST23, Alg. 1] and [Pav24, Alg. 1] failed to compute this example.  $\diamond$

### 5.1.7. Quantum information projections

Given an arbitrary quantum state  $\rho$  we can ask for the member  $\tilde{\rho} \in \mathcal{Q}$  of some quantum exponential family  $\mathcal{Q} = \text{GM}(\mathcal{L})$  “closest” to  $\rho$ . Here, “closest” means minimising the *quantum relative entropy*.

**Definition 5.1.32.** The *quantum relative entropy*  $D(\rho||\sigma)$  between a state  $\rho$  and a positive semidefinite operator  $\sigma$  is

$$D(\rho||\sigma) := \begin{cases} \text{Tr}(\rho(\log(\rho) - \log(\sigma))) & \text{if } \text{Supp}(\rho) \subseteq \text{Supp}(\sigma) \\ +\infty & \text{otherwise.} \end{cases}$$

Here, the support of a linear operator is the subspace orthogonal to the kernel with respect to the standard inner product on  $\mathbb{C}^n$ , and all logarithms are taken to have base two.

This is a quantum generalisation of the Kullback–Leibler divergence in classical information theory. Note that, similarly to the Kullback–Leibler divergence, the quantum relative entropy is *not* an actual metric as it is not symmetric and does not satisfy the triangle inequality. However, it does satisfy nonnegativity (quantum Gibbs’ inequality). More precisely, if  $\text{Tr}(\sigma) \leq 1$  we have  $D(\rho||\sigma) \geq 0$  with equality if and only if  $\rho = \sigma$ . See [Wil13, §11.8] for an extensive reference. Since minimising the Kullback–Leibler divergence is equivalent to maximum likelihood estimation, minimising the quantum relative entropy is the quantum analogue of maximum likelihood estimation. See Chapter 3 for the geometric study of MLE.

**Definition 5.1.33.** The *quantum information projection*  $\tilde{\rho}$  of a quantum state  $\rho$  to a quantum exponential family  $\mathcal{Q}$  is the element of  $\mathcal{Q}$  which is the closest to  $\rho$  with respect to the quantum relative entropy

$$\tilde{\rho} = \underset{\rho' \in \mathcal{Q}}{\operatorname{argmin}} D(\rho||\rho').$$

The quantum information projection is unique and has been characterised in the case where  $\mathcal{Q}$  consists of exponentials of local Hamiltonians [NGKG13, Lem. 2]. Since the Gibbs manifold considered in the previous subsection is semialgebraic, we can use algebraic techniques to find the quantum information projection in this case. The following theorem gives an algebraic characterisation of the quantum information projection for a quantum exponential family  $\mathcal{Q}$  of commuting Hamiltonians, in particular for  $\mathcal{Q} = \text{GM}(\mathfrak{H}_G)$ .

**Theorem 5.1.34.** Let  $\mathfrak{H} = \langle H_1, \dots, H_k \rangle$  be a linear span of commuting Hamiltonians in  $\mathbb{S}_{\mathbb{R}}^d$ , fix a positive definite matrix  $\rho \in \mathbb{S}_{\mathbb{R}}^d$  and let  $b_i := \langle H_i, \rho \rangle = \text{Tr}(H_i \rho)$  for  $i = 1, \dots, k$ . Let  $M_\rho$  be the affine linear space defined by

$$M_\rho := \{A \in \mathbb{S}_{\mathbb{R}}^d : \langle H_i, A \rangle = b_i \text{ for } i = 1, \dots, k\}.$$

Then  $M_\rho \cap \text{GM}(\mathfrak{H})$  consists of a unique point  $\rho^*$ . It is the maximiser of the von Neumann entropy inside  $M_\rho$  and the quantum information projection of  $\rho$  to  $\text{GM}(\mathfrak{H})$ .

**Remark 5.1.35.** This result generalises Birch’s Theorem (Theorem 2.2.11) to quantum exponential families that become toric after an orthogonal change of coordinates. It is an interesting problem in algebraic statistics to determine which statistical models become toric after some linear change of coordinates, see e.g. [MP23, KV24].

*Proof.* The fact that  $M_\rho \cap \text{GM}(\mathfrak{H}) = \{\rho^*\}$  and  $\rho^*$  is the unique point maximising the von Neumann entropy is a direct consequence of [PST23, Thm. 5.1]. It remains to show that this point is the quantum information projection of  $\rho$  to  $\text{GM}(\mathfrak{H})$ .

Let  $\tilde{\rho}$  be the quantum information projection of  $\rho$  to  $\text{GM}(\mathfrak{H})$ . As in the discussion preceding Remark 5.1.27, let  $U$  be the matrix diagonalising  $\mathfrak{H}$  into  $\mathfrak{D} = \langle D_1, \dots, D_k \rangle$ , i.e.  $H_i = UD_iU^{-1}$  for  $i = 1, \dots, k$ , so  $\tilde{\rho} \in U\text{GM}(\mathfrak{D})U^{-1}$ . Minimising the quantum relative entropy between  $\rho$  and  $\text{GM}(\mathfrak{H})$  is then equivalent to minimising the quantum relative entropy between  $U^{-1}\rho U$  and  $\text{GM}(\mathfrak{D})$ . Let  $\sigma := U^{-1}\rho U$ ; then we want to maximise  $\text{Tr}(\sigma \log(\Delta))$  over diagonal matrices  $\Delta \in \text{GM}(\mathfrak{D})$ , i.e. find  $\tilde{\rho}' = \text{diag}(\hat{\delta})$  such that

$$\tilde{\rho}' = \underset{\Delta \in \text{GM}(\mathfrak{D})}{\text{argmax}} \sum_j \sigma_{jj} \log \Delta_{jj}.$$

This is the same problem as finding the maximum likelihood estimator on the exponential family  $\text{GM}(\mathfrak{D})$  given data  $u = (\sigma_{11}, \sigma_{22}, \dots, \sigma_{dd})$ . Note that every coordinate of  $u$  is nonzero. By Birch's Theorem,  $\hat{\delta}$  is the unique point on  $\text{GM}(\mathfrak{D})$  satisfying  $A\hat{\delta} = Au$  where  $A$  is the matrix whose  $i$ th row is the diagonal of  $D_i$  as in (5.1.7). Observe that

$$(Au)_i = \sum_j (D_i)_{jj} \sigma_{jj} = \text{Tr}(D_i \sigma) = \text{Tr}(D_i U^{-1} \rho U) = \text{Tr}(UD_i U^{-1} \rho) = \text{Tr}(H_i \rho) = b_i;$$

analogously,  $(A\hat{\delta})_i = \text{Tr}(H_i \hat{\delta})$ . This shows  $\tilde{\rho} \in M_\rho \cap \text{GM}(\mathfrak{H})$  and thus  $\tilde{\rho} = \rho^*$ .  $\square$

Theorem 5.1.34 provides a way to compute the quantum information projection to  $\text{GM}(\mathfrak{H}_G)$  algorithmically by using numerical algebraic geometry; concretely, one can first compute  $M_\rho \cap \text{GV}(\mathfrak{H}_G)$  and then choose the unique point lying in the PSD cone.

**Example 5.1.36.** Consider the positive definite matrix

$$\rho = \begin{pmatrix} 84 & -22 & 11 & -51 & -15 & -8 & -26 & 4 \\ -22 & 51 & -5 & -7 & 23 & -13 & 17 & 40 \\ 11 & -5 & 51 & 25 & -16 & -3 & 9 & 28 \\ -51 & -7 & 25 & 70 & -19 & 17 & 18 & -26 \\ -15 & 23 & -16 & -19 & 92 & 32 & 23 & 24 \\ -8 & -13 & -3 & 17 & 32 & 62 & 2 & -36 \\ -26 & 17 & 9 & 18 & 23 & 2 & 94 & 10 \\ 4 & 40 & 28 & -26 & 24 & -36 & 10 & 109 \end{pmatrix}$$

and the 3-chain graph  $G$ . The intersection  $M_\rho \cap \text{GV}(\mathfrak{H}_G)$  consists of six real matrices. Only one of them is positive semidefinite, namely the matrix

$$\tilde{\rho} = \begin{pmatrix} 20.5417 & -12.5 & -20.5 & -12.4746 & -5.5 & 3.34685 & -5.48884 & -3.34006 \\ -12.5 & 20.5417 & 12.4746 & 20.5 & 3.34685 & -5.5 & 3.34006 & 5.48884 \\ -20.5 & 12.4746 & 20.5417 & 12.5 & 5.48884 & -3.34006 & 5.5 & 3.34685 \\ -12.4746 & 20.5 & 12.5 & 20.5417 & 3.34006 & -5.48884 & 3.34685 & 5.5 \\ -5.5 & 3.34685 & 5.48884 & 3.34006 & 20.5417 & -12.5 & 20.5 & 12.4746 \\ 3.34685 & -5.5 & -3.34006 & -5.48884 & -12.5 & 20.5417 & -12.4746 & -20.5 \\ -5.48884 & 3.34006 & 5.5 & 3.34685 & 20.5 & -12.4746 & 20.5417 & 12.5 \\ -3.34006 & 5.48884 & 3.34685 & 5.5 & 12.4746 & -20.5 & 12.5 & 20.5417 \end{pmatrix}.$$

This matrix is the quantum information projection of  $\rho$  to  $\text{GM}(\mathfrak{H}_G)$ .  $\diamond$

## 5.2. Exponential integrals and edge-bicoloured graphs

In this section, we study the following family of bivariate integrals,

$$I(z) = \frac{z}{2\pi} \int_D \exp(z g(x, y)) \, dx \, dy, \quad (5.2.1)$$

where  $D$  is a certain subset of  $\mathbb{R}^2$ ,  $g$  is a function  $D \rightarrow \mathbb{R}$  satisfying specific conditions (see Subsection 5.2.1) and  $z$  is a large positive number. Integrals as  $I(z)$  arise naturally in two important applications. First, they appear in Bayesian statistics as *marginal likelihood integrals* (see e.g. [Wat09, §1]). Second, they are *path integrals* associated to a zero-dimensional quantum system with two interacting fields parametrised by  $x$  and  $y$ , whose action is given by  $g(x, y)$  (see e.g. [Ski18, §2] or [BIZ80]). The setups might differ, however, in the integration domain  $D$ , leading to different asymptotic behaviours (see [Lin11] for an asymptotic analysis in the realm of statistics). Edge-bicoloured graphs play a classical role in Ramsey theory (see e.g. [Bol01, Ch. 12]) and their (asymptotic) enumeration is a subject with a long history (see e.g. [Wor18] and the references therein).

We will explain that the coefficients of the large- $z$  asymptotic expansion of  $I(z)$  count weighted edge-bicoloured graphs. Each graph is weighted by the reciprocal of the order of its automorphism group and the product of an arbitrary set of parameters assigned to each bicoloured incidence structure of a vertex. We do so by proving a bivariate version of the Laplace method in Subsection 5.2.1, before interpreting the coefficients of the asymptotic expansion combinatorially in Subsection 5.2.2. Therefore, we may interpret  $I(z)$  as a *generating function* of edge-bicoloured graphs (Theorem 5.2.10). From a physical perspective, these are *Feynman graphs* of the corresponding path integral. In Subsection 5.2.3, we derive an effective algorithm for the computation of those coefficients. In the final Subsection 5.2.4, we prove an asymptotic formula for the weighted number of *regular* edge-bicoloured graphs, in the limit where the number of edges and vertices goes to infinity. Our main result Theorem 5.2.17 relates this asymptotic formula to the critical points of the polynomial  $g(x, y)$  whose shape is governed by the vertex incidence structure of the graphs. We showcase that, unlike the monochromatic case (see e.g. [Bor18, Ch. 3]), only critical points satisfying some reality constraints contribute to the asymptotics.

Throughout the section we illustrate our statements through the example of the Ising model on a random 4-regular graph. The Ising model is a central object of study in mathematical physics (see e.g. [Kaz86, DC23]). The relationship between our combinatorial approach and this model is explained in Remark 5.2.12.

### 5.2.1. Laplace method and asymptotic expansions

We start by using the Laplace method to study the large- $z$  behaviour of the integral  $I(z)$  defined in (5.2.1). We require the data  $D$  and  $g : D \rightarrow \mathbb{R}$  to be chosen such that

1. the integral  $I(z)$  exists for  $z > 0$ ,
2.  $D$  is a neighbourhood of the origin,
3.  $g$  attains its unique supremum  $\sup_{(x,y) \in D} g(x, y) = g(\mathbf{0})$  at the origin,
4. near the origin,  $g$  is analytic with absolutely converging expansion

$$g(x, y) = -\frac{x^2}{2} - \frac{y^2}{2} + \sum_{\substack{u, w \geq 0 \\ u+w \geq 3}} \Lambda_{u,w} \frac{x^u y^w}{u! w!}. \quad (5.2.2)$$

The last condition ensures that (5.2.1) resembles a Gaussian integral when  $x$  and  $y$  are small. This observation allows approximating  $I(z)$  by a slightly perturbed Gaussian integral when  $z$  is large.

We define a family of polynomials  $a_{s,t}$  indexed by integers  $s, t \geq 0$  in a two-fold infinite set of variables  $\lambda_{u,w}$  indexed by  $u, w \geq 0$  with  $u + w \geq 1$ . Let  $\mathfrak{R}$  be the ring of polynomials in these variables, i.e.  $\mathfrak{R} = \mathbb{Q}[\lambda_{0,1}, \lambda_{1,0}, \lambda_{1,1}, \lambda_{0,2}, \dots]$ . The polynomials  $a_{s,t}(\lambda) \in \mathfrak{R}$  are defined by the generating function

$$\sum_{s,t \geq 0} a_{s,t}(\lambda) x^s y^t = \exp \left( \sum_{\substack{u,w \geq 0 \\ u+w \geq 1}} \lambda_{u,w} \frac{x^u y^w}{u! w!} \right) \in \mathfrak{R}[[x,y]]. \quad (5.2.3)$$

For instance,  $a_{0,0}(\lambda) = 1$ ,  $a_{1,0}(\lambda) = \lambda_{1,0}$ , and  $a_{2,0}(\lambda) = \frac{1}{2}(\lambda_{2,0} + \lambda_{1,0}^2)$ .

We will relate the *asymptotic expansion* of  $I(z)$  for large  $z$  to the polynomials  $a_{s,t}$ . For a given function  $h(z)$ , the set  $\mathcal{O}(h(z))$  consists of all functions  $f(z)$  for which

$$\limsup_{z \rightarrow \infty} |f(z)/h(z)|$$

is finite. The notation  $f(z) = g(z) + \mathcal{O}(h(z))$  means that  $f(z) - g(z) \in \mathcal{O}(h(z))$ . We use the asymptotic expansion notation  $f(z) \sim \sum_{n \geq 0} g_n(z)$  to denote  $f(z) - \sum_{n=0}^{R-1} g_n(z) \in \mathcal{O}(g_R(z))$  for all  $R \geq 0$ .

**Proposition 5.2.1.** *If  $I(z)$ ,  $g$ ,  $D$  and the coefficients  $\Lambda_{u,w}$  are related as above, then*

$$I(z) \sim \sum_{n \geq 0} A_n z^{-n},$$

for large  $z$ , where  $A_n$  is the coefficient of  $z^{-n}$  in the formal power series

$$\sum_{s,t \geq 0} z^{-(s+t)} (2s-1)!! \cdot (2t-1)!! \cdot a_{2s,2t}(z \cdot \Lambda) \in \mathbb{R}[[z^{-1}]],$$

where  $(2s-1)!! = (2s-1)(2s-3) \cdots 3 \cdot 1$  and  $a_{2s,2t}(z \cdot \Lambda)$  is the polynomial  $a_{2s,2t}(\lambda)$  defined in (5.2.3), with

$$\lambda_{u,w} = \begin{cases} 0 & \text{for } u, w \geq 0 \text{ and } 1 \leq u + w < 3, \\ z \Lambda_{u,w} & \text{for } u, w \geq 0 \text{ and } u + w \geq 3. \end{cases} \quad (5.2.4)$$

The proof of this proposition uses the classical *Laplace method* which gives an expression for the asymptotic expansion of the integral  $I(z)$ . See e.g. [BV20, App. A] for the proof of the one-dimensional case.

*Proof.* Fix an integer  $R \geq 0$  and any value for  $\gamma \in (\frac{1}{3}, \frac{1}{2})$ . We first prove that the integral  $I(z)$  is concentrated in the square  $B(z) = [-z^{-\gamma}, z^{-\gamma}]^2 \subset D$  that shrinks for growing  $z$ . Let  $M(z) = \max_{(x,y) \in D \setminus B(z)} g(x, y)$ , then

$$\begin{aligned} \left| I(z) - \frac{z}{2\pi} \int_{B(z)} \exp(zg(x, y)) dx dy \right| &= \frac{z}{2\pi} \int_{D \setminus B(z)} \exp(zg(x, y)) dx dy \\ &\leq \frac{z}{2\pi} \exp((z-1)M(z)) \int_D \exp(g(x, y)) dx dy. \end{aligned}$$

The last integral is finite by requirement. As the origin is the unique global maximum of  $g$  in  $D$ , the maximal value  $M(z)$  will be attained on the boundary of the square  $B(z)$  if  $z$  is

sufficiently large. Near the origin,  $g(x, y)$  behaves as  $-\frac{x^2}{2} - \frac{y^2}{2} + (\text{higher order terms})$ , so  $M(z) = -\frac{1}{2}z^{-2\gamma} + \mathcal{O}(z^{-3\gamma})$ . Hence,

$$I(z) = \frac{z}{2\pi} \int_{B(z)} \exp(zg(x, y)) dx dy + \mathcal{O}(z \exp(-z^{1-2\gamma})). \quad (5.2.5)$$

As  $\gamma < \frac{1}{2}$ , we have, in particular,  $\mathcal{O}(z \exp(-z^{1-2\gamma})) \subset \mathcal{O}(z^{-R})$ . So, for the purpose of finding the asymptotic expansion of  $I(z)$  in decreasing powers  $z^0, z^{-1}, z^{-2}, \dots, z^{-R+1}$ , integrating only over  $B(z)$  as in (5.2.5) is sufficient.

Note that, by (5.2.4),  $a_{s,t}(z \cdot \Lambda)$  is a polynomial of degree at most  $\frac{s+t}{3}$  in  $z$ . The function  $\exp(z(\frac{1}{2}x^2 + \frac{1}{2}y^2 + g(x, y)))$  is analytic for all  $(x, y) \in B(z)$ . Therefore, for each  $K \geq 0$ , there is a constant  $C > 0$  such that

$$\left| \exp \left( z \sum_{\substack{u,w \geq 0 \\ u+w \geq 3}} \Lambda_{u,w} \frac{x^u y^w}{u! w!} \right) - \sum_{\substack{s,t \geq 0 \\ s+t < K}} a_{s,t}(z \cdot \Lambda) x^s y^t \right| \leq C z^{\frac{1}{3}K - \gamma K} \text{ for } (x, y) \in B(z).$$

Next, we fix  $K = \frac{3R}{3\gamma-1} \geq 0$  so that  $z^{\frac{1}{3}K - \gamma K} = z^{-R}$ , and use (5.2.5) to get

$$I(z) = \frac{z}{2\pi} \sum_{\substack{s,t \geq 0 \\ s+t < K}} a_{s,t}(z \cdot \Lambda) \int_{B(z)} e^{-z \frac{x^2}{2} - z \frac{y^2}{2}} x^s y^t dx dy + \mathcal{O}(z^{-R}). \quad (5.2.6)$$

We want now to extend the integration domain to the whole real plane. For any integer  $s \geq 0$ , consider the integral

$$\int_{z^{-\gamma}}^{\infty} e^{-z \frac{x^2}{2}} x^s dx = \exp\left(-\frac{z^{1-2\gamma}}{2}\right) \int_0^{\infty} \exp\left(-z \frac{x^2}{2} - z^{1-\gamma} x\right) (z^{-\gamma} + x)^s dx.$$

For fixed  $z$ , the function  $x \mapsto \exp(-z^{1-\gamma}x)(z^{-\gamma} + x)^s$  attains its unique maximum at  $x = x_{\max} = sz^{\gamma-1} - z^{-\gamma}$ . If  $z$  is sufficiently large we have  $x_{\max} \leq 0$ . Hence, in the range we are interested in, the integral is decreasing in  $x$ , and using  $\sqrt{\frac{z}{2\pi}} \int_{\mathbb{R}} e^{-z \frac{x^2}{2}} dx = 1$  we get

$$\sqrt{\frac{z}{2\pi}} \int_{z^{-\gamma}}^{\infty} e^{-z \frac{x^2}{2}} x^s dx \in \mathcal{O}\left(z^{-\gamma s} \exp\left(-\frac{z^{1-2\gamma}}{2}\right)\right) \subset \mathcal{O}(z^{-R}).$$

Combining this with (5.2.6) shows that

$$I(z) = \frac{z}{2\pi} \sum_{\substack{s,t \geq 0 \\ s+t < K}} a_{s,t}(z \cdot \Lambda) \int_{\mathbb{R}^2} e^{-z \frac{x^2}{2} - z \frac{y^2}{2}} x^s y^t dx dy + \mathcal{O}(z^{-R}).$$

Using the Gaussian integral identities

$$\sqrt{\frac{z}{2\pi}} \int_{\mathbb{R}} e^{-z \frac{x^2}{2}} x^{2s} dx = z^{-s} \cdot (2s-1)!! \quad \text{and} \quad \int_{\mathbb{R}} e^{-z \frac{x^2}{2}} x^{2s+1} dx = 0$$

for all integers  $s \geq 0$ , proves the statement.  $\square$

**Example 5.2.2.** Fix  $D = [-1, 1]^2$  and  $g(x, y) = -\frac{x^2}{2} - \frac{y^2}{2} + \frac{x^4}{4!} + \lambda \frac{x^2 y^2}{4!} + \lambda^2 \frac{y^4}{4!}$  with  $\lambda \in \mathbb{R}_{>0}$  some arbitrary positive constant. The conditions for Proposition 5.2.1 are fulfilled and the associated integral  $I(z)$  as defined in (5.2.1) has an asymptotic expansion  $I(z) \sim$



Figure 5.2: An edge-bicoloured graph with two connected components.

$\sum_{n \geq 0} A_n z^{-n}$ . Using the formula from Proposition 5.2.1 and the generating function for the polynomials  $a_{s,t}$  from (5.2.3), we find that  $A_0 = 1$  and

$$\begin{aligned} A_1 &= \frac{1}{8} + \frac{1}{4}\lambda + \frac{1}{8}\lambda^2, \\ A_2 &= \frac{35}{384} + \frac{5}{32}\lambda + \frac{19}{64}\lambda^2 + \frac{5}{32}\lambda^3 + \frac{35}{384}\lambda^4, \\ A_3 &= \frac{385}{3072} + \frac{105}{512}\lambda + \frac{1295}{3072}\lambda^2 + \frac{175}{256}\lambda^3 + \frac{1295}{3072}\lambda^4 + \frac{105}{512}\lambda^5 + \frac{385}{3072}\lambda^6. \end{aligned} \quad \diamond$$

In the next subsection, we endow the obtained analytic expressions with a combinatorial interpretation. This process is inspired by quantum field theory, where perturbative expansions of observables, which are combinatorially controlled via *Feynman graphs*, relate to *path integrals*. The integral in (5.2.1) can be seen as a path integral for a zero-dimensional space-time: the integral is then taken over all two-parameter fields on a point, hence an integral over  $\mathbb{R}^2$ . The associated Feynman graphs are *edge-bicoloured graphs*. See [Ski18] for more details on quantum field theory.

## 5.2.2. Edge-bicoloured graphs

For the purpose of this subsection, a *graph* is a one-dimensional, finite CW complex, sometimes also called multigraph in the literature. It is *edge-bicoloured* if each edge has one of two different colours. We will represent graphs using only discrete data. A (*set*) *partition*  $P$  of a finite set  $H$  is a set of non-empty and mutually disjoint subsets of  $H$  whose union equals  $H$ . The elements of  $P$  are called *blocks*.

**Definition 5.2.3.** Given two disjoint finite sets  $S$  and  $T$  of labels, an  $[S, T]$ -labelled *edge-bicoloured graph* is a tuple  $\Gamma = (V, E_S, E_T)$ , where

1. the *vertex set*  $V$  is a partition of  $S \sqcup T$ ,
2.  $E_S$  is a partition of  $S$  into blocks of size 2,
3.  $E_T$  is a partition of  $T$  into blocks of size 2.

We think of the elements of  $S$  and  $T$  as *half-edge labels* coloured red and yellow, respectively. These half-edges are bundled together in vertices via the partition  $V$ . The edge sets  $E_S$  and  $E_T$  pair the half-edges into edges of the respective colour. Every edge-bicoloured graph without isolated vertices can be represented by at least one  $[S, T]$ -labelled graph. All graphs in this section will be edge-bicoloured, so we will drop this adjective from now on.

**Example 5.2.4.** Let  $S = \{s_1, s_2, \dots, s_6\}$  and  $T = \{t_1, t_2\}$ . The partitions

$$\begin{aligned} V &= \{\{s_1, s_2, s_3, s_4\}, \{s_5, s_6, t_1, t_2\}\}, \\ E_S &= \{\{s_1, s_2\}, \{s_3, s_4\}, \{s_5, s_6\}\}, \\ E_T &= \{\{t_1, t_2\}\}, \end{aligned}$$

form an  $[S, T]$ -labelled graph representing the graph depicted in Figure 5.2.  $\diamond$

An *isomorphism* from an  $[S_1, T_1]$ -labelled graph  $(V^1, E_S^1, E_T^1)$  to an  $[S_2, T_2]$ -labelled graph  $(V^2, E_S^2, E_T^2)$  is a pair of bijections  $j_S : S_1 \rightarrow S_2$ ,  $j_T : T_1 \rightarrow T_2$  such that  $j(V^1) = V^2$ ,  $j(E_S^1) = E_S^2$ , and  $j(E_T^1) = E_T^2$  with  $j$  being the map that  $j_S$  and  $j_T$  canonically induce on the subsets of  $S$ ,  $T$ , and  $S \sqcup T$ . An *automorphism* of an  $[S, T]$ -labelled graph  $\Gamma$  is an isomorphism to itself. Those form the group  $\text{Aut}(\Gamma)$ .

**Lemma 5.2.5.** *Each  $[\{1, \dots, 2s\}, \{1, \dots, 2t\}]$ -labelled graph  $\Gamma$  belongs to an isomorphism class of such graphs of size  $\frac{(2s)!(2t)!}{|\text{Aut}(\Gamma)|}$ .*

*Proof.* For given  $\Gamma$ , let  $\text{lab}(\Gamma)$  be the set of  $[\{1, \dots, 2s\}, \{1, \dots, 2t\}]$ -labelled graphs that are isomorphic to  $\Gamma$ . The product of symmetric groups  $\mathfrak{S}_{2s} \times \mathfrak{S}_{2t}$  acts on  $\text{lab}(\Gamma)$  by permuting the half-edge labels of the respective colour.  $\text{Aut}(\Gamma)$  is the subgroup of  $\mathfrak{S}_{2s} \times \mathfrak{S}_{2t}$  stabilising  $\Gamma$ . The lemma follows from the orbit stabiliser theorem.  $\square$

**Example 5.2.6.** An  $[S, T]$ -labelled graph  $\Gamma$  representing the graph depicted in Figure 5.2 has automorphism group isomorphic to  $(\mathfrak{S}_2 \times \mathfrak{S}_2 \rtimes \mathfrak{S}_2 \times \mathfrak{S}_2) \times \mathfrak{S}_2 \leq \mathfrak{S}_6 \times \mathfrak{S}_2$ , where  $\rtimes$  denotes the semidirect product of groups,  $\mathfrak{S}_6$  refers to the six red half-edges in  $S$  and  $\mathfrak{S}_2$  to the two yellow half-edges in  $T$ .  $\diamond$

We write  $\mathcal{G}$  for the set of isomorphism classes of graphs. For each  $G \in \mathcal{G}$ , we write  $V^G, E_S^G, E_T^G, E^G = E_S^G \sqcup E_T^G$  and  $\text{Aut}(G)$  for the respective set or group of some  $[S, T]$ -labelled representative of  $G$ . The *Euler characteristic* of  $G$  is defined by  $\chi(G) = |V^G| - |E^G|$ , and does not depend on the colouring. The *bidegree* of a graph's vertex  $v \in V^G$  is the pair of integers  $\deg(v) = (u, w)$  where  $u$  counts the number of half-edges in  $v$  that lie in the red-coloured set  $S$  and  $w$  the half-edges in the yellow-coloured part  $T$ . The *vertex degree* of  $v$  is  $|\deg(v)| = u + w$ .

**Proposition 5.2.7.** *The generating function for graphs with marked bidegrees is*

$$\sum_{G \in \mathcal{G}} \frac{\eta^{|E^G|}}{|\text{Aut}(G)|} \prod_{v \in V^G} \lambda_{\deg(v)} = \sum_{s,t \geq 0} \eta^{s+t} \cdot (2s-1)!! \cdot (2t-1)!! \cdot a_{2s,2t}(\lambda) \in \mathfrak{R}[[\eta]],$$

where  $a_{s,t}$  is defined as in (5.2.3).

We postpone the proof to after Lemma 5.2.9 and first illustrate the result.

**Example 5.2.8.** The formula in Proposition 5.2.7 provides a recipe to count our graphs for a given number of edges, grouping them according to their bidegrees. For instance, the coefficient of  $\eta^1$  counts graphs with one edge:

$$\begin{aligned} \sum_{\substack{G \in \mathcal{G}, \\ |E^G|=1}} \frac{1}{|\text{Aut}(G)|} \prod_{v \in V^G} \lambda_{\deg(v)} &= \textcolor{red}{\bullet} \textcolor{red}{\circ} + \textcolor{red}{\bullet} \textcolor{red}{\bullet} + \textcolor{yellow}{\bullet} \textcolor{yellow}{\circ} + \textcolor{yellow}{\bullet} \textcolor{yellow}{\bullet} \\ &= \frac{1}{2}\lambda_{2,0} + \frac{1}{2}\lambda_{1,0}^2 + \frac{1}{2}\lambda_{0,2} + \frac{1}{2}\lambda_{0,1}^2. \end{aligned}$$

Using the power series on the right-hand side of Proposition 5.2.7, this can be obtained simply as  $a_{2,0} + a_{0,2}$ , and by expanding the exponential in (5.2.3), we get exactly the above expression. If for  $|E^G| = 1$  these two approaches may seem equally complicated, already for graphs with two edges it is clear that the use of the generating function speeds up the computation. In fact, there are seven (monochromatic) graphs with two edges:



which turn into 23 edge-bicoloured graphs. On the other hand, a simple expansion of the exponential function gives

$$\begin{aligned}
& \sum_{\substack{G \in \mathcal{G}, \\ |E^G| = 2}} \frac{1}{|\text{Aut}(G)|} \prod_{v \in V^G} \lambda_{\deg(v)} = 3a_{4,0} + a_{2,2} + 3a_{0,4} \\
&= \lambda_{0,1}\lambda_{1,0}\lambda_{1,1} + \frac{\lambda_{0,1}^4}{8} + \frac{3\lambda_{0,1}^2\lambda_{0,2}}{4} + \frac{\lambda_{0,1}^2\lambda_{1,0}^2}{4} + \frac{\lambda_{0,1}^2\lambda_{2,0}}{4} \\
&\quad + \frac{\lambda_{0,1}\lambda_{0,3}}{2} + \frac{\lambda_{0,1}\lambda_{2,1}}{2} + \frac{3\lambda_{0,2}^2}{8} + \frac{\lambda_{0,2}\lambda_{1,0}^2}{4} + \frac{\lambda_{0,2}\lambda_{2,0}}{4} + \frac{\lambda_{0,4}}{8} + \frac{\lambda_{1,0}^4}{8} \\
&\quad + \frac{3\lambda_{1,0}^2\lambda_{2,0}}{4} + \frac{\lambda_{1,0}\lambda_{1,2}}{2} + \frac{\lambda_{1,0}\lambda_{3,0}}{2} + \frac{\lambda_{1,1}^2}{2} + \frac{3\lambda_{2,0}^2}{8} + \frac{\lambda_{2,2}}{4} + \frac{\lambda_{4,0}}{8}. \quad \diamond
\end{aligned}$$

To prove Proposition 5.2.7, we use the following lemma on the number of partitions of a set where elements come in two different colours. Let  $S = \{1, \dots, s\}$  and  $T = \{1, \dots, t\}$  and  $\mathcal{P}_{s,t}$  the set of partitions of the disjoint union  $S \sqcup T$ . For each block  $B$  of a partition  $P \in \mathcal{P}_{s,t}$  we define the *bidegree*  $\deg(B)$  of the block to be the pair of integers  $(u, w)$  where  $u$  is the number of elements from  $S$  and  $w$  the number of elements from  $T$  in  $B$ .

**Lemma 5.2.9.** *Given  $s, t \geq 0$ , consider a set of nonnegative integers  $n_{u,w}$  indexed by pairs  $u, w$  with  $0 \leq u \leq s, 0 \leq w \leq t, u + w \geq 1$ , such that*

$$\sum_u u \cdot n_{u,w} = s, \quad \sum_w w \cdot n_{u,w} = t.$$

*The number of partitions in  $\mathcal{P}_{s,t}$  with exactly  $n_{u,w}$  blocks of bidegree  $(u, w)$  is*

$$\frac{s!t!}{\prod_{u,w} n_{u,w}!(u!)^{n_{u,w}}(w!)^{n_{u,w}}}.$$

*Proof.* The group  $\mathfrak{S}_s \times \mathfrak{S}_t$  acts on  $\mathcal{P}_{s,t}$  by permuting the elements of  $S$  and  $T$ , respectively. This action is transitive if we restrict to partitions with specific block bidegree set  $\{n_{u,w}\}_{u,w}$ . A specific partition with given block bidegrees is stabilised by the subgroup that permutes the elements inside each block and blocks of the same size. This subgroup is isomorphic to  $(\mathfrak{S}_u \times \mathfrak{S}_w)^{n_{u,w}} \rtimes \mathfrak{S}_{n_{u,w}}$ . The claim follows from the orbit stabiliser theorem.  $\square$

*Proof of Proposition 5.2.7.* From (5.2.3), and  $e^X = \sum_{n \geq 0} \frac{X^n}{n!}$ , we get

$$s! \cdot t! \cdot a_{s,t}(\lambda) = \sum_{\{n_{u,w}\}} \frac{s!t!}{\prod_{u,w} n_{u,w}!(u!)^{n_{u,w}}(w!)^{n_{u,w}}} \prod_{\substack{u,w \geq 0 \\ u+w \geq 1}} \lambda_{u,w}^{n_{u,w}}, \quad (5.2.7)$$

where the sum is over all sets of integers  $\{n_{u,w}\}$  that satisfy the conditions for Lemma 5.2.9 with respect to  $s$  and  $t$ .

We can match the elements of the set  $S = \{1, \dots, 2s\}$  among each other in  $(2s - 1)!!$  ways and the ones of  $T = \{1, \dots, 2t\}$  analogously. So, by Definition 5.2.3, Lemma 5.2.9, and (5.2.7), the number of  $[S, T]$ -labelled graphs with exactly  $n_{u,w}$  vertices of bidegree  $u, w$  is  $(2s - 1)!! \cdot (2t - 1)!!$  times the coefficient of the monomial  $\prod_{u,w} \lambda_{u,w}^{n_{u,w}}$  in the polynomial  $(2s)! \cdot (2t)! \cdot a_{2s,2t}(\lambda) \in \mathfrak{R}$ . The statement follows then from Lemma 5.2.5.  $\square$

Our first main result follows from combining Propositions 5.2.1 and 5.2.7.

**Theorem 5.2.10.** *If  $I(z), g, D$  and the coefficients  $\Lambda_{u,w}$  are related as in Subsection 5.2.1, then the integral in (5.2.1) has the asymptotic expansion*

$$I(z) \sim \sum_{n \geq 0} A_n z^{-n},$$

for large  $z$ , with the coefficients  $A_n$  given by

$$A_n = \sum_{G \in \mathcal{G}_{-n}^*} \frac{1}{|\text{Aut}(G)|} \prod_{v \in V^G} \Lambda_{\deg(v)},$$

where we sum over the set  $\mathcal{G}_{-n}^*$  of all isomorphism classes of edge-bicoloured graphs with vertex degrees at least 3 and Euler characteristic equal to  $-n$ .

*Proof.* Using the fact that  $\chi(G) = |V^G| - |E^G|$  we rewrite

$$\sum_{n \geq 0} A_n z^{-n} = \sum_{G \in \mathcal{G}_{-n}^*} \frac{z^{\chi(G)}}{|\text{Aut}(G)|} \prod_{v \in V^G} \Lambda_{\deg(v)} = \sum_{G \in \mathcal{G}_{-n}^*} \frac{z^{-|E^G|}}{|\text{Aut}(G)|} \prod_{v \in V^G} z \cdot \Lambda_{\deg(v)}.$$

Applying Proposition 5.2.7 for  $\lambda_{u,w}$  as defined in (5.2.4), this is further equal to

$$\sum_{s,t \geq 0} z^{-(s+t)} \cdot (2s-1)!! \cdot (2t-1)!! \cdot a_{2s,2t}(\lambda).$$

By Proposition 5.2.1, this is the large- $z$  asymptotic expansion of  $I(z)$ .  $\square$

**Example 5.2.11.** Continuing Example 5.2.2, let  $c_n^{(k)}$  be the coefficient of  $\lambda^k$  in  $A_n$ . By Theorem 5.2.10,  $c_n^{(k)}$  counts automorphism-weighted graphs with Euler characteristic  $-n$  and vertex degree four, such that  $k_1$  vertices have exactly two yellow half-edges and  $k_2$  vertices have four yellow half-edges, so that  $k_1 + k_2 = k$ . We can view this explicitly for  $n = 2$ , as follows. Among the 21 (monochromatic) graphs with  $\chi = -2$ , there are only three 4-regular graphs. These are . Considering all bicolourings, we get

$$\begin{aligned} c_2^{(0)} &= \text{square} + \text{square with central dot} + \text{square with diagonal lines} = \frac{1}{128} + \frac{1}{48} + \frac{1}{16} = \frac{35}{384}, \\ c_2^{(1)} &= \text{square with one yellow edge} + \text{square with two yellow edges} = \frac{1}{32} + \frac{1}{8} = \frac{5}{32}, \\ c_2^{(2)} &= \text{square with one yellow edge} + \text{square with two yellow edges} + \text{square with three yellow edges} + \text{square with four yellow edges} = \frac{1}{64} + \frac{1}{32} + \frac{1}{8} + \frac{1}{16} + \frac{1}{16} = \frac{19}{64}, \\ c_2^{(3)} &= \text{square with three yellow edges} + \text{square with four yellow edges} = \frac{1}{32} + \frac{1}{8} = \frac{5}{32}, \\ c_2^{(4)} &= \text{square with four yellow edges} = \frac{1}{128} + \frac{1}{48} + \frac{1}{16} = \frac{35}{384}. \end{aligned} \quad \diamond$$

**Remark 5.2.12 (Ising model).** Our examples are motivated from the physical Ising model. The partition function of the *critical Ising model* on a specific monochromatic graph  $G$  (not necessarily lattice-like) is defined by

$$Z(G, \lambda) = \sum_{\substack{\gamma \subset G \\ \gamma \text{ Eulerian}}} \lambda^{|\mathcal{E}(\gamma)|},$$

where we sum over all Eulerian subgraphs  $\gamma$  of  $G$  (see e.g. [Cim12]). This means that if we delete all edges of  $G$  that are not in  $\gamma$ , then the resulting graph shall only have vertices of even degree. A pair  $(G, \gamma)$  of a monochromatic graph  $G$  and an Eulerian subgraph  $\gamma \subset G$  is equivalent to an edge-bicoloured graph in which an even number of yellow edges belongs to each vertex.

Notice that we effectively designed the polynomial  $g(x, y)$  from Example 5.2.2 and equivalently the coefficients  $\Lambda_{u,w}$ , such that the coefficient of  $\lambda^k$  in  $A_n$  is the automorphism-weighted number of 4-regular graphs with  $k$  yellow edges where an even number of yellow edges belong to each vertex.

Hence, with  $A_n$  as defined in Example 5.2.2, we find that

$$A_n = \sum_G \frac{Z(G, \lambda)}{|\text{Aut } G|},$$

where we sum over all monochromatic graphs  $G$  that are 4-regular and which have Euler characteristic  $-n$ . We can thus interpret  $A_n$  as the partition function of the critical Ising model of a *random* 4-regular monochromatic graph of fixed Euler characteristic. Here, random means that each monochromatic graph  $G$  is sampled with probability  $1/|\text{Aut } G|$ .

### 5.2.3. Efficient computation of the coefficients $A_n$

In this subsection, we describe an effective algorithm to compute the coefficients  $A_n$  that encode the asymptotic expansion of the integral (5.2.1), and the weighted numbers of edge-bicoloured graphs of Euler characteristic  $-n$ , by Theorem 5.2.10. The algorithm is implemented in Julia and is available at [KLW24b].

**Proposition 5.2.13.** *For a given integer  $n \geq 1$ , and the coefficients  $\Lambda_{u,w}$  as required by Theorem 5.2.10, the following algorithm correctly computes  $A_0, \dots, A_n$ :*

Step 1: Define the polynomials

$$F_k(x, y) = \sum_{\substack{u, w \geq 0 \\ u+w=k+2}} \Lambda_{u,w} \frac{x^u y^w}{u! w!} \text{ for } k \in \{1, \dots, 2n\};$$

Step 2: set  $Q_0(x, y) = 1$  and recursively compute  $Q_1, \dots, Q_{2n}$  using

$$Q_m(x, y) = \frac{1}{m} \sum_{k=1}^m k F_k(x, y) Q_{m-k}(x, y) \text{ for } m \in \{1, \dots, 2n\};$$

Step 3: let  $q_{s,t}^{(k)}$  be the coefficients of  $Q_k(x, y) = \sum_{s,t \geq 0} q_{s,t}^{(k)} x^s y^t$ . Then,

$$A_k = \sum_{s,t \geq 0} (2s-1)!! \cdot (2t-1)!! \cdot q_{2s,2t}^{(2k)} \text{ for } k \in \{0, \dots, n\}.$$

To run the algorithm with a fixed  $n$ , it is sufficient to know  $\Lambda_{u,w}$  for all  $u, w \geq 0$  with  $u + w \leq 2n + 2$ . Also, recall that we require  $\Lambda_{u,w} = 0$  if  $u + w < 3$ .

*Proof.* By Proposition 5.2.1, we have this identity of power series in  $z^{-n}$ :

$$\sum_{n \geq 0} A_n z^{-n} = \sum_{s,t \geq 0} z^{-(s+t)} \cdot (2s-1)!! \cdot (2t-1)!! \cdot a_{2s,2t}(z \cdot \Lambda),$$

where  $a_{s,t}(z \cdot \Lambda)$  is as described in Proposition 5.2.1 and (5.2.3):

$$\sum_{s,t \geq 0} a_{s,t}(z \cdot \Lambda) x^s y^t = \exp \left( z \sum_{\substack{u, w \geq 0 \\ u+w \geq 3}} \Lambda_{u,w} \frac{x^u y^w}{u! w!} \right).$$

Rescaling  $(x, y) \mapsto (x/\sqrt{z}, y/\sqrt{z})$  in the above formula gives the following identity of power series in  $\mathbb{R}[x, y][[z^{-1/2}]]$ ,

$$\sum_{s,t \geq 0} z^{-\frac{s+t}{2}} a_{s,t}(z \cdot \Lambda) x^s y^t = \exp \left( \sum_{k \geq 1} z^{-\frac{k}{2}} F_k(x, y) \right), \quad (5.2.8)$$

where we used the definition of  $F_k(x, y)$  in the statement. Let  $q_{s,t}^{(k)}$  be the coefficients  $\sum_{k \geq 0} q_{s,t}^{(k)} z^{-\frac{k}{2}} = z^{-\frac{s+t}{2}} a_{s,t}(z \cdot \Lambda)$ . With this definition, (5.2.8) and the formula under Step 3 in the statement correctly compute  $A_k$ .

It remains to prove that the coefficients  $q_{s,t}^{(k)}$  are computed correctly by Step 2 in the statement. Rewrite (5.2.8) using the definition of  $Q_k(x, y)$ , before applying the derivative operator  $z \frac{\partial}{\partial z}$  on both sides. This gives

$$\begin{aligned} z \frac{\partial}{\partial z} \left( \sum_{k \geq 0} Q_k(x, y) z^{-\frac{k}{2}} \right) &= z \frac{\partial}{\partial z} \exp \left( \sum_{k \geq 1} z^{-\frac{k}{2}} F_k(x, y) \right) \\ &\Rightarrow - \sum_{k \geq 0} \frac{k}{2} Q_k(x, y) z^{-\frac{k}{2}} = - \left( \sum_{m \geq 0} Q_m(x, y) z^{-\frac{m}{2}} \right) \sum_{k \geq 1} \frac{k}{2} z^{-\frac{k}{2}} F_k(x, y). \end{aligned}$$

The recursive relation between  $Q_m$  and  $F_k$  follows by comparing the  $z^{-\frac{m}{2}}$  coefficients on both sides of this equation.  $\square$

## 5.2.4. Asymptotics and critical points

In this subsection, we study the asymptotic behaviour of the coefficients  $A_n$  in Theorem 5.2.10 for large  $n$ . Here, we will restrict ourselves to *regular* edge-bicoloured graphs, meaning that each vertex has a fixed degree  $k \geq 3$ . For fixed coefficients  $\Lambda_{u,w}$  given for  $u, w \geq 0$  with  $u + w = k$ , we study the weighted sum over graphs

$$A_n = \sum_{G \in \mathcal{G}_{-n}^k} \frac{1}{|\text{Aut}(G)|} \prod_{v \in V^G} \Lambda_{\deg(v)},$$

where  $\mathcal{G}_{-n}^k$  is the set of all regular (edge-bicoloured) graphs with vertex degree  $k$  and Euler characteristic  $-n$ . As for each  $k$ -regular graph  $G$  we have  $k|V^G| = 2|E^G|$ , all graphs in  $\mathcal{G}_{-n}^k$  have  $\frac{2n}{k-2}$  vertices and  $\frac{nk}{k-2}$  edges. It is convenient to define the homogeneous polynomial

$$V(x, y) = g(x, y) + \frac{x^2}{2} + \frac{y^2}{2} = \sum_{\substack{u, w \geq 0 \\ u+w=k}} \Lambda_{u,w} \frac{x^u y^w}{u! w!} \in \mathbb{R}[x, y].$$

Let  $\Phi$  be the set of global maxima of the function

$$S^1 = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 = 1\} \rightarrow \mathbb{R}_{\geq 0}, \quad (x, y) \mapsto |V(x, y)|.$$

A point  $(x, y) \in \Phi$  is *nondegenerate* if  $k^2 V(x, y) \neq \left( \frac{\partial^2 V}{\partial x^2}(x, y) + \frac{\partial^2 V}{\partial y^2}(x, y) \right)$ .

**Proposition 5.2.14.** *Let  $M = \frac{k}{k-2}$  and  $K = \frac{2}{k-2}$ . If  $A_n$ ,  $V$ ,  $\Lambda_{u,w}$  and  $\Phi$  are related as described above and all extrema in  $\Phi$  are nondegenerate, then*

$$A_n \sim \begin{cases} \frac{1}{2\sqrt{2}\pi} k^{nM+\frac{1}{2}} K^{n-\frac{1}{2}} \Gamma(n) \sum_{(x,y) \in \Phi} \frac{V(x,y)^{nK}}{\sqrt{B(x,y)}} & \text{if } nK, nM \in \mathbb{Z}, \\ 0 & \text{else,} \end{cases}$$

where  $\Gamma$  denotes the Gamma function and

$$B(x, y) = k^2 - \frac{\frac{\partial^2 V}{\partial x^2}(x, y) + \frac{\partial^2 V}{\partial y^2}(x, y)}{V(x, y)} \quad \text{for } (x, y) \in S^1.$$

We will prove this theorem by first proving an integral representation of the coefficients  $A_n$ . Afterwards, we apply the one-dimensional Laplace method to provide an asymptotic expression for this integral in the large- $n$  limit.

**Lemma 5.2.15.** *Let  $M = \frac{k}{k-2}$  and  $K = \frac{2}{k-2}$ . For a given integer  $n \geq 0$  such that  $nK$  and  $nM$  are integers, we have*

$$A_n = \frac{2^{nM}(nM)!}{2\pi \cdot (nK)!} \int_{-\pi}^{\pi} V(\cos \varphi, \sin \varphi)^{nK} d\varphi.$$

If  $nK$  or  $nM$  is not an integer then  $A_n = 0$ .

*Proof.* A  $k$ -regular graph has  $nM$  edges and  $nK$  vertices, so  $nM$  and  $nK$  must be integers; otherwise  $A_n = 0$ . We will assume the former. By Proposition 5.2.7,

$$A_n = \sum_{\substack{s, t \geq 0 \\ s+t=nM}} (2s-1)!! \cdot (2t-1)!! \cdot a_{2s, 2t}(\Lambda).$$

If  $s+t = nM$ , it follows from (5.2.3) that  $a_{2s, 2t}(\Lambda)$  is a homogeneous polynomial of degree  $nK$ . Because  $\exp(X) = \sum_{N \geq 0} \frac{X^N}{N!}$ , it also follows that  $a_{2s, 2t}(\Lambda)$  is the coefficient in front of  $x^{2s}y^{2t}$  in the quotient  $V(x, y)^{nK}/(nK)!$ , since  $V$  is homogeneous. Using  $\frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-\frac{x^2}{2}} x^{2s} dx = (2s-1)!!$  and  $\int_{\mathbb{R}} e^{-\frac{x^2}{2}} x^{2s+1} dx = 0$  for integers  $s$ , we obtain

$$A_n = \frac{1}{2\pi \cdot (nK)!} \int_{\mathbb{R}^2} e^{-\frac{x^2}{2} - \frac{y^2}{2}} V(x, y)^{nK} dx dy.$$

We can pass to polar coordinates and use  $V(rx, ry) = r^k V(x, y)$ , together with

$$\int_0^\infty e^{-\frac{r^2}{2}} r^{nkK+1} dr = \int_0^\infty e^{-q} (2q)^{nkK/2} dq = 2^{nM}(nM)!$$

to prove the lemma.  $\square$

*Proof of Proposition 5.2.14.* We are interested in the cases in which  $A_n \neq 0$ . When  $n$  is large, the main contribution to the integral in the statement of Lemma 5.2.15 comes from angles  $\varphi$  where  $|V(\cos \varphi, \sin \varphi)|$  is maximal. Let  $\varphi_c$  be the location of such a maximum. By definition, we have  $(\cos \varphi_c, \sin \varphi_c) \in \Phi$ . Near this maximum, we get the Taylor expansion

$$f_{\varphi_c}(\varphi) := \log \frac{V(\cos \varphi, \sin \varphi)}{V(\cos \varphi_c, \sin \varphi_c)} = -B(\cos \varphi_c, \sin \varphi_c) \frac{(\varphi - \varphi_c)^2}{2} + \mathcal{O}((\varphi - \varphi_c)^3),$$

where  $B(\cos \varphi_c, \sin \varphi_c)$  is defined as in the statement. Because  $\varphi_c$  is a maximum of  $|V(\cos \varphi_c, \sin \varphi_c)|$ , we have  $B(\cos \varphi_c, \sin \varphi_c) \geq 0$ . Our assumption that all the maxima are nondegenerate hence implies that  $B(\cos \varphi_c, \sin \varphi_c) > 0$ .

We may therefore write, for some sufficiently small  $\epsilon > 0$ ,

$$A_n = \frac{2^{nM}(nM)!}{2\pi \cdot (nK)!} \left( \sum_{(\cos \varphi_c, \sin \varphi_c) \in \Phi} V(\cos \varphi_c, \sin \varphi_c)^{nK} \int_{\varphi_c-\epsilon}^{\varphi_c+\epsilon} e^{nK f_{\varphi_c}(\varphi)} d\varphi \right) + R_1(n, \epsilon).$$

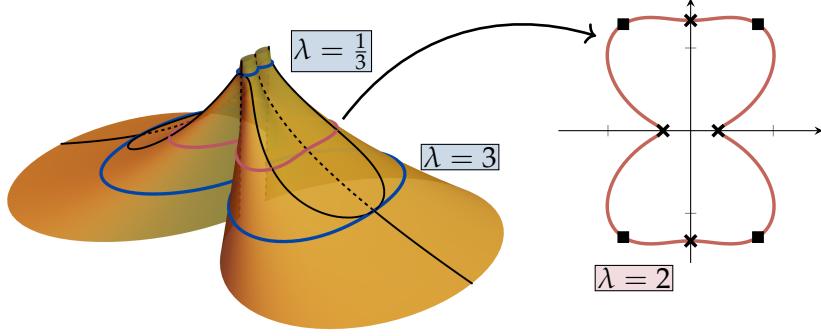


Figure 5.3: The system (5.2.9) for the function  $V$  from Example 5.2.16. Left: all values of  $\lambda \in (0, 4)$  on the (reversed) vertical axis. At each level  $\lambda = \text{const.}$  the black continuous curves are the maxima; the dashed curves are the minima. In blue, the curves for  $\lambda = \frac{1}{3}$  and  $\lambda = 3$  where the behaviour of the maxima changes. Right: the section  $\lambda = 2$ , with its maxima (squares) and its minima (crosses).

From the Taylor expansion of the function  $f_{\varphi_c}(\varphi)$  and Lemma 5.2.15, it follows by the same reasoning as in the proof of Proposition 5.2.1 that the remainder term respects the bound  $|R_1(n, \varepsilon)| \leq C_1 \exp(-C_2 n \varepsilon^2)$  with some constants  $C_1, C_2 > 0$ . Specifying  $\varepsilon = n^{-\gamma}$  with  $\gamma \in (\frac{1}{3}, \frac{1}{2})$  allows us to truncate the Taylor expansion of  $f_{\varphi_c}$  after the second term without changing asymptotic behaviour in the  $n \rightarrow \infty$  limit. Hence,

$$\int_{\varphi_c - \varepsilon}^{\varphi_c + \varepsilon} e^{nKf_{\varphi_c}(\varphi)} d\varphi = \int_{-\varepsilon}^{\varepsilon} \exp\left(-nK \frac{B(\cos \varphi_c, \sin \varphi_c)}{2} \varphi^2\right) d\varphi + R_2(n, \varepsilon).$$

The remainder term satisfies  $|R_2(n, \varepsilon)| < C_3 n^{\frac{1}{2}} \varepsilon^3$  for some  $C_3 > 0$ . Again, as in the proof of Proposition 5.2.1, we may complete the Gaussian integral to find that

$$\int_{\varphi_c - \varepsilon}^{\varphi_c + \varepsilon} e^{nKf_{\varphi_c}(\varphi)} d\varphi = \sqrt{\frac{2\pi}{nK \cdot B(\cos \varphi_c, \sin \varphi_c)}} + \mathcal{O}(n^{-1}).$$

The result follows from Stirling's formula  $\Gamma(n) \sim \sqrt{2\pi n^{-1}} n^n e^{-n}$  as  $n \rightarrow \infty$ .  $\square$

**Example 5.2.16.** To continue the running example of the Ising model (cf. Example 5.2.11), let  $V(x, y) = \frac{x^4}{4!} + \lambda \frac{x^2 y^2}{4} + \lambda^2 \frac{y^4}{4!}$ . We want to find the critical points of  $V$  on the circle, that means the points  $(x, y) \in \mathbb{R}^2$  with  $x^2 + y^2 = 1$  satisfying

$$y \frac{\partial V}{\partial x}(x, y) = x \frac{\partial V}{\partial y}(x, y). \quad (5.2.9)$$

We get the following eight critical points:

$$(\pm 1, 0), (0, \pm 1), \left( \pm \frac{\sqrt{\lambda(\lambda-3)}}{\sqrt{\lambda^2 - 6\lambda + 1}}, \pm \frac{\sqrt{1-3\lambda}}{\sqrt{\lambda^2 - 6\lambda + 1}} \right) \in \mathbb{R}^2. \quad (5.2.10)$$

Our case of interest is  $\lambda > 0$ . Then, the last four points are real if and only if  $\lambda \in [\frac{1}{3}, 3]$ , and in that interval those are the maxima of  $|V|$  on  $S^1$ . For  $\lambda < \frac{1}{3}$ , the maxima are  $(\pm 1, 0)$ , whereas for  $\lambda > 3$ , the maxima are  $(0, \pm 1)$ . Figure 5.3 displays the function  $(V(x, y)x, V(x, y)y)$  and its critical points, for  $\lambda \in (0, 4)$ .

We can now use Proposition 5.2.14 to find  $A_n \sim c \Gamma(n) \alpha^n$ , where  $c = c(\lambda)$  and  $\alpha = \alpha(\lambda)$  are piecewise defined as

|                             | $\alpha(\lambda)$                             | $c(\lambda)$                                                    |
|-----------------------------|-----------------------------------------------|-----------------------------------------------------------------|
| $0 < \lambda < \frac{1}{3}$ | $\frac{2}{3}$                                 | $\frac{1}{\pi} \sqrt{\frac{1}{2-6\lambda}}$                     |
| $\frac{1}{3} < \lambda < 3$ | $\frac{-16\lambda^2}{3\lambda^2-18\lambda+3}$ | $\frac{1}{\pi} \sqrt{\frac{8\lambda}{-3\lambda^2+10\lambda-3}}$ |
| $\lambda > 3$               | $\frac{2\lambda^2}{3}$                        | $\frac{1}{\pi} \sqrt{\frac{\lambda}{2\lambda-6}}$               |

The function  $\alpha$  is continuous, it is not  $C^1$ -differentiable at  $\lambda = \frac{1}{3}$ , and it is  $C^1$ - but not  $C^2$ -differentiable at  $\lambda = 3$ . On the other hand, the limits of  $c(\lambda)$  at  $\frac{1}{3}$  and 3 go to infinity from both sides. This can be observed in Figure 5.4. The points  $\lambda = \frac{1}{3}$  and  $\lambda = 3$  where the functions  $\alpha(\lambda)$  and  $c(\lambda)$  are non-analytic are *phase transition points*. Phase transitions are of pivotal interest in statistical physics. Here, we find the phase transitions of the Ising model on a random 4-regular graph. In each of the three regions for the parameter  $\lambda$ , the statistical system is expected to exhibit intrinsically different behaviours.

Note that using Proposition 5.2.13 we can also compute  $A_n$  for large  $n$  and solve for  $\alpha$  and  $c$  numerically. For details see our implementation at [KLW24b].  $\diamond$

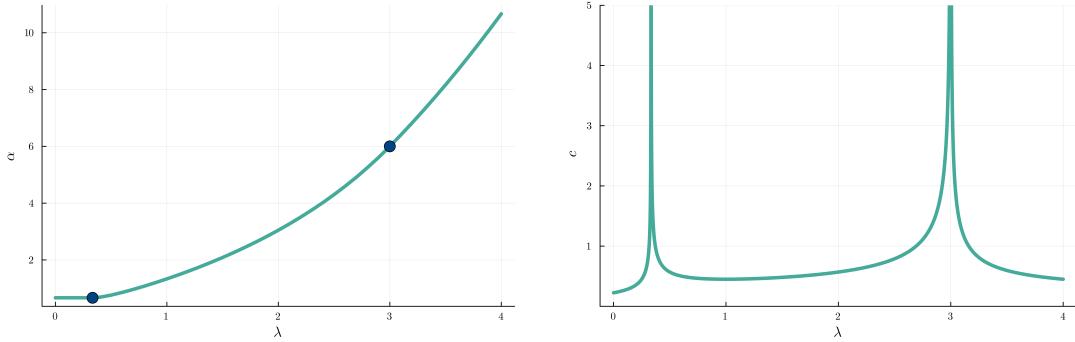


Figure 5.4: The behaviour of  $\alpha(\lambda)$  and  $c(\lambda)$  in the Ising model from Example 5.2.16. The phase transitions at  $\lambda = \frac{1}{3}, 3$  can be detected in both quantities. At  $\lambda = 3$ , the function  $\alpha$  is  $C^1$ - but not  $C^2$ -differentiable.

It is common belief in physics that the asymptotic behaviour of  $A_n$  depends on the critical points of  $g(x, y) = -\frac{x^2}{2} - \frac{y^2}{2} + V(x, y)$  (see e.g. [LGZJ12]). Moreover, it is well-known, also in applied mathematics, that identifying the critical point which contributes most to the asymptotics is a complicated *connection problem* [BH90]. Therefore, we rephrase Proposition 5.2.14 in terms of the critical points of  $g$  instead of those of  $V$  restricted to the sphere. We write  $\text{crit}_D f$  for the set of critical points of  $f$  restricted to the domain  $D$ . Let

$$\Psi = \{(w, z) \in \text{crit}_{\mathbb{C} \cdot \mathbb{R}^2} g \setminus \{\mathbf{0}\} : \|(w, z)\| \leq \|(w', z')\| \quad \forall (w', z') \in \text{crit}_{\mathbb{C} \cdot \mathbb{R}^2} g \setminus \{\mathbf{0}\}\},$$

where  $\mathbb{C} \cdot \mathbb{R}^2$  is the set of complex points  $(w, z)$  whose ratio (if well-defined) is real. Points in  $\Psi$  are called *nondegenerate* if the Hessian matrix  $H_g(w, z)$  of  $g$  at  $(w, z)$  has full rank.

**Theorem 5.2.17.** *Assume that  $A_n$ ,  $g$ , and  $\Psi$  are related as described above and all extrema in  $\Psi$  are nondegenerate. Then*

$$A_n \sim \frac{1}{2\pi} \Gamma(n) \sum_{(w,z) \in \Psi} \frac{(-g(w, z))^{-n}}{\sqrt{-\det H_g(w, z)}}. \quad (5.2.11)$$

*Proof.* Our goal is to express  $A_n$  from Proposition 5.2.14 in terms of the critical points of  $g$ . The first step is to associate the critical points of  $V$  to those of  $g$ . Given  $(x, y) \in \text{crit}_{S^1}(V)$ , we look for some  $\ell \in \mathbb{C}^*$  such that  $(\ell x, \ell y) \in \text{crit}_{\mathbb{C} \cdot \mathbb{R}^2}(g)$ . Imposing the conditions  $\ell x = \frac{\partial V}{\partial x}(\ell x, \ell y)$ ,  $\ell y = \frac{\partial V}{\partial y}(\ell x, \ell y)$ , and using homogeneity of  $V$ , we get

$$\ell^{2-k} = kV(x, y). \quad (5.2.12)$$

Therefore, as  $k \geq 3$ ,

$$\max_{(x,y) \in \text{crit}_{S^1}(V)} V(x, y) = \frac{1}{k} \left( \min_{(w,z) \in \text{crit}_{\mathbb{C} \cdot \mathbb{R}^2}(g)} \| (w, z) \| \right)^{2-k},$$

so every element  $(w, z) \in \Psi \subset \mathbb{C} \cdot \mathbb{R}^2$  arises as  $(\ell x, \ell y)$  for some  $(x, y) \in \Phi$ .

Using these considerations, we write the result from Proposition 5.2.14 in terms of the critical points of  $g$ . At a point  $(w, z) = (\ell x, \ell y) \in \Psi$ , by (5.2.12), we have

$$g(w, z) = -\frac{\ell^2}{2} + V(w, z) = -\frac{\ell^2}{2} + \ell^k V(x, y) = \ell^2 \frac{2-k}{2k}. \quad (5.2.13)$$

Let  $K = \frac{2}{k-2}$  and  $M = \frac{k}{k-2}$ . Then, we have

$$V(x, y)^{nK} = k^{-nK} \ell^{-2n} = k^{-nK} (-k K g(w, z))^{-n} = k^{-nM} K^{-n} (-g(w, z))^{-n}.$$

This allows to cancel prefactors in the asymptotic expression for  $A_n$  from Proposition 5.2.14. We are left to rewrite  $B$  in terms of  $(w, z)$ . Notice that the determinant of the Hessian of  $g(w, z)$  can be expressed, using (5.2.12), as

$$\begin{aligned} \det H_g(w, z) &= \frac{1}{\ell^2} \det H_g(\ell x, \ell y) = B(x, y) \ell^{k-2} V(x, y) (1 - k(k-1) \ell^{k-2} V(x, y)) \\ &= -B(x, y) \frac{k-2}{k}, \end{aligned}$$

where  $(\ell, x, y)$  are new coordinates on  $\mathbb{C}^* \times S^1$ , and  $(w, z) \in \Psi$ . Hence,

$$\begin{aligned} A_n &\sim \frac{1}{2\sqrt{2}\pi} k^{nM+\frac{1}{2}} K^{n-\frac{1}{2}} \Gamma(n) \sum_{(x,y) \in \Phi} \frac{V(x, y)^{nK}}{\sqrt{B(x, y)}} \\ &= \frac{1}{4\pi} \sqrt{k(k-2)} \Gamma(n) \frac{2}{k-2} \sum_{(w,z) \in \Psi} \frac{(-g(w, z))^{-n}}{\sqrt{-\frac{k \det H_g(w, z)}{k-2}}} \\ &= \frac{1}{2\pi} \Gamma(n) \sum_{(w,z) \in \Psi} \frac{(-g(w, z))^{-n}}{\sqrt{-\det H_g(w, z)}}, \end{aligned}$$

where the factor  $\frac{2}{k-2}$  appears since each of the  $k-2$  points  $\{(w, z) = (\ell x, \ell y)\}$  in  $\Psi$  is counted twice by the corresponding points  $\{(x, y), (-x, -y)\} \in \Phi$ .  $\square$

**Remark 5.2.18.** The formula (5.2.11) yields 0 if  $nM$  or  $nK$  are not integers. Indeed, using (5.2.13) from the proof above, we can write, for  $(w, z) \in \Psi$ ,

$$g(w, z) = (l \cdot \zeta_i)^2 \frac{2-k}{2k}, \quad i \in \{1, \dots, k-2\},$$

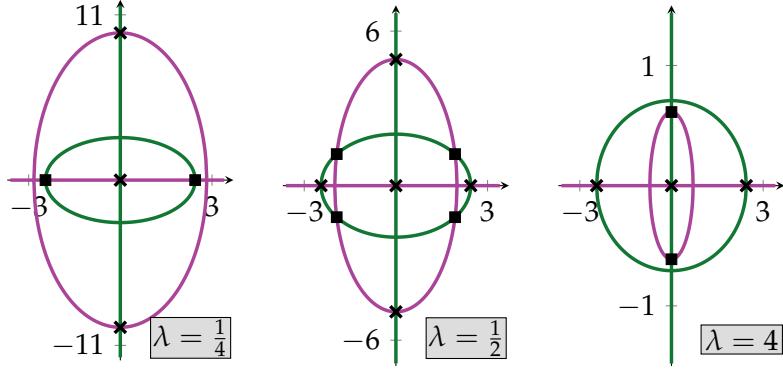


Figure 5.5: The system (5.2.14) for the function  $V$  from Example 5.2.19, for the values  $\lambda = \frac{1}{2}, \frac{1}{4}, 4$ , from left to right. The solutions are marked in black. The solutions that are (equally) closest to (but distinct from) the origin, are marked with squares. Notice the different scaling in the  $y$ -axis, for the sake of clarity.

where  $l \in \mathbb{R}$  and  $\zeta_i$  is a  $(k-2)$ th root of unity, so  $(w, z) = (l\zeta_i x, l\zeta_i y)$  for some  $(x, y) \in \Phi$ . Also,  $(l\zeta_j x, l\zeta_j y) \in \Psi$  for all  $j \in \{1, \dots, k-2\}$ . Therefore, the sum in (5.2.11) becomes

$$\begin{aligned} \sum_{(w,z) \in \Psi} \frac{(-g(w,z))^{-n}}{\sqrt{-\det H_g(w,z)}} &\propto \sum_{j=1}^{k-2} \frac{(l \cdot \zeta_j)^{-2n} \left(\frac{k-2}{k}\right)^{-n}}{\sqrt{-\det H_g(l\zeta_j x, l\zeta_j y)}} \\ &= \begin{cases} (k-2) \frac{l^{-2n} \left(\frac{k-2}{k}\right)^{-n}}{\sqrt{-\det H_g(lx, ly)}} & \text{if } (k-2) \mid 2n, \\ 0 & \text{else.} \end{cases} \end{aligned}$$

The condition  $(k-2) \mid 2n$  is equivalent to  $nK \in \mathbb{Z}$ , which also implies  $nM \in \mathbb{Z}$ .

We exhibit the connection between the two collections  $\Phi$  and  $\Psi$  of critical points explicitly in our running example.

**Example 5.2.19.** Let  $V(x, y) = \frac{x^4}{4!} + \lambda \frac{x^2 y^2}{4!} + \lambda^2 \frac{y^4}{4!}$  and  $g(x, y) = -\frac{x^2}{2} - \frac{y^2}{2} + V(x, y)$ . Consider the system of critical equations for  $g$

$$w = \frac{\partial V}{\partial w}(w, z), \quad z = \frac{\partial V}{\partial z}(w, z), \quad (5.2.14)$$

and its complex nontrivial solutions, for  $\lambda > 0$ :

$$(\pm\sqrt{6}, 0), \left(0, \pm\frac{\sqrt{6}}{\lambda}\right), \left(\pm\sqrt{\frac{9-3\lambda}{4\lambda}}, \pm\frac{\sqrt{9\lambda-3}}{2\lambda}\right).$$

Among these solutions, some are real for every  $\lambda > 0$ . The last type of singular points is real if and only if  $\lambda \in [\frac{1}{3}, 3]$ . We get

$$\Psi = \begin{cases} (\pm\sqrt{6}, 0) & 0 < \lambda < \frac{1}{3}, \\ \left(\pm\sqrt{\frac{9-3\lambda}{4\lambda}}, \pm\frac{\sqrt{9\lambda-3}}{2\lambda}\right) & \frac{1}{3} < \lambda < 3, \\ \left(0, \pm\frac{\sqrt{6}}{\lambda}\right) & \lambda > 3. \end{cases}$$

This is displayed in Figure 5.5. The reader may check that rescaling each point in  $\Psi$  to unit vector gives precisely two of the points in (5.2.10).  $\diamond$

**Remark 5.2.20.** Lee–Yang theory studies the location of the roots of the polynomials  $A_n$ , when  $n$  becomes large. This fascinating theory touches combinatorics, statistics and physics (see e.g. [BB09] for an overview). In the spirit of Lee–Yang theory, the two phase transitions  $\lambda = \frac{1}{3}, 3$  in the running example can be detected also by looking at the asymptotic behaviour of the roots of  $A_n(\lambda)$  as  $n \rightarrow \infty$ . Using our algorithm from Proposition 5.2.13, we can compute the polynomials  $A_n(\lambda)$  and find their roots numerically. This is the content of Figure 5.6. The roots of these polynomials are all complex (except for  $\lambda = -1$ , for odd  $n$ ) but they get closer and closer to the real values  $\lambda = \frac{1}{3}$  and  $\lambda = 3$ .

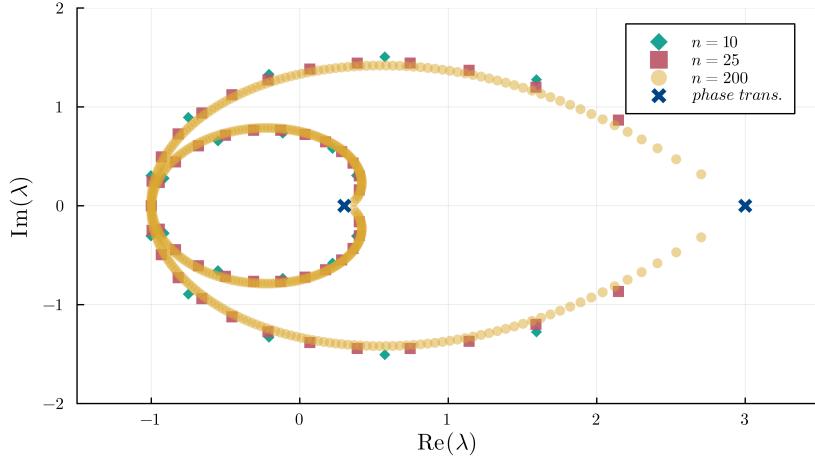


Figure 5.6: The roots of  $A_n(\lambda)$  under the assumptions of Example 5.2.2, for  $n = 10, 25, 200$ . The blue crosses are the phase transitions  $\lambda = \frac{1}{3}, 3$ .

Although Proposition 5.2.14 and Theorem 5.2.17 assume  $V$  to be homogeneous, the following example shows that this condition does not seem to be necessary.

**Example 5.2.21.** Take the inhomogeneous polynomial  $V(x, y) = \frac{x^3}{3!} + \lambda \frac{xy^2}{2} + \lambda^2 \frac{y^4}{4!}$ , with  $\lambda > 0$ . For  $\lambda < \frac{1}{2}$ , one can compute that  $\Psi = \{(2, 0)\}$ ; for  $\lambda > \frac{1}{2}$ , one gets

$$\Psi = \left\{ \left( \frac{4\lambda - \sqrt{2\lambda(8\lambda-3)}}{\lambda}, \pm \sqrt{6} \sqrt{\frac{1-(4\lambda - \sqrt{2\lambda(8\lambda-3)})}{\lambda^2}} \right) \right\}.$$

The formula for  $A_n$  from Theorem 5.2.17 would give

|                             | $\alpha(\lambda)$                                                           | $c(\lambda)$                                                                                             |
|-----------------------------|-----------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------|
| $0 < \lambda < \frac{1}{2}$ | $\frac{3}{2}$                                                               | $\frac{1}{2\pi\sqrt{1-2\lambda}}$                                                                        |
| $\lambda > \frac{1}{2}$     | $\frac{6\lambda^2}{(8\lambda-3)(16\lambda-3-4\sqrt{2\lambda(8\lambda-3)})}$ | $\frac{1}{\pi} \sqrt{\frac{\lambda}{32\lambda^2-12\lambda+2\sqrt{2\lambda(8\lambda-3)(1-16\lambda^2)}}}$ |

This matches our numerical computations, see [KLW24b].  $\diamond$

Based on the previous example and similar computations, we conjecture that Theorem 5.2.17 is also valid for inhomogeneous  $V(x, y)$ , i.e. graphs that are not necessarily regular. In this setting, the univariate Laplace method as used in the proof of Proposition 5.2.14 does not work any more, also due to the failure of Lemma 5.2.15. Instead, a multivariate saddle point method shall be required.

**Conjecture 5.2.22.** Let  $g$  and  $A_n$  be related as in the beginning of Subsection 5.2.1 (i.e.  $g(x, y) + \frac{x^2}{2} + \frac{y^2}{2}$  is not necessarily homogeneous). Let  $\Psi$  be defined as before and assume that all points in  $\Psi$  are nondegenerate. Then,

$$A_n \sim \frac{1}{2\pi} \Gamma(n) \sum_{(w,z) \in \Psi} \frac{(-g(w,z))^{-n}}{\sqrt{-\det H_g(w,z)}} \quad \text{as } n \rightarrow \infty.$$

## 5.3. Conclusion

In this chapter we have explored several connections between nonlinear algebra and quantum physics. In the context of quantum information theory, we have seen how we can associate algebraic varieties to quantum graphical models, providing a new approach to studying the structure of quantum states satisfying constraints on the mutual information between subsystems or their locality. The different approaches to construct such algebraic varieties presented in Section 5.1 yield different varieties; it is yet to be seen which one of them is most useful from the physics perspective. Moreover, all of these approaches pose very challenging computational problems. It would be interesting to try applying recently improved algorithms for numerical implicitisation [CH23] to compute algebraic varieties associated to quantum graphical models with at least four nodes. Our algebraic perspective proves useful in analysing quantum information projections to quantum exponential families (Theorem 5.1.34). It remains an open problem to extend this result to more general families of Hamiltonians giving rise to quantum exponential families.

In the very last section we have seen how techniques from perturbative quantum field theory inspire enumeration methods of edge-bicoloured graphs. Besides Conjecture 5.2.22 generalising our main result about the asymptotic number of regular edge-bicoloured graphs to nonregular graphs, there is another open problem for future research: generalising 5.2.17 to graphs with more than two edge colours. The latter problem causes challenges because critical loci might be higher dimensional. This analysis then leads to Picard–Lefschetz theory [DH02].

# Bibliography

- [AC06] Abdelmalek Abdesselam and Jaydeep Chipalkatti. The bipartite Brill–Gordan locus and angular momentum. *Transformation groups*, 11(3):341–370, 2006.
- [AKMM23] Takuro Abe, Lukas Kühne, Paul Mücksch, and Leonie Mühlherr. Projective dimension of weakly chordal graphic arrangements. [arXiv:2307.06021](https://arxiv.org/abs/2307.06021), 2023. To appear in Algebraic Combinatorics.
- [ABF<sup>+</sup>23] Daniele Agostini, Taylor Brysiewicz, Claudia Fevola, Lukas Kühne, Bernd Sturmfels, Simon Telen, and Thomas Lam. Likelihood degenerations. *Advances in Mathematics*, 414:108863, 2023.
- [AFST24] Daniele Agostini, Claudia Fevola, Anna-Laura Sattelberger, and Simon Telen. Vector spaces of generalized Euler integrals. *Communications in Number Theory and Physics*, 18(2):327–370, 2024.
- [AH95] James Alexander and André Hirschowitz. Polynomial interpolation in several variables. *Journal of Algebraic Geometry*, 4(4):201–222, 1995.
- [Alh23] Álvaro M Alhambra. Quantum many-body systems in thermal equilibrium. *PRX Quantum*, 4(4):040201, Nov 2023.
- [ABB<sup>+</sup>19] Carlos Améndola, Nathan Bliss, Isaac Burke, Courtney R Gibbons, Martin Helmer, Serkan Hoşten, Evan D Nash, Jose Israel Rodriguez, and Daniel Smolkin. The maximum likelihood degree of toric varieties. *Journal of Symbolic Computation*, 92:222–242, 2019.
- [AKK20] Carlos Améndola, Dimitra Kosta, and Kaie Kubjas. Maximum likelihood estimation of toric fano varieties. *Algebraic Statistics*, 11(1):5–30, 2020.
- [AO24] Carlos Améndola and Janike Oldekop. Likelihood geometry of reflexive polytopes. *Algebraic Statistics*, 15(1):113–143, 2024.
- [ACGH19] Sanjeev Arora, Nadav Cohen, Noah Golowich, and Wei Hu. A convergence analysis of gradient descent for deep linear neural networks. In *7th International Conference on Learning Representations, ICLR 2019, New Orleans, LA, USA, May 6–9, 2019*. OpenReview.net, 2019.
- [AM18] Michael Atiyah and Ian MacDonald. *Introduction to Commutative Algebra*. CRC Press, 2018.
- [BCF23] David Barnhill, John Cobb, and Matthew Faust. Likelihood correspondence of statistical models. [arXiv:2312.08501](https://arxiv.org/abs/2312.08501), 2023.

- [BP22] Marko Berghoff and Erik Panzer. Hierarchies in relative Picard–Lefschetz theory. [arXiv:2212.06661](https://arxiv.org/abs/2212.06661), 2022.
- [BBO18] Alessandra Bernardi, Grigoriy Blekherman, and Giorgio Ottaviani. On real typical ranks. *Bollettino dell’Unione Matematica Italiana*, 11:293–307, 2018.
- [BH90] Michael V Berry and Chris J Howls. Hyperasymptotics. *Proceedings of the Royal Society of London. Series A: Mathematical and Physical Sciences*, 430(1880):653–668, 1990.
- [BIZ80] Daniel Bessis, Claude Itzykson, and Jean-Bernard Zuber. Quantum field theory techniques in graphical enumeration. *Advances in Applied Mathematics*, 1(2):109–157, 1980.
- [BN06] Christopher M Bishop and Nasser M Nasrabadi. *Pattern Recognition and Machine Learning*, volume 4. Springer, 2006.
- [BBKP19] Thomas Bitoun, Christian Bogner, René Pascal Klausen, and Erik Panzer. Feynman integral relations from parametric annihilators. *Letters in Mathematical Physics*, 109:497–564, 2019.
- [BCR13] Jacek Bochnak, Michel Coste, and Marie-Françoise Roy. *Real algebraic geometry*, volume 36. Springer Science & Business Media, 2013.
- [Bol01] Béla Bollobás. *Random Graphs*. Cambridge University Press, 2 edition, 2001.
- [BB09] Julius Borcea and Petter Brändén. The Lee–Yang and Pólya–Schur programs. I. Linear operators preserving stability. *Inventiones Mathematicae*, 177(3):541–569, 2009.
- [Bor18] Michael Borinsky. *Graphs in Perturbation Theory: Algebraic Structure and Asymptotics*. Springer, 2018.
- [BMW24] Michael Borinsky, Chiara Meroni, and Maximilian Wiesmann. Bivariate exponential integrals and edge-bicolored graphs. [arXiv:2409.18607](https://arxiv.org/abs/2409.18607), 2024. To appear in *Le Matematiche*.
- [BV20] Michael Borinsky and Karen Vogtmann. The Euler characteristic of  $\text{Out}(F_n)$ . *Commentarii Mathematici Helvetici*, 95(4):703–748, 2020.
- [BBG04] Jonathan M Borwein, David H Bailey, and Roland Girgensohn. *Experimentation in mathematics: Computational paths to discovery*. AK Peters/CRC Press, 2004.
- [BNT20] Nicolas Boullé, Yuji Nakatsukasa, and Alex Townsend. Rational neural networks. *Advances in Neural Information Processing Systems*, 33:14243–14253, 2020.
- [BO08] Maria C Brambilla and Giorgio Ottaviani. On the Alexander–Hirschowitz theorem. *Journal of Pure and Applied Algebra*, 212(5):1229–1251, 2008.
- [BPAM23] Pierre Bréchet, Katerina Papagiannouli, Jing An, and Guido Montúfar. Critical points and convergence analysis of generative deep linear networks trained with Bures–Wasserstein loss. In *International Conference on Machine Learning*, pages 3106–3147. PMLR, 2023.

- [BKS18] Paul Breiding, Sara Kališnik, Bernd Sturmfels, and Madeleine Weinstein. Learning algebraic varieties from samples. *Revista Matemática Complutense*, 31:545–593, 2018.
- [BKS24] Paul Breiding, Kathlén Kohn, and Bernd Sturmfels. *Metric algebraic geometry*. Springer Nature, 2024.
- [BT18] Paul Breiding and Sascha Timme. HomotopyContinuation.jl: A package for homotopy continuation in Julia. In *Mathematical Software—ICMS 2018: 6th International Conference, South Bend, IN, USA, July 24–27, 2018, Proceedings* 6, pages 458–465. Springer, 2018.
- [BS11] Sarah B Brodsky and Bernd Sturmfels. Tropical quadrics through three points. *Linear algebra and its applications*, 435(7):1778–1785, 2011.
- [Bro09] Francis Brown. On the periods of some Feynman integrals. [arXiv:0910.0114](https://arxiv.org/abs/0910.0114), 2009.
- [BP12] Winton Brown and David Poulin. Quantum Markov networks and commuting Hamiltonians. [arXiv:1206.0755](https://arxiv.org/abs/1206.0755), 2012.
- [BV06] Winfried Bruns and Udo Vetter. *Determinantal Rings*, volume 1327. Springer, 2006.
- [CEGM19] Freddy Cachazo, Nick Early, Alfredo Guevara, and Sebastian Mizera. Scattering equations: from projective spaces to tropical Grassmannians. *Journal of High Energy Physics*, 2019(6):1–33, 2019.
- [CHY14] Freddy Cachazo, Song He, and Ellis Ye Yuan. Scattering equations and Kawai–Lewellen–Tye orthogonality. *Physical Review D*, 90(6):065001, 2014.
- [CHKS06] Fabrizio Catanese, Serkan Hoşten, Amit Khetan, and Bernd Sturmfels. The maximum likelihood degree. *American Journal of Mathematics*, 128(3):671–697, 2006.
- [CC99] Luca Chiantini and Ciro Ciliberto. On the Severi varieties of surfaces in  $\mathbb{P}^3$ . *Journal of Algebraic Geometry*, 8:67–83, 1999.
- [Chi03] Jaydeep V Chipalkatti. On equations defining coincident root loci. *Journal of Algebra*, 267(1):246–271, 2003.
- [Chi04] Jaydeep V Chipalkatti. Invariant equations defining coincident root loci. *Archiv der Mathematik*, 83:422–428, 2004.
- [CMB<sup>+</sup>20] Grigorios G Chrysos, Stylianos Moschoglou, Giorgos Bouritsas, Yannis Panagakis, Jiankang Deng, and Stefanos Zafeiriou. P-nets: Deep polynomial neural networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7325–7335, 2020.
- [Cim12] David Cimasoni. The critical Ising model via Kac–Ward matrices. *Communications in Mathematical Physics*, 316(1):99–126, 2012.
- [CHKO24] Oliver Clarke, Serkan Hoşten, Nataliia Kushnerchuk, and Janike Oldekop. Matroid stratification of ML degrees of independence models. *Algebraic Statistics*, 15(2):199–223, 2024.

- [CC20] Patrick Clarke and David A Cox. Moment maps, strict linear precision, and maximum likelihood degree one. *Advances in Mathematics*, 370:107233, 2020.
- [CDFV11] Daniel Cohen, Graham Denham, Michael Falk, and Alexander Varchenko. Critical points and resonance of hyperplane arrangements. *Canadian Journal of Mathematics*, 63(5):1038–1057, 2011.
- [CGLM08] Pierre Comon, Gene Golub, Lek-Heng Lim, and Bernard Mourrain. Symmetric tensors and symmetric tensor rank. *SIAM Journal on Matrix Analysis and Applications*, 30(3):1254–1279, 2008.
- [CM96] Pierre Comon and Bernard Mourrain. Decomposition of quantics in sums of powers of linear forms. *Signal Processing*, 53(2-3):93–107, 1996.
- [CO12] Pierre Comon and Giorgio Ottaviani. On the typical rank of real binary forms. *Linear and multilinear algebra*, 60(6):657–667, 2012.
- [CK99] David A Cox and Sheldon Katz. *Mirror symmetry and algebraic geometry*, volume 68. American Mathematical Society Providence, RI, 1999.
- [CLS11] David A Cox, John B Little, and Henry K Schenck. *Toric varieties*, volume 124. American Mathematical Society, 2011.
- [CH23] Joseph Cummings and Benjamin Hollering. Computing implicitizations of multi-graded polynomial maps. [arXiv:2311.07678](https://arxiv.org/abs/2311.07678), 2023.
- [dD15] Giacomo d’Antonio and Emanuele Delucchi. Minimality of toric arrangements. *Journal of the European Mathematical Society*, 17(3):483–521, 2015.
- [DR72] John N Darroch and Douglas Ratcliff. Generalized iterative scaling for log-linear models. *The Annals of Mathematical Statistics*, 43(5):1470–1480, 1972.
- [DDPS23] Isobel Davies, Eliana Duarte, Irem Portakal, and Miruna-Ştefana Sorea. Families of polytopes with rational linear precision in higher dimensions. *Foundations of Computational Mathematics*, 23(6):2151–2202, 2023.
- [DH02] Eric Delabaere and Christopher J Howls. Global asymptotics for multiple integrals with boundaries. *Duke Mathematical Journal*, 112(2):199–266, 2002.
- [DSS<sup>+</sup>13] Graham Denham, Hal Schenck, Mathias Schulze, Max Wakefield, and Uli Walther. Local cohomology of logarithmic forms. In *Annales de l’Institut Fourier*, volume 63, pages 1177–1203, 2013.
- [DS09] Graham Denham and Mathias Schulze. Complexes, duality and Chern classes of logarithmic forms along hyperplane arrangements. *Advanced Studies in Pure Mathematics*, 99:27–57, 2009.
- [DNM<sup>+</sup>16] Sudip Dey, Susmita Naskar, Tanmoy Mukhopadhyay, Uwe Gohs, Axel Spick-enheuer, Lars Bittrich, Srinivas Sriramula, Sondipon Adhikari, and Gert Heinrich. Uncertain natural frequency analysis of composite plates including effect of noise—a polynomial neural network approach. *Composite Structures*, 143:130–142, 2016.
- [DGM21] Serena Di Giorgio and Paulo Mateus. On the complexity of finding the maximum entropy compatible quantum state. *Mathematics*, 9(2), 2021.

- [DGMM20] Serena Di Giorgio, Paulo Mateus, and Bruno Mera. Recoverability from direct quantum correlations. *Journal of Physics A: Mathematical and Theoretical*, 53(18):185301, 2020.
- [DH88] Steven Diaz and Joe Harris. Geometry of the Severi variety. *Transactions of the American Mathematical Society*, 309(1):1–34, 1988.
- [DHT17] Alicia Dickenstein, Maria I Herrero, and Luis F Tabera. Arithmetics and combinatorics of tropical Severi varieties of univariate polynomials. *Israel Journal of Mathematics*, 221(2):741–777, 2017.
- [DP03] Alexandru Dimca and Stefan Papadima. Hypersurface complements, Milnor fibers and higher homotopy groups of arrangements. *Annals of Mathematics*, 158(2):473–507, 2003.
- [DM07] Valery Dolotin and Alexei Morozov. *Introduction to non-linear algebra*. World Scientific, 2007.
- [Dra24] Jan Draisma. Erratum: A counterexample to Comon’s conjecture. *SIAM Journal on Applied Algebra and Geometry*, 8(1):225–225, 2024.
- [DHO<sup>+</sup>15] Jan Draisma, Emil Horobet, Giorgio Ottaviani, Bernd Sturmfels, and Rekha R Thomas. The Euclidean distance degree of an algebraic variety. *Foundations of Computational Mathematics*, 16(1):99–149, January 2015.
- [DSS08] Mathias Drton, Bernd Sturmfels, and Seth Sullivant. *Lectures on algebraic statistics*, volume 39. Springer Science & Business Media, 2008.
- [DHW23] Eliana Duarte, Benjamin Hollering, and Maximilian Wiesmann. Toric fiber products in geometric modeling. In Frank Nielsen and Frédéric Barbaresco, editors, *Geometric Science of Information*, pages 494–503. Springer Nature Switzerland, 2023. Extended version at [arXiv:2303.08754](https://arxiv.org/abs/2303.08754).
- [DMS21] Eliana Duarte, Orlando Marigliano, and Bernd Sturmfels. Discrete statistical models with rational maximum likelihood estimator. *Bernoulli*, 27(1):135–154, 2021.
- [DPW23a] Eliana Duarte, Dmitrii Pavlov, and Maximilian Wiesmann. Algebraic geometry of quantum graphical models. [arXiv:2308.11538](https://arxiv.org/abs/2308.11538), 2023. Submitted to Advances in Applied Mathematics.
- [DPW23b] Eliana Duarte, Dmitrii Pavlov, and Maximilian Wiesmann. MathRepo QuantumGraphicalModels. <https://mathrepo.mis.mpg.de/QuantumGraphicalModels>, 2023.
- [DC23] Hugo Duminil-Copin. 100 years of the (critical) Ising model on the hypercubic lattice. In *ICM—International Congress of Mathematicians. Vol. 1. Prize lectures*, pages 164–210. EMS Press, Berlin, 2023.
- [Dup15] Clément Dupont. The Orlik–Solomon model for hypersurface arrangements. In *Annales de l’institut Fourier*, volume 65, pages 2507–2545, 2015.
- [EPS24] Nick Early, Anaëlle Pfister, and Bernd Sturmfels. Minimal kinematics on  $\mathcal{M}_{0,n}$ . [arXiv:2402.03065](https://arxiv.org/abs/2402.03065), 2024.

- [ER93] Paul H Edelman and Victor Reiner. A counterexample to Orlik’s conjecture. *Proceedings of the American Mathematical Society*, 118(3):927–929, 1993.
- [Efr22] Bradley Efron. *Exponential Families in Theory and Practice*. Institute of Mathematical Statistics Textbooks. Cambridge University Press, 2022.
- [Eis13] David Eisenbud. *Commutative algebra: with a view toward algebraic geometry*, volume 150. Springer Science & Business Media, 2013.
- [Eis18] David Eisenbud. The ReesAlgebra package in Macaulay2. *Journal of Software for Algebra and Geometry*, 8:49–60, 2018.
- [EHU03] David Eisenbud, Craig Huneke, and Bernd Ulrich. What is the Rees algebra of a module? *Proceedings of the American Mathematical Society*, 131(3):701–708, 2003.
- [EKS14] Alexander Engström, Thomas Kahle, and Seth Sullivant. Multigraded commutative algebra of graph decompositions. *Journal of Algebraic Combinatorics*, 39:335–372, 2014.
- [Est13] Alexander Esterov. The discriminant of a system of equations. *Advances in Mathematics*, 245:534–572, 2013.
- [Est17] Alexander Esterov. Characteristic classes of affine varieties and Plücker formulas for affine morphisms. *Journal of the European Mathematical Society*, 20(1):15–59, 2017.
- [Eva20] Robin J Evans. Model selection and local geometry. *The Annals of Statistics*, 48(6):3513 – 3544, 2020.
- [FNR06] László M Fehér, András Némethi, and Richárd Rimányi. Coincident root loci of binary forms. *Michigan Mathematical Journal*, 54(2):375–392, 2006.
- [Fel81] Joseph Felsenstein. Evolutionary trees from DNA sequences: a maximum likelihood approach. *Journal of Molecular Evolution*, 17:368–376, 1981.
- [FMT24] Claudia Fevola, Sebastian Mizera, and Simon Telen. Principal Landau determinants. *Computer Physics Communications*, 303:109278, 2024.
- [FRWY24] Bella Finkel, Jose I Rodriguez, Chenxi Wu, and Thomas Yahl. Activation thresholds and expressiveness of polynomial neural networks. [arXiv:2408.04569](https://arxiv.org/abs/2408.04569), 2024.
- [Ful82] William Fulton. On nodal curves. In *Algebraic geometry: Open problems*, volume 997 of *Lecture notes in Mathematics*, pages 146–155. Springer Berlin-Heidelberg-New York, 1982.
- [Ful13] William Fulton. *Intersection theory*, volume 2. Springer Science & Business Media, 2013.
- [GPS10] Luis D Garcia-Puente and Frank Sottile. Linear precision for parametric patches. *Advances in Computational Mathematics*, 33(2):191–214, 2010.
- [GMS06] Dan Geiger, Christopher Meek, and Bernd Sturmfels. On the toric algebra of graphical models. *The Annals of Statistics*, 34(3):1463 – 1492, 2006.

- [GKZ08] Israel M Gelfand, Mikhail Kapranov, and Andrei Zelevinsky. *Discriminants, resultants, and multidimensional determinants*. Springer Science & Business Media, 2008.

[GHL11] Rozaida Ghazali, Abir J Hussain, and Panos Liatsis. Dynamic ridge polynomial neural network: Forecasting the univariate non-stationary and stationary trading signals. *Expert Systems with Applications*, 38(4):3765–3776, 2011.

[GBC16] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016. [www.deeplearningbook.org](http://www.deeplearningbook.org).

[GS02] Daniel R Grayson and Michael E Stillman. Macaulay2, a software system for research in algebraic geometry, 2002. Available at <https://macaulay2.com/>.

[GLS07] Gert-Martin Greuel, Christoph Lossen, and Eugenii I Shustin. *Introduction to singularities and deformations*. Springer Science & Business Media, 2007.

[GLMW22] J Elisenda Grigsby, Kathryn Lindsey, Robert Meyerhoff, and Chenxi Wu. Functional dimension of feedforward ReLU neural networks. [arXiv:2209.04036](https://arxiv.org/abs/2209.04036), 2022.

[Hab74] Shelby J Haberman. Log-linear models for frequency tables derived by indirect observation: Maximum likelihood equations. *The Annals of Statistics*, 2(5):911–924, 1974.

[HTF09] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning*. Springer New York, 2009.

[HJPW04] Patrick Hayden, Richard Jozsa, Dénes Petz, and Andreas Winter. Structure of states which satisfy strong subadditivity of quantum entropy with equality. *Communications in Mathematical Physics*, 246:359–374, 2004.

[Hay98] Simon Haykin. *Neural networks: A comprehensive foundation*. Prentice Hall PTR, 1998.

[HEB04] Marc Hein, Jens Eisert, and Hans J Briegel. Multiparty entanglement in graph states. *Physical Review A*, 69(6):062311, 2004.

[HN23] Martin Helmer and Vidit Nanda. Conormal spaces and Whitney stratifications. *Foundations of Computational Mathematics*, 23(5):1745–1780, 2023.

[HPT24] Martin Helmer, Georgios Papathanasiou, and Felix Tellander. Landau singularities from Whitney stratifications. [arXiv:2402.14787](https://arxiv.org/abs/2402.14787), 2024.

[Hil87] David Hilbert. Singularitäten der Diskriminantenfläche. *Mathematische Annalen*, 30:437–441, 1887.

[HJ85] Roger A Horn and Charles R Johnson. *Matrix Analysis*. Cambridge University Press, 1985.

[HSW89] Kurt Hornik, Maxwell Stinchcombe, and Halbert White. Multilayer feedforward networks are universal approximators. *Neural networks*, 2(5):359–366, 1989.

[HKS05] Serkan Hoşten, Amit Khetan, and Bernd Sturmfels. Solving the likelihood equations. *Foundations of Computational Mathematics*, 5:389–407, 2005.

- [HSHK03] Lin-Lin Huang, Akinobu Shimizu, Yoshihiro Hagihara, and Hidefumi Kobatake. Face detection from cluttered images using a polynomial neural network. *Neurocomputing*, 51:197–211, 2003.
- [Huh12] June Huh. Milnor numbers of projective hypersurfaces and the chromatic polynomial of graphs. *Journal of the American Mathematical Society*, 25(3):907–927, 2012.
- [Huh13] June Huh. The maximum likelihood degree of a very affine variety. *Compositio Mathematica*, 149(8):1245–1266, 2013.
- [Huh14] June Huh. Varieties with maximum likelihood degree one. *Journal of Algebraic Statistics*, 5(1):1–17, 2014.
- [HS14] June Huh and Bernd Sturmfels. Likelihood geometry. *Combinatorial algebraic geometry*, 2108:63–117, 2014.
- [Hun81] Craig Huneke. On the symmetric algebra of a module. *Journal of Algebra*, 69:113–119, 1981.
- [Jam06] Frederick James. *Statistical methods in experimental physics*. World Scientific Publishing Company, 2006.
- [KKM<sup>+</sup>24a] Thomas Kahle, Lukas Kühne, Leonie Mühlherr, Bernd Sturmfels, and Maximilian Wiesmann. Arrangements and likelihood. [arXiv:2411.09508](https://arxiv.org/abs/2411.09508), 2024. Submitted to Vietnam Journal of Mathematics.
- [KKM<sup>+</sup>24b] Thomas Kahle, Lukas Kühne, Leonie Mühlherr, Bernd Sturmfels, and Maximilian Wiesmann. MathRepo page ArrangementsLikelihood. <https://mathrepo.mis.mpg.de/ArrangementsLikelihood>, 2024.
- [KV24] Thomas Kahle and Julian Vill. Efficiently deciding if an ideal is toric after a linear coordinate change. [arXiv:2408.14323](https://arxiv.org/abs/2408.14323), 2024.
- [Kaz86] Vladimir A Kazakov. Ising model on a dynamical planar random lattice: Exact solution. *Physics Letters A*, 119:140–144, 1986.
- [KTB] Joe Kileel, Matthew Trager, and Joan Bruna. Github repository [https://github.com/mtrager/polynomial\\_networks](https://github.com/mtrager/polynomial_networks).
- [KTB19] Joe Kileel, Matthew Trager, and Joan Bruna. On the expressive power of deep polynomial neural networks. *Advances in neural information processing systems*, 32, 2019.
- [KF09] Daphne Koller and Nir Friedman. *Probabilistic Graphical Models: Principles and Techniques*. MIT press, 2009.
- [Kra02] Rimvydas Krasauskas. Toric surface patches. *Advances in Computational Mathematics*, 17(1):89–113, 2002.
- [KLW24a] Kaie Kubjas, Jiayi Li, and Maximilian Wiesmann. Geometry of polynomial neural networks. *Algebraic Statistics*, 15(2):295–328, 2024.
- [KLW24b] Kaie Kubjas, Jiayi Li, and Maximilian Wiesmann. MathRepo page PolynomialNeuralNetworks. <https://mathrepo.mis.mpg.de/PolynomialNeuralNetworks>, 2024.

- [Kur12] Simon Kurmann. Some remarks on equations defining coincident root loci. *Journal of Algebra*, 352(1):223–231, 2012.
- [Lam24] Thomas Lam. Moduli spaces in positive geometry. [arXiv:2405.17332](https://arxiv.org/abs/2405.17332), 2024. To appear in *Le Matematiche*.
- [Lan11] Joseph M Landsberg. *Tensors: geometry and applications*, volume 128. American Mathematical Society, 2011.
- [Lan19] Joseph M Landsberg. A very brief introduction to quantum computing and quantum information theory for mathematicians. In Edoardo Ballico, Alessandra Bernardi, Iacopo Carusotto, Sonia Mazzucchi, and Valter Moretti, editors, *Quantum Physics and Geometry*, pages 5–41. Springer, 2019.
- [Lau96] Steffen L Lauritzen. *Graphical models*, volume 17 of *Oxford Statistical Science Series*. The Clarendon Press, Oxford University Press, New York, 1996. Oxford Science Publications.
- [LGZJ12] Jean-Claude Le Guillou and Jean Zinn-Justin. *Large-order behaviour of perturbation theory*, volume 7. Elsevier, 2012.
- [Leb74] Karsten Lebelt. Torsion äußerer Potenzen von Moduln der homologischen Dimension 1. *Mathematische Annalen*, 211:183–197, 1974.
- [Leb77] Karsten Lebelt. Freie Auflösungen äußerer Potenzen. *manuscripta mathematica*, 21:341–355, 1977.
- [LS16] Hwangrae Lee and Bernd Sturmfels. Duality of multiple root loci. *Journal of Algebra*, 446:499–526, 2016.
- [LP08] Matthew S Leifer and David Poulin. Quantum graphical models and belief propagation. *Annals of Physics*, 323(8):1899–1946, 2008.
- [LR73] Elliott H Lieb and Mary B Ruskai. Proof of the strong subadditivity of quantum-mechanical entropy. *Les rencontres physiciens-mathématiciens de Strasbourg-RCP25*, 19:36–55, 1973.
- [Lin11] Shaowei Lin. *Algebraic methods for evaluating integrals in Bayesian statistics*. University of California, Berkeley, 2011.
- [LUSB14] Shaowei Lin, Caroline Uhler, Bernd Sturmfels, and Peter Bühlmann. Hypersurfaces and their singularities in partial correlation testing. *Foundations of Computational Mathematics*, 14(5):1079–1116, 2014.
- [MDLW18] Marloes Maathuis, Mathias Drton, Steffen L Lauritzen, and Martin Wainwright. *Handbook of Graphical Models*. CRC Press, 2018.
- [MS21] Diane Maclagan and Bernd Sturmfels. *Introduction to tropical geometry*, volume 161. American Mathematical Society, 2021.
- [MP23] Aida Maraj and Arpan Pal. Symmetry Lie algebras of varieties with applications to algebraic statistics. [arXiv:2309.10741](https://arxiv.org/abs/2309.10741), 2023.
- [MM19] Pierpaolo Mastrolia and Sebastian Mizera. Feynman integrals and intersection theory. *Journal of High Energy Physics*, 2019(2):1–25, 2019.

- [Mat12] John Mather. Notes on topological stability. *Bulletin of the American Mathematical Society*, 49(4):475–506, 2012.
- [MHMT23] Saiei-Jaeyeong Matsubara-Heo, Sebastian Mizera, and Simon Telen. Four lectures on Euler integrals. *SciPost Physics Lecture Notes*, page 75, 2023.
- [MMSV16] Mateusz Michałek, Hyunsuk Moon, Bernd Sturmfels, and Emanuele Ventura. Real rank geometry of ternary forms. *Annali di Matematica Pura ed Applicata*, 196(3):1025–1054, August 2016.
- [MS21] Mateusz Michałek and Bernd Sturmfels. *Invitation to nonlinear algebra*, volume 211. American Mathematical Society, 2021.
- [MS05] Ezra Miller and Bernd Sturmfels. *Combinatorial commutative algebra*, volume 227. Springer Science & Business Media, 2005.
- [MT22] Sebastian Mizera and Simon Telen. Landau discriminants. *Journal of High Energy Physics*, 2022(8):1–57, 2022.
- [Mol21] Sam Molcho. Universal stacky semistable reduction. *Israel Journal of Mathematics*, 242:55–82, 2021.
- [Müh] Leonie Mühlherr. Separator-based derivations of graphic arrangements. In preparation.
- [NM18] Sarat C Nayak and Bijan B Misra. Estimating stock closing indices using a GA-weighted condensed polynomial neural network. *Financial Innovation*, 4(1):21, 2018.
- [NGKG13] Sönke Niekamp, Tobias Galla, Matthias Kleinmann, and Otfried Gühne. Computing complexity measures for quantum states based on exponential families. *Journal of Physics A: Mathematical and Theoretical*, 46(12):125301, 2013.
- [Nie15] Michael A Nielsen. *Neural Networks and Deep Learning*. Determination Press, 2015.
- [NC02] Michael A Nielsen and Isaac Chuang. *Quantum Computation and Quantum Information*. American Association of Physics Teachers, 2002.
- [OPP03] Sung-Kwun Oh, Witold Pedrycz, and Byoung-Jun Park. Polynomial neural networks architecture: analysis and design. *Computers & Electrical Engineering*, 29(6):703–725, 2003.
- [OT13] Peter Orlik and Hiroaki Terao. *Arrangements of hyperplanes*, volume 300. Springer Science & Business Media, 2013.
- [OSC24] OSCAR – open source computer algebra research system, version 1.0.4, 2024.
- [PS05] Lior Pachter and Bernd Sturmfels. *Algebraic Statistics for Computational Biology*, volume 13. Cambridge University Press, 2005.
- [Pan15] Erik Panzer. Algorithms for the symbolic integration of hyperlogarithms with applications to Feynman integrals. *Computer Physics Communications*, 188:148–166, 2015.

- [Par88] Adam Parusiński. A generalization of the Milnor number. *Mathematische Annalen*, 281:247–254, 1988.
- [Pav24] Dmitrii Pavlov. Logarithmically sparse symmetric matrices. *Beiträge zur Algebra und Geometrie/Contributions to Algebra and Geometry*, pages 1–16, 2024.
- [PST23] Dmitrii Pavlov, Bernd Sturmfels, and Simon Telen. Gibbs manifolds. *Information Geometry*, pages 1–27, 2023.
- [Pet86] Dénes Petz. Sufficient subalgebras and the relative entropy of states of a von neumann algebra. *Communications in Mathematical Physics*, 105:123–131, 1986.
- [PH11] David Poulin and Matthew B Hastings. Markov entropy decomposition: a variational dual for quantum belief propagation. *Physical Review Letters*, 106(8):080403, 2011.
- [RPK<sup>+</sup>17] Maithra Raghu, Ben Poole, Jon Kleinberg, Surya Ganguli, and Jascha Sohl-Dickstein. On the expressive power of deep neural networks. In *International Conference on Machine Learning*, pages 2847–2854. PMLR, 2017.
- [Rot97] Gian-Carlo Rota. *Indiscrete Thoughts*. Modern Birkhäuser Classics. Birkhäuser Boston, MA, 1997.
- [Sai80] Kyoji Saito. Theory of logarithmic differential forms and logarithmic vector fields. *Journal of the Faculty of Science, University of Tokyo, Section IA Mathematics*, 27(2):265–291, 1980.
- [SvdV23] Anna-Laura Sattelberger and Robin van der Veer. Maximum likelihood estimation from a tropical and a Bernstein–Sato perspective. *International Mathematics Research Notices*, 2023(6):5263–5292, 2023.
- [SUV03] Aron Simis, Bernd Ulrich, and Wolmer V Vasconcelos. Rees algebras of modules. *Proceedings of the London Mathematical Society*, 87(3):610–646, 2003.
- [SV81] Aron Simis and Wolmer V Vasconcelos. On the dimension and integrality of symmetric algebras. *Mathematische Zeitschrift*, 177(3):341–358, 1981.
- [Ski18] David Skinner. Quantum field theory II. *Lecture notes, Part III of the Mathematical Tripos, University of Cambridge*, 2018.
- [SW<sup>+</sup>05] Andrew J Sommese, Charles W Wampler, et al. *The Numerical solution of systems of polynomials arising in engineering and science*. World Scientific, 2005.
- [SV86] Jürgen Stückrad and Wolfgang Vogel. *Buchsbaum rings and applications: an interaction between algebra, geometry, and topology*, volume 21. Springer Berlin, 1986.
- [ST21] Bernd Sturmfels and Simon Telen. Likelihood equations and scattering amplitudes. *Algebraic Statistics*, 12(2):167–186, 2021.
- [Sul07] Seth Sullivant. Toric fiber products. *Journal of Algebra*, 316(2):560–577, 2007.
- [Sul18] Seth Sullivant. *Algebraic statistics*, volume 194. American Mathematical Society, 2018.
- [Tel22] Simon Telen. Introduction to toric geometry. [arXiv:2203.01690](https://arxiv.org/abs/2203.01690), 2022.

- [TW24a] Simon Telen and Maximilian Wiesmann. Euler stratifications of hypersurface families. [arXiv:2407.18176](https://arxiv.org/abs/2407.18176), 2024. Submitted to the Journal of the European Mathematical Society.
- [TW24b] Simon Telen and Maximilian Wiesmann. MathRepo page EulerStratifications. <https://mathrepo.mis.mpg.de/EulerStratifications>, 2024.
- [Ter80] Hiroaki Terao. Arrangements of hyperplanes and their freeness I. *Journal of the Faculty of Science, University of Tokyo, Section IA Mathematics*, 27(2):293–312, 1980.
- [Tev06] Evgeni A Tevelev. *Projective duality and homogeneous spaces*, volume 133. Springer Science & Business Media, 2006.
- [Tho64] René Thom. Local topological properties of differentiable mappings. In *Differential Analysis, Bombay Colloquium, 1964*, pages 191–202. Tata Institute of Fundamental Research, Bombay, 1964.
- [TKB20] Matthew Trager, Kathlén Kohn, and Joan Bruna. Pure and spurious critical points: a geometric study of linear networks. In *8th International Conference on Learning Representations, ICLR 2020, Addis Ababa, Ethiopia, April 26-30, 2020*. OpenReview.net, 2020.
- [TV15] Tomáš Tyc and Jan Vlach. Quantum marginal problems. *The European Physical Journal D*, 69:1–6, 2015.
- [URBY13] Caroline Uhler, Garvesh Raskutti, Peter Bühlmann, and Bin Yu. Geometry of the faithfulness assumption in causal inference. *The Annals of Statistics*, 41(2):436–463, 2013.
- [Wal06] Charles T C Wall. Regular stratifications. In *Dynamical Systems—Warwick 1974: Proceedings of a Symposium held at the University of Warwick 1973/74*, pages 332–344. Springer, 2006.
- [Wat09] Sumio Watanabe. *Algebraic geometry and statistical learning theory*, volume 25. Cambridge University Press, 2009.
- [Wei94] Charles A Weibel. *An Introduction to Homological Algebra*. Cambridge Studies in Advanced Mathematics. Cambridge University Press, 1994.
- [Wei22] Stefan Weinzierl. *Feynman Integrals*. UNITEXT for Physics. Springer Cham, 2022.
- [WG23] Stephan Weis and João Gouveia. The face lattice of the set of reduced density matrices and its coatoms. *Information Geometry*, pages 1–34, 2023.
- [Wer74] Paul Werbos. Beyond regression: New tools for prediction and analysis in the behavioral sciences. *PhD thesis, Committee on Applied Mathematics, Harvard University, Cambridge, MA*, 1974.
- [Wey89] Jerzy Weyman. The equations of strata for binary forms. *Journal of Algebra*, 122(1):244–249, 1989.
- [Whi65] Hassler Whitney. Tangents to an analytic variety. *Annals of Mathematics*, 81(3):496–549, 1965.

- [Wil13] Mark M Wilde. *Quantum Information Theory*. Cambridge University Press, 2013.
- [Wor18] Nicholas Wormald. Asymptotic enumeration of graphs with given degree sequence. In *Proceedings of the International Congress of Mathematicians—Rio de Janeiro 2018. Vol. IV. Invited lectures*, pages 3245–3264. World Scientific Publishing, Hackensack, NJ, 2018.
- [Yan13] Jihyeon J Yang. Tropical Severi varieties. *Portugaliae Mathematica*, 70(1):59–91, 2013.
- [YHN<sup>+</sup>21] Mohsen Yavartanoo, Shih-Hsuan Hung, Reyhaneh Neshatavar, Yue Zhang, and Kyoung Mu Lee. Polynet: Polynomial neural network for 3d shape recognition with polyshape representation. In *International Conference on 3D Vision, 3DV 2021, London, United Kingdom, December 1-3, 2021*, pages 1014–1023. IEEE, 2021.
- [Yok86] Shoji Yokura. Polar classes and Segre classes on singular projective varieties. *Transactions of the American Mathematical Society*, 298(1):169–191, 1986.
- [Zha18] Xiping Zhang. Chern classes and characteristic cycles of determinantal varieties. *Journal of Algebra*, 497:55–91, 2018.
- [Zho08] Duanlu L Zhou. Irreducible multiparty correlations in quantum states without maximal rank. *Physical Review Letters*, 101:180505, Oct 2008.

# Index

- A*-discriminant, 55  
*A*-resultant, 68  
abstract normal toric variety, 44  
activation function, 34  
ambient dimension, 104  
architecture, 34  
Aronhold invariant, 59  
asymptotic bottleneck, 113  
asymptotic expansion, 142  
backpropagation, 35  
Bell state, 37  
binary optics, 72  
Birch's Theorem, 27  
blending function, 42  
braid arrangement, 79  
Brill–Gordon locus, 60  
categorical fibre product, 44  
Chern–Mather class, 118  
coincident root locus, 60  
conditional independence, 25  
constructible set, 57  
data, 26  
defect, 104  
density matrix, 36  
derivation, 32  
discrete statistical model, 24  
edge-bicoloured graph, 144  
empirical risk minimisation, 34  
entangled state, 37  
equisingular locus, 54  
essential arrangement, 79  
Euclidean distance degree, *see* generic Euclidean distance degree  
Euler characteristic, 28  
Euler derivation, 32  
Euler discriminant polynomial, 55  
Euler discriminant variety, 55  
Euler integral, 74  
Euler stratification, 54  
expected dimension, 103  
exponential family, discrete regular, 24  
expressivity, 102  
feedforward neural network, 33  
Feynman integral, 73  
filling architecture, 102  
free arrangement, 33  
Gauss map, 57  
generalised characteristic polynomial, 106  
generating function, 141  
generic arrangement, 88  
generic Euclidean distance degree, 116  
gentle arrangement, 81  
Gibbs manifold, 134  
Gibbs variety, 134  
global Markov statements, 25  
Gorenstein toric Fano variety, 75  
graphic arrangement, 33  
half-edge labelled graph, 144  
Hamiltonian, 37  
Hammersley–Clifford Theorem, 26  
hidden layer, 34  
Hirzebruch surface, 72  
Horn matrix, 28  
Horn pair, 29  
Horn parametrisation, 29  
hypersurface arrangement, 31  
independence model, 25  
input layer, 34  
intersection lattice, 33  
Ising model, 147  
Jacobian syzygy module, 79

- Kähler differential, 32  
 Kouchnirenko's Theorem, 24  
 Landau analysis, 74  
 Landau variety, 56  
 Laplace method, 141  
 learning degree, 116  
 learning rate, 35  
 Lee–Yang theory, 155  
 likelihood correspondence, 26  
 likelihood degeneration, 69  
 likelihood function, 78  
 likelihood ideal, 80  
 likelihood module, 79  
 linear type, 30  
 local Hamiltonian, 134  
 localisation of hyperplane arrangement, 33  
 log-affine model, 24  
 log-derivation module, 79  
 log-likelihood function, 26  
 log-linear model, 24  
 logarithmic  $\mathcal{A}$ -derivation, 32  
 logarithmic differential form, 32  
 logarithmic ideal, 82  
 Mandelstam invariant, 74  
 marginal likelihood integral, 141  
 master function, 78  
 master integral, 73  
 matroid stratification, 72  
 maximally mixed state, 37  
 maximum likelihood degree, 26  
 maximum likelihood estimate, 26  
 maximum likelihood estimation, 26  
 maximum likelihood estimator, 26  
 meromorphic ideal, 82  
 Milnor algebra, 59  
 Milnor number, 59  
 MLE, *see* maximum likelihood estimate  
 multidegree, 83  
 multigrading, 42  
 multiple root locus, 60  
 Nash blow-up, 118  
 Nash tangent bundle, 118  
 neural network, 33  
 neuromanifold, 101  
 neuron, 34  
 neurovariety, 101  
 normalised volume, 24  
 orbit-cone-correspondence, 23  
 orbit-face-correspondence, 23  
 output layer, 34  
 pairwise Markov statements, 26  
 parametric likelihood correspondence, 80  
 parametric ML degree, 83  
 partial trace, 37  
 partition function, 147  
 path integral, 141  
 Pauli matrices, 37  
 Petz recovery map, 131  
 Petz variety, 132  
 phase transition, 152  
 polar discriminant, 66  
 polynomial neural network, 101  
 pre-likelihood ideal, 81  
 principal  $A$ -determinant, 55  
 principal Landau determinant, 56  
 probability simplex, 24  
 pruned module of logarithmic forms, 90  
 QCMI variety, 128  
 quantum conditional mutual information, 127  
 quantum conditional mutual information axioms, 127  
 quantum relative entropy, 139  
 quantum state, 36  
 quasi-projective variety, 56  
 qubit, 36  
 qudit, 36  
 rational linear precision, 42, 48  
 Rees algebra, 29  
 reflexive polygon, 75  
 Saito derivation, 95  
 Saito's criterion, 33  
 scaled projective toric variety, 22  
 scattering amplitude, 86  
 scattering correspondence, 86  
 scattering potential, 78  
 semialgebraic set, 24  
 separator-based derivation, 95  
 Severi variety, 54  
 shallow network, 34  
 Smith normal form, 23  
 stabiliser, 38  
 stabiliser formalism, 37  
 strict linear precision, 50  
 symmetric algebra, 29

symmetric rank, 31  
symmetric tensor, 30  
  
tame, 33  
Terao conjecture, 33  
Terao module, 79  
TFP, *see* toric fibre product  
thick architecture, 102  
toric blending function, 49  
toric fibre product, 42  
toric patch, 49  
toric variety, 22  
total Milnor number, 59  
twisted cohomology, 73  
typical symmetric rank, 107  
  
undirected graphical model, 25  
  
very affine variety, 28  
von Neumann entropy, 127  
  
Whitney stratification, 55

# Bibliographische Daten

Nonlinear Algebra in Likelihood, Neurocomputing and Quantum Physics  
Wiesmann, Maximilian  
Universität Leipzig, Dissertation, 2025  
172 Seiten, 19 Abbildungen, 3 Tabellen, 201 Referenzen