GEOMETRY IN ALGORITHMS AND COMPLEXITY:

Holographic Algorithms and Valiant's Conjecture

Sitan Chen

submitted in partial fulfillment of the requirements for the degree of Bachelor of Arts with Honors in Mathematics and Computer Science

Advisor: Leslie Valiant

Math Advisor: Joseph Harris

Harvard University
April 1, 2016

Acknowledgments

Firstly, I would like to thank my advisor, Prof. Leslie Valiant. I owe so much to his inspiring guidance throughout the course of this project, for his suggestions on papers to read, and for his countless hours of helpful discussions and feedback.

I also thank Prof. Joseph Landsberg for posing the problem addressed in Chapter 5 of this thesis, for patiently answering questions I had about geometric complexity theory, and for helpful feedback on earlier versions of Chapters 4 and 5. I am grateful to Prof. Joseph Harris for being my math advisor for this thesis, for recommending that I reach out to Prof. Landsberg, and for teaching two superb courses in algebraic geometry that proved invaluable to this work. I would also like to thank Prof. Jin-Yi Cai for his incredibly extensive and invaluable feedback on the proof of my main result in Chapter 3. Lastly, I would like to thank Prof. Luke Oeding for sending me notes from his and collaborators' earlier attempted proof of Theorem 5.3.1 via the technique of Section 5.4.

More generally, I cannot overstate how much Harvard's mathematics and computer science departments have helped me grow as a student. Special thanks go to Prof. Salil Vadhan for putting me on the path to pursuing theoretical computer science research and for being a singularly inspiring mentor, and to Prof. Yum-Tong Siu for serving as my academic advisor and for teaching two incredible courses that really got me excited about several complex variables.

Needless to say, I am also deeply indebted to all the friends and mentors who have shaped my life in mathematics and beyond. I'm especially grateful for David Ding and Aaron Landesman for being frequent collaborators on problem sets and sporting victims to all of my silly questions about geometry and representation theory, Thomas Steinke for many hours of discussion and mentorship on a previous research project, Yakir Reshef for introducing me to complexity theory, Jesse Geneson and Tanya Khovanova for getting me excited about research in mathematics, and Janice Wong for inspiring me to pursue music. In addition, I'd like to thank Andrew Sanchez, Courtney Noh, Kevin Wang, Matt Rauen, and Michelle Deng for bearing with my endless rants about this thesis and for the countless trips to the MAC, Eliot grille, Qdoba, and Toscanini's that kept me sane during the course of this project.

Lastly, I would like to thank my mother and father for bringing me into this world and making everything possible. They were the reason I first fell in love with learning, and I am grateful every day for what they have done to raise me up to be the best version of myself.

Contents

Chapter 1

Introduction

1.1 History

Theoretical computer science is concerned with studying the ultimate capabilities of computing and can be divided into two areas: algorithms and complexity theory. The study of algorithms focuses on finding resource-efficient procedures for solving computational problems, while complexity theory focuses on proving that such problems require a certain amount of such computational resources as time, space, randomness, parallelism, etc. Put differently, whereas algorithms establish upper bounds on the hardness of problems, results in complexity theory give lower bounds.

The search for such upper and lower bounds has its origins in Hilbert's *Entscheidungsproblem*: find an algorithm that, given any statement in first-order logic, either produces a proof or concludes that none exists. Church and Turing notably showed that if Turing machines are a universal model of computation, this problem is undecidable [17, 52].

But computer scientists are typically concerned with settings where agents are resource-bounded, in which case the *Entscheidungsproblem* would ask: does there exist an *efficient* algorithm that, given any statement in first-order logic, either produces a *short* proof or concludes that no short proof exists? In passing to this finitary formulation, we effectively recover the famous P vs. NP problem, and indeed this was its original formulation in Gödel's famous 1956 letter to von Neumann. Equivalently, it asks whether the class of problems for which one can quickly verify a proposed solution (NP) is the same as the class of problems for which one can quickly find such a solution (P).

While it is strongly believed that $\mathsf{P} \neq \mathsf{NP}$, i.e. there is a significant lower bound on the time required to solve the hardest problems in NP, progress towards proving such a lower bound has been notoriously minimal, to the extent that some of the most famous theorems in this direction have been so-called "barrier" results showing that an overwhelming majority of techniques for proving lower bounds in complexity theory cannot separate P and NP [1, 3, 46]. And whereas people observed in the 1980's that proving $\mathsf{P} \neq \mathsf{NP}$ is equivalent to proving that the hardest problems in NP require exponentially large Boolean circuits to compute, the best known lower bound on circuit size for any problem [28] is only linear in the size of the input!

This lack of progress might compel an upper bounds enthusiast to seek new algorithmic techniques for solving problems in NP, or a lower bounds enthusiast to seek complexity theoretic techniques that evade the existing barrier results. To a reader of either persuasion, the relative lack of structure in the problem statement of P vs. NP makes it unclear where such techniques might come from. As we will see in this thesis, it turns out that some of the most promising techniques in both areas have been rooted in geometry.

1.2 Holographic Algorithms

Over the past fifty years, the CS theory community has accumulated a rich arsenal of deterministic polynomial-time algorithms, ranging from linear and semidefinite programming to primality testing and dynamic programming. In Part 1 of this thesis, we focus on one of the more recent additions to this collection, Valiant's framework of holographic algorithms. Originally motivated by the problem of classically simulating certain components of quantum computation [58], holographic algorithms have since been used to solve a number of counting problems that previously would have been conjectured to be intractable [59, 60, 62].

At the core of every holographic algorithm is the fact that the determinant of a matrix is easy to compute. Specifically, holographic algorithms exploit the fact that counting perfect matchings in planar graphs is as easy as computing determinants, a fact statistical physicists discovered some decades ago [29, 50]. The idea then is to show other counting problems are easy by reducing them to counting perfect matchings in planar graphs. To this end, a classical reduction would take a given problem instance ϕ and map it to some Ω for which the number of solutions to ϕ is in one-to-one correspondence with the number of perfect matchings of Ω, but the issue is that the range of problems for which this strategy would work is impossibly narrow. Instead, a holographic reduction takes ϕ to an Ω for which the number of solutions to ϕ agrees with the number of perfect matchings of Ω but for which no one-to-one correspondence necessarily exists. Holographic algorithms get their name from the way they allow multiple strands of computation to come together in a custom-built mixture reminiscent of quantum interference in order to produce the answer.

This mixture is made possible by a so-called *basis change*. Roughly speaking, $\mathrm{PerfMatch}(\Omega)$ can be realized as an inner product $\langle u, v\rangle$ for vectors u, v encoding the local constraints of ϕ, so applying a linear transformation to u and the dual transformation to v preserves this inner product, extending the range of vectors that can be used to encode ϕ's constraints. One of the primary questions in the past few years regarding the ultimate capabilities of holographic algorithms has been to understand the full power this change of basis affords.

In Part 1 of this thesis, we explore the problem of quantifying the power of basis change and see that this problem and holographic algorithms more generally have intriguing connections to the geometry of spinor varieties. By design, a problem admits a holographic solution if and only if a particular system of polynomial equations is solvable, and understanding the kinds of counting problems that can be solved by holographic algorithms turns out to be related to understanding the $\mathrm{SL}(n)$-orbits of spinor varieties.

1.3 Permanent vs. Determinant

On the complexity side of things, a starting point is to pass from the Boolean setting to an algebraic one. In 1979, Valiant [54, 55] introduced algebraic analogues of P and NP together with polynomials that captured these algebraic complexity classes, namely the determinant and permanent polynomials. Both are exponentially large sums of products, yet whereas the determinant is easy to compute, the permanent appears to be very difficult to compute.

Valiant posed the following algebraic analogue of $\mathsf{P} \neq \mathsf{NP}$:

Conjecture 1.3.1 (Valiant's conjecture)**.** The permanent of an $m \times m$ matrix M cannot be computed as the determinant of an $n \times n$ matrix M' whose entries are affine linear forms in those of M unless n is super-polynomial in m.

Whereas holographic algorithms take advantage of how special the determinant polynomial is in order to show certain problems are easy to solve, attempts to prove Valiant's conjecture try to take advantage of how special the determinant polynomial is in order to show other polynomials are very far from sharing the same properties.

After homogenizing, one can formulate Valiant's conjecture geometrically as a claim about the relationship between orbits of polynomials: for a given m, does the orbit of $\det_n$ under the action of $\mathrm{End}(n^2)$ contain that of $\ell^{n-m}\,\mathrm{perm}_m$ for some n super-polynomial in m? In the first half of Part 2 of this thesis, we explore some previous approaches to this problem.

One difficulty with trying to separate these orbits is that they are not cut out by polynomial equations, so naturally one can relax the problem by comparing the Zariski closures of these orbits instead. In [44], Mulmuley/Sohoni put forth an approach, their so-called *geometric complexity theory* (GCT) program, to separate the closures of these orbits by finding distinguishing representation-theoretic "obstructions" in their coordinate rings. Notably, their techniques seem to avoid all known barriers by making use of properties of the permanent and determinant that do not hold for general polynomials.

While passing to the closure puts the tools of algebraic geometry at one's disposal, the extent to which this relaxation differs from Valiant's original conjecture remains somewhat mysterious. Understanding this necessitates understanding the geometry of the boundaries of these orbit closures, which we explore in the second half of Part 2 of this thesis.

1.4 Our Contributions

In our exploration of holographic algorithms, our main result is to resolve the following decade-old open problem dating back to at least [59]: for a given domain size of counting problems (e.g. counting the number of k-colorings in a graph is a problem over domain size k), what is the smallest basis transformation matrix needed to simulate all holographic algorithms over that domain size? This is crucial to classifying what holographic algorithms are capable of accomplishing for counting problems over all domain sizes. Specifically, we prove the following in Chapter 3:

Theorem 1.4.1. *Any holographic algorithm over domain size k using a full rank signature can be simulated by a holographic algorithm using a $2^{\lceil \log_2 k \rceil} \times k$ basis matrix.*

A major step in our approach is to introduce a new coordinate-free interpretation of the algebraic properties characterizing holographic algorithms. A preliminary version of this result appeared in [15].

For free, this result also gives a new translation of the holographic framework over domain sizes $k = 2^K$ into the language of spinor varieties, extending the characterization given in [34] over domain size $k = 2$.

In our study of geometric techniques for separating the permanent and determinant, we introduce a new combinatorial technique for analyzing irreducible components on the boundary of $\overline{\mathrm{GL}(n^2) \cdot [\det_n]}$ and use this to completely classify boundary components for $n = 3$ in Section 5.4. This reproduces the following recent result that was coincidentally proven in [27] during the course of this project:

Theorem 1.4.2. *The only irreducible components of the $GL(9)$-orbit closure of det_3, are the component consisting of degenerate* $\mathrm{End}(9)$*-translates of det_3, and the $GL(9)$-orbit of the polarization*

$$\partial det_{2,1}(W_{skew}, W_{sym}),$$

where W_{skew} and W_{sym} are the generic skew-symmetric and symmetric 3×3 matrix respectively.

The third result of this work is to construct the first known infinite family of boundary components of Det_n for even n in Section 5.5. The rest of this thesis is expository, devoted to presenting what the author views as particularly beautiful connections between geometry and the P vs. NP problem.

1.5 Organization

In Chapter 1, we center our exposition around the three most substantial lower bounds for the determinantal complexity of the permanent, which draw on tools from classical differential geometry and highest weight theory, and briefly introduce Mulmuley/Sohoni's GCT program of finding representation-theoretic obstructions in coordinate rings.

In Chapter 2, we focus on a specific problem in algebraic geometry that arises from the GCT program: classifying components on the boundary of $\overline{\mathrm{GL}(n^2) \cdot [det_n]}$. We motivate this with a negative result of Kumar [31] that says this orbit closure is non-normal, so the *extension problem* of determining which functions on the orbit extend to the orbit closure is difficult. The boundary components problem turns out to be related to the long-standing problem of finding maximal linear subspaces on $V(\det_n)$, which we discuss at length. We then present the proof of Theorem 1.4.2 due to [27], which uses resolution of singularities, before presenting our own proof of the same result and our result on infinite families of boundary components.

In Chapter 3, we develop Valiant's holographic framework and introduce new coordinate-free interpretations of the algebraic properties characterizing holographic algorithms. In addition, we review the connections established in [34] between holographic algorithms over domain size 2 and the geometry of spinor varieties and their $\mathrm{SL}(n)$-orbits.

In Chapter 4, we present our proof of Theorem 1.4.1. In [34], Landsberg et al. gave an interpretation of holographic algorithms over the Boolean domain in terms of orbits of spinor varieties, and as a straightforward consequence of our Theorem 1.4.1, we are able to extend their interpretation to $k = 2^K$.

The only assumptions on background that this work makes are a rudimentary understanding of classical complexity theory, representation theory, invariant theory, and algebraic geometry. In the appendix, we include reviews of the basic complexity classes, the representation theory of the general linear group, the algebraic Peter-Weyl theorem, affine complex algebraic groups, the basic algebraic theory of spinors, and proofs of some minor details mentioned in the body of the thesis.

Part I

Algorithms

Chapter 2

Holographic Algorithms

Holographic algorithms were originally introduced by Valiant in [57, 58] as a method for classically simulating certain quantum gates in polynomial time. Roughly, they work by reducing counting problems to the problem of counting perfect matchings in planar graphs, which is known to be tractable. The unique aspect of these reductions is that they are "many-to-many," i.e. for a given problem instance, they do not establish a one-to-one bijection between the set of solutions and the corresponding set of perfect matchings, even though the quantities of both happen to agree. Instead, the intuition is that multiple strands of computation get combined in a "holographic" mixture with exponential, custom-built cancellations specified by a choice of basis vectors to produce the final answer.

Valiant asked whether matchgates could be used to derive other polynomial-time algorithms, and in recent years the framework has been applied to obtain a wide array of surprising polynomial-time algorithms for seemingly intractable problems [59, 60, 61, 62], the only criterion for their existence being the solvability of certain finite systems of polynomial equations. The subtext is that while there is an overwhelming consensus in the complexity theory community that $\mathsf{P} \neq \mathsf{NP}$, the justifications are somewhat tenuous for being based largely on the intuition that the algorithmic methods available to us for solving problems deterministically in polynomial time don't seem sufficient for solving NP-complete problems. Indeed, the problems holographic algorithms can solve would have been deemed intractable for such reasons prior to the introduction of holographic algorithms. Understanding the ultimate limitations of this framework therefore seems like a basic prerequisite for understanding why the separation $\mathsf{P} \neq \mathsf{NP}$ should hold.

In this chapter, we introduce the basic theory of holographic algorithms. We work through an illustrative example in Section 2.1 before formalizing the framework in Section 2.2. Then in Section 2.3, we discuss the defining algebraic properties of holographic algorithms and provide an original re-formulation of these properties that will prove essential to proving our main theorem in the next chapter. Finally, we give a geometric interpretation of holographic algorithms in terms of spinor varieties in Section 2.4, drawing upon ideas from [34].

2.1 Motivation

2.1.1 FKT algorithm

In [55], Valiant showed that the problem of computing the permanent for a matrix from $\mathcal{M}_{n\times n}(\{0,1\})$ is $\#\mathsf{P}$-complete. Because the number of perfect matchings in a bipartite graph is equal to the permanent of its adjacency matrix $M \in \mathcal{M}_{n\times n}(\{0,1\})$, this implies that counting the number of perfect matchings even in a bipartite graph is a $\#\mathsf{P}$-complete problem. More generally, for weighted directed graphs G, we can define the *perfect matching polynomial* $\mathrm{PerfMatch}(G)$ by

$$\mathrm{PerfMatch}(G) = \sum_{M} \prod_{(i,j)\in M} A^i_j,$$

where A denotes the (skew-symmetric) weighted adjacency matrix of G. The problem of computing PerfMatch is likewise #P-complete.

A famous result in statistical mechanics states however that the same problem for planar graphs can be solved in polynomial time. The eponymous algorithm due jointly to Fisher and Kasteleyn [29] and Temperley [50] takes advantage of the fact that 1) $\mathrm{PerfMatch}(G)$ looks quite similar to the polynomial

$$\mathrm{Pf}(A) = \sum_{M} \mathrm{sgn}(\pi_M) \prod_{(i,j)\in M} A^i_j, \tag{2.1}$$

called the *Pfaffian* of A, where $\mathrm{sgn}(\pi_M)$ denotes the sign of the permutation corresponding to perfect matching M, and 2) $\mathrm{Pf}(A)^2 = \det(A)$ [43], which we know how to compute quickly. Their observation was that for planar graphs, the edges can be directed in such a way that the Pfaffian and perfect matching polynomial agree.

Theorem 2.1.1 ([29, 50])**.** *Let G' be a weighted undirected graph. There exists an orientation of the edges giving rise to a directed graph G with adjacency matrix A for which* $\mathrm{PerfMatch}(G) = \mathrm{Pf}(A)$*. In particular, the problem of counting the number of perfect matchings in a planar graph can be solved in polynomial time.*

The FKT algorithm is a key ingredient in the holographic framework, under which one seeks to reduce counting problems to the problem of counting perfect matchings in planar graphs.

2.1.2 $\#_7$Pl-Rtw-Mon-3CNF

To illustrate the nature of these reductions before formally defining all the parts that go into the holographic approach, we provide a fairly informal exploration of one of its more notable success stories. That said, this subsection may be skipped if the reader prefers to dive straight into definitions.

Consider the canonical #P-complete problem of counting the number of satisfying assignments to a 3-CNF ϕ[1] If ϕ is *monotone*, i.e. if ϕ has no negations, we may associate to it a bipartite graph $G_\phi = (V_\phi, E_\phi)$ with a left vertex for every unique literal v, a right vertex for every clause C, and an edge connecting any v and C if v occurs in C.

To simplify things, we will insist that ϕ be monotone, *read-twice* (every literal appears at most twice in ϕ), and *planar* (G_ϕ has a planar embedding). We denote the problem of counting the number of satisfying assignments to such ϕ by #Pl-Rtw-Mon-3CNF. According to the following result whose proof we omit in this work, this is still intractable.

Theorem 2.1.2 ([65])**.** *#Pl-Rtw-Mon-3CNF is #P-complete.*

Even if we just ask for the parity of the number of satisfying assignments, this problem remains intractable. In general, denote the corresponding problem modulo k by $\#_k$Pl-Rtw-Mon-3CNF.

Theorem 2.1.3 ([59])**.** *$\#_2$Pl-Rtw-Mon-3CNF is $\bigoplus$P-complete.*

One of the early miracles of the holographic approach was the following:

Theorem 2.1.4 ([59, 12])**.** *$\#_7$Pl-Rtw-Mon-3CNF has a polynomial-time solution.*

We now sketch a proof of this. Intuitively, imagine that every left vertex v in G_ϕ emits signals along the one or two edges (called *wires*) connected to it indicating whether the corresponding literal in ϕ is assigned 0 or 1. $\#_7$Pl-Rtw-Mon-3CNF is characterized by two local constraints on these signals: 1) for each left vertex v, the signals along the edges connected to v must agree, 2) for each right vertex C, at least one of the three incoming signals must be 1 in order for clause C in ϕ to be satisfied.

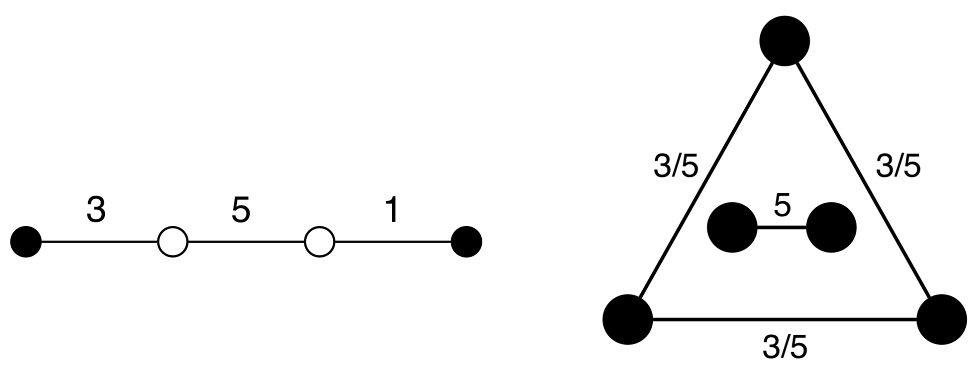
We'd like to replace the left and right vertices of G_ϕ respectively with small planar graphs G and R, called *matchgates*, whose perfect matching properties encode these local constraints in such a way that the number of perfect matchings of the graph G'_ϕ obtained after these replacements is equal to the number of satisfying assignments to ϕ.

[1]Recall that a 3-CNF is a Boolean formula ϕ of the form $\bigwedge_i (x^i_1 \vee x^i_2 \vee x^i_3)$, where x^i_j are (not necessarily distinct) $\{0,1\}$-valued variables, and $\vee$ and $\wedge$ denote the Boolean OR and AND operations. A *satisfying assignment* is an assignment of $\{0,1\}$ to each variable for which ϕ evaluates to 1.

So for any choice of signals $x \in \{0,1\}^{|E_\phi|}$ along the wires of G_ϕ, denote by Z the set of wires w_i of G_ϕ for which $x_i = 1$. Consider the induced subgraph $G'_\phi \backslash Z$. We must have that $\text{PerfMatch}(G'_\phi \backslash Z)$ equals 1 if x satisfies the two local conditions and 0 otherwise. But $\text{PerfMatch}(G'_\phi \backslash Z)$ is the product of all copies of $\text{PerfMatch}(G \backslash Z)$ and $\text{PerfMatch}(R \backslash Z)$ across left vertices v and right vertices C. This vanishes if and only if $\text{PerfMatch}(G \backslash Z) = 0$ and/or $\text{PerfMatch}(R \backslash Z) = 0$. On the other hand, if $\text{PerfMatch}(G \backslash Z) = \text{PerfMatch}(R \backslash Z) = 1$, then $\text{PerfMatch}(G'_\phi \backslash Z) = 1$.

Then it suffices to construct a G for which the number of perfect matchings upon removing either both vertices or neither vertex incident to a wire is 1, and upon removing exactly one vertex incident to a wire is 0; and an R for which the number of perfect matchings upon removing at least one vertex incident to a wire is 1, and upon removing no vertices is 0. We would then say that such graphs G and R have *standard signatures* $(1,0,0,1)$ and $(1,1,1,1,1,1,1,0)^T$.

The issue is that the latter vector cannot possibly be a standard signature: removing an odd number of vertices and an even number of vertices cannot both give a nonzero number of perfect matchings. This is the so-called *parity condition* of standard signatures. But the situation is still salvageable: Valiant's insight was to use a change of basis to extend the range of feasible signatures. The intuition is that $\text{PerfMatch}(\Omega)$ can be regarded as a vector pairing $\langle u, v\rangle$ between a tensor product $u \in \mathbb{C}^{\otimes 2^n} \simeq (\mathbb{C}^2)^{\otimes n}$ of copies of $(1,0,0,1)$ and a tensor product $v \in \mathbb{C}^{*2^n} \simeq (\mathbb{C}^{*2})^{\otimes n}$ of copies of $(1,1,1,1,1,1,1,0)$. Transforming u by the action of some *basis matrix* $M \in \text{GL}_2(\mathbb{C})$ and v by the dual action preserves the pairing, so in fact it suffices to find an appropriate M taking both G and R to valid standard signatures. In [12], Cai and Lu found that remarkably, this is possible over $\mathcal{F}_7$: take $M = \begin{pmatrix} 1 & 3 \\ 6 & 5 \end{pmatrix}$ and standard signatures $(3,0,0,5)$ and $(0,3,3,0,3,0,0,5)$, corresponding to the matchgates in Figures 2.1a and 2.1b.



(a) Standard signature $(3,0,0,5)$ (b) Standard signature $(0,3,3,0,3,0,0,5)$

Figure 2.1: Gadgets for $\#_7$Pl-Rtw-Mon-3CNF

The point Valiant makes in [57, 58] is that the same reasoning by which people believe $\mathsf{P} \neq \mathsf{NP}$, namely that none of the existing frameworks for obtaining efficient algorithms can solve an NP-complete problem, would have suggested that $\#_7$Pl-Rtw-Mon-3CNF is intractable prior to the introduction of holographic algorithms. This suggests that to arrive upon the desired separation of P and NP, we need a better understanding of the possibilities of polynomial-time computation. For this reason, determining the ultimate capabilities of holographic algorithms appears to be a crucial step.

2.2 Preliminaries

In this section, we formalize the notions introduced in the previous section.

2.2.1 Background

Denote the Hamming weight of string α by $\mathrm{wt}(\alpha)$, and define the parity of α to be the parity of $\mathrm{wt}(\alpha)$. Given $1 \le i \le m$, define $e_i \in \{0,1\}^m$ to be the bitstring with a single nonzero bit in position i. The parameter m is implicit, and when this notation is used, m will be clear from the context. Denote by 1^m the length-m bitstring consisting solely of 1's.

We review some basic definitions and results about holographic algorithms. For a comprehensive introduction to this subject, see [60].

Definition 2.2.1. A *matchgate* $\Gamma = (G, X, Y)$ is defined by a planar embedding of a planar graph $G = (V, E, W)$, *input nodes* $X \subseteq V$, and *output nodes* $Y \subseteq V$, where $X \cap Y = \varnothing$. We refer to $X \cup Y$ as the *external nodes* of Γ.

We say that Γ has *arity* $|X| + |Y|$. In the planar embedding of G, the input and output nodes are arranged such that if one travels counterclockwise around the outer face of G, one encounters first the input nodes labeled 1,2,...,$|X|$ and then the output nodes $|Y|$,...,2,1.

If Γ has exclusively output (resp. input) nodes, we say that Γ is a generator (resp. recognizer). Otherwise, we say that Γ is an *$|X|$-input, $|Y|$-output transducer.*

Definition 2.2.2. A *basis matrix* with *basis size* ℓ over *domain size* k is a $2^\ell \times k$ matrix $M = (a_i^\alpha)$, where rows and columns are indexed by $\alpha \in \{0,1\}^\ell$ and $i \in [k]$ respectively. The domain size should be interpreted as the range over which variables in the counting problem can take values, so for instance, problems related to counting certain k-colorings in a graph are problems over domain size k. The basis size should then be interpreted roughly as the number of bits needed to encode each of these k colors.

Definition 2.2.3. The *standard signature* of a matchgate Γ of arity $n\ell$ is a vector of dimension $2^{n\ell}$ which will be denoted by $\underline{\Gamma}$, where for $\alpha_i \in \{0,1\}^\ell$, $\underline{\Gamma}^{\alpha_1 \cdots \alpha_n}$ denotes the entry of $\underline{\Gamma}$ indexed by $\alpha_1 \circ \cdots \circ \alpha_n$. If Z is the subset of the external nodes of Γ for which $\alpha_1 \circ \cdots \circ \alpha_n$ is the indicator string, then

$$\underline{\Gamma}^{\alpha_1 \cdots \alpha_n} = \mathrm{PerfMatch}(\Gamma \backslash Z).$$

Here, if $A = (A_{ij})$ is the adjacency matrix of Γ, PerfMatch is the polynomial in the entries of A defined by

$$\mathrm{PerfMatch}(A) = \sum_M \prod_{(i,j) \in M} A_{ij},$$

with the sum taken over the set M of all perfect matchings of Γ.

The following lemma follows from the definition of standard signatures.

Lemma 2.2.4. *Suppose $\underline{R}$ is the standard signature of a recognizer of arity $n\ell$ and T the standard signature of a transducer with s inputs and ℓ outputs. Then $\underline{R}' = \underline{R}T^{\otimes n}$ is the standard signature of a recognizer matchgate of arity ns.*

Definition 2.2.5. A column (resp. row) vector of dimension k^n is said to be a *generator (resp. recognizer) signature realizable over a basis M* if there exists a generator (resp. recognizer) matchgate Γ satisfying $M^{\otimes n} G = \underline{G}$ (resp. $\underline{R} M^{\otimes n} = R$). We say that a collection of recognizer and generator signatures $R_1, \ldots, R_a, G_1, \ldots, G_b$ is *simultaneously realizable* if they are realizable over a common basis M.

In particular, if M is square, the signature of a matchgate with respect to the standard basis is the standard signature. Also note that in terms of coordinates, we have that

$$\underline{G}^{\alpha_1 \cdots \alpha_n} = \sum_{j_1, \ldots, j_n \in [k]} G^{j_1 \cdots j_n} a_{j_1}^{\alpha_1} \cdots a_{j_n}^{\alpha_n}$$

and

$$R_{j_1 \cdots j_n} = \sum_{\alpha_1, \ldots, \alpha_n \in \{0,1\}^\ell} \underline{R}_{\alpha_1 \cdots \alpha_n} a_{j_1}^{\alpha_1} \cdots a_{j_n}^{\alpha_n}.$$

Definition 2.2.6. A *matchgrid* $\Omega = (G, R, W)$ is a weighted planar graph consisting of a set of g generators $G = \{G_1, ..., G_g\}$, a set of r recognizers $R = \{R_1, ..., R_r\}$, and a set of w wires $W = \{W_1, ..., W_w\}$, each of which has weight 1 and connects the output node of a generator to the input node of a recognizer so that every input and output node among the matchgates in $G \cup R$ lies on exactly one wire.

Define the *underlying graph* of Ω to be the graph with $g + r$ vertices and w edges constructed from Ω by replacing each matchgate with a vertex and each edge between external nodes of matchgates G_i and R_j with an edge between the new pair of vertices corresponding to these matchgates.

Definition 2.2.7. Suppose $\Omega = (G, R, W)$ is a matchgrid with g generators, r recognizers, and w wires, and let M be a basis for Ω. Define the *Holant* to be the following quantity:

$$\text{Holant}(\Omega) = \sum_{z \in [k]^w} \left(\prod_{i=1}^{g} G_i^{y_i} \prod_{j=1}^{r} R_j^{x_j} \right).$$

Here, $z = y_1 \circ \cdots \circ y_g = x_1 \circ \cdots \circ x_r$ such that $y_i \in [k]^{|Y_i|}$ and $x_j \in [k]^{|X_j|}$ for Y_i the output nodes of G_i and X_j the input nodes of R_j, and G_i and R_j denote the signatures of their respective matchgates under basis M.

Valiant's Holant theorem states the following.

Theorem 2.2.8 (Theorem 4.1, [60])**.** *If Ω is a matchgrid over a basis M, then* $\text{Holant}(\Omega) = \text{PerfMatch}(\Omega)$.

As the Fisher-Kasteleyn-Temperley algorithm [29, 50] can compute the number of perfect matchings of a planar graph in polynomial time, $\text{Holant}(\Omega)$ can be computed in polynomial time as long as Ω is planar.

The following observation about the Holant, alluded to in Section 2.1.2, will prove useful in Section 2.4 when we give a geometric interpretation of the holographic framework.

Observation 1. $\text{Holant}(\Omega)$ is a vector pairing $\langle \mathbf{G}, \mathbf{R} \rangle$, where $\mathbf{G} = \otimes G_i$ and $\mathbf{R} = \otimes R_i$. Here, the tensor products are ordered in a way compatible with the wiring of Ω.

2.2.2 Matrix form of signatures

It will be convenient to regard signatures not as vectors but as matrices.

Definition 2.2.9. For generator signature G, the t-th matrix form $G(t)$ $(1 \le t \le n)$ is a $k \times k^{n-1}$ matrix where the rows are indexed by $1 \le j_t \le k$ and the columns are indexed by $j_1 \cdots j_{t-1} j_{t+1} \cdots j_n$ in lexicographic order.

Definition 2.2.10. For recognizer signature R, the t-th matrix form $R(t)$ $(1 \le t \le n)$ is a $k^{n-1} \times k$ matrix where the rows are indexed by $j_1 \cdots j_{t-1} j_{t+1} \cdots j_n$ in lexicographic order and the columns are indexed by $1 \le j_t \le k$.

We would also like to regard standard signatures as matrices; if basis M is square, the following definitions are special cases of the above.

Definition 2.2.11. For standard signature $\underline{G}$, the t-th matrix form $\underline{G}(t)$ $(1 \le t \le n)$ is a $2^\ell \times 2^{(n-1)\ell}$ matrix where the rows are indexed by α_t and the columns are indexed by $\alpha_1 \cdots \alpha_{t-1} \alpha_{t+1} \cdots \alpha_n$.

Definition 2.2.12. For standard signature $\underline{R}$, the t-th matrix form $\underline{R}(t)$ $(1 \le t \le n)$ is a $2^{(n-1)\ell} \times 2^\ell$ matrix where the rows are indexed by $\alpha_1 \cdots \alpha_{t-1} \alpha_{t+1} \cdots \alpha_n$ and the columns are indexed by α_t.

One can check that $\underline{G}(t)$ and $G(t)$, and $\underline{R}(t)$ and $R(t)$, are related as follows.

Lemma 2.2.13. *If $\underline{G} = M^{\otimes n} G$, then*

$$\underline{G}(t) = M G(t) (M^T)^{\otimes (n-1)}.$$

Lemma 2.2.14. *If* $R = \underline{R}M^{\otimes n}$*, then*

$$R(t) = (M^T)^{\otimes(n-1)}\underline{R}(t)M.$$

We will denote by $\underline{G}(t)^\sigma$ the row vector indexed by σ, $\underline{G}(t)_\zeta$ the column vector indexed by ζ, and $\underline{G}(t)^\sigma_\zeta$ the entry of $\underline{G}$ in row σ and column ζ. We use analogous notation for matrices $\underline{R}$, G, and R. In general, if Γ is any matrix, we will sometimes refer to the entry Γ^σ_ζ as the "entry (indexed by) (σ,ζ)."

In general, if Γ is a matrix with rows indexed by $\{0,1\}^a$ and columns indexed by $\{0,1\}^b$, and $S \subset \{0,1\}^a$ (resp. $S \subset \{0,1\}^b$), we will let Γ^S (resp. Γ_S) denote the submatrix of Γ consisting of rows (resp. columns) indexed by S. Where Γ is clear from context, we will denote the row span of Γ^S (resp. column span of Γ_S) by $\text{span}(S)$.

Lastly, a column/row is called *odd* (resp. *even*) if its index is odd (resp. even).

2.2.3 Degenerate and full rank signatures

Definition 2.2.15. A signature G (generator or recognizer) is degenerate iff there exist vectors γ_i $(1 \le i \le n)$ of dimension k for which $G = \gamma_1 \otimes \cdots \otimes \gamma_n$.

Lemma 2.2.16 (Lemma 3.1, [8])**.** *A signature G is degenerate iff* $\text{rank}(G(t)) \le 1$ *for* $1 \le t \le n$*.*

Definition 2.2.17. A signature G is of full rank iff there exists some $1 \le t \le n$ for which $\text{rank}(G(t)) = k$.

By Lemma 2.2.13, if signature G is of full rank, then for the corresponding standard signature $\underline{G}$, we have that $\text{rank}(\underline{G}(t)) = k$ for some t. Over domain size 2, by Lemma 2.2.16, all signatures not of full rank are degenerate, and holographic algorithms exclusively using such signatures are trivial because degenerate generators can by definition be decoupled into arity-1 generators. Over domain size $k \ge 3$ however, it is unknown to what extent holographic algorithms exclusively using signatures not of full rank trivialize. In [8], the collapse theorems over domain sizes 3 and 4 were proved under the assumption that at least one signature is of full rank, so we too make that assumption.

2.2.4 Clusters

One of the key results in our proof of the general collapse theorem is the existence within any matrix-form standard signature of a full-rank square submatrix whose entries have indices satisfying certain properties. In this section we make precise those properties.

Definition 2.2.18. A set of 2^m distinct bitstrings $Z = \{x^1, ..., x^{2^m}\} \subset \{0,1\}^n$ is an (m,n)*-cluster* if there exists $s \in \{0,1\}^n$ and positions $p_1, ..., p_m \in [n]$ such that for each $i \in [2^m]$, $x^i = s \oplus \left(\bigoplus_{j\in J} e_{p_j}\right)$ for some $J \subset \{p_1, ..., p_m\}$. We write Z as $s + \{e_{p_1}, ..., e_{p_m}\}$. Note that for a fixed cluster, s is only unique up to the bits outside of positions $p_1, ..., p_m$. If a cluster Z' is a subset of another cluster Z, we say that Z' is a *subcluster* of Z.

Definition 2.2.19. In Γ, a $2^m \times 2^m$ submatrix Γ' is a m*-cluster submatrix* if there exist (m,ℓ)- and $(m,(n-1)\ell)$-clusters Σ and Z such that $\Gamma' = (\Gamma^\sigma_\zeta)^{\sigma\in\Sigma}_{\zeta\in Z}$ (here we omit the parameters n and ℓ in the notation as they will be clear from context).

2.3 Matchgate identities and consequences

2.3.1 Parity condition and matchgate identities

The most obvious property that standard signatures satisfy is the *parity condition*: because a graph with an odd number of vertices has no perfect matchings, the indices of the nonzero entries in $\underline{G}$ have the same parity. It trivially follows that in $\underline{G}(t)$, columns $\underline{G}(t)_\zeta$ and $\underline{G}(t)_\eta$, if both nonzero, are linearly independent if ζ and η are of opposite parities. The same holds for the rows of $\underline{G}(t)$.

In [9] it is shown algebraically that the parity condition in fact follows from the so-called *matchgate identities* which we present below, quadratic relations which together form a necessary and sufficient condition for a vector to be the standard signature of some matchgate.

As in Cai and Fu's proof of the collapse theorem for domain size 4, we will make heavy use of the matchgate identities stated below. Wherever we invoke them, they will be for generator matchgates, so we focus on this case.

Theorem 2.3.1 (Theorem 2.1, [9])**.** *A $2^\ell \times 2^{(n-1)\ell}$ matrix Γ is the t-th matrix form of the standard signature of some generator matchgate iff for all $\zeta, \eta \in \{0,1\}^{(n-1)\ell}$ and $\sigma, \tau \in \{0,1\}^\ell$, the following* matchgate identity *holds. Let $\zeta \oplus \eta = e_{q_1} \oplus \cdots \oplus e_{q_{d'}}$ and $\sigma \oplus \tau = e_{p_1} \oplus \cdots \oplus e_{p_d}$, where $q_1 < \cdots < q_{d'}$ and $p_1 < \cdots < p_d$. Then*

$$\sum_{i=1}^{d}(-1)^{i+1}\Gamma_{\zeta}^{(\sigma\oplus e_{p_1}\oplus e_{p_i})}\Gamma_{\eta}^{(\tau\oplus e_{p_1}\oplus e_{p_i})} = \sum_{j=1}^{d'} \pm\Gamma_{(\zeta\oplus e_{q_j})}^{(\sigma\oplus e_{p_1})}\Gamma_{(\eta\oplus e_{q_j})}^{(\tau\oplus e_{p_1})}. \tag{2.2}$$

Here the $\pm$ signs depend on both j and, if q_j is after the t-th block, the parity of d. If d is even,

$$\sum_{i=1}^{d}(-1)^{i+1}\Gamma_{\zeta}^{(\sigma\oplus e_{p_1}\oplus e_{p_i})}\Gamma_{\eta}^{(\tau\oplus e_{p_1}\oplus e_{p_i})} = \epsilon_{\zeta,\eta}\sum_{j=1}^{d'} (-1)^{j+1}\Gamma_{(\zeta\oplus e_{q_j})}^{(\sigma\oplus e_{p_1})}\Gamma_{(\eta\oplus e_{q_j})}^{(\tau\oplus e_{p_1})}, \tag{2.3}$$

where $\epsilon_{\zeta,\eta} \in \{\pm 1\}$ is positive (resp. negative) if the number of q_j preceding the t-th block is odd (resp. even).

Remark 2.3.2. If $d + d'$ is odd, (2.2) is trivial by the parity condition.

We will be making extensive use of the matchgate identities in this paper, but we will typically not care about the $\epsilon_{\zeta,\eta}$ sign on the right-hand side of (2.3). For this reason, it will be convenient to make the following definition.

Definition 2.3.3. A $2^\ell \times 2^m$ matrix M is a *pseudo-signature* if for all σ, τ for which $\mathrm{wt}(\sigma \oplus \tau)$ is even, its entries satisfy the corresponding identity (2.3) up to a factor of ± 1 on the right-hand side.

Standard signatures and cluster submatrices are all examples of pseudo-signatures.

Observation 2. If M is a pseudo-signature, then its transpose M^T is a pseudo-signature.

2.3.2 Matchgate identities and determinants

We now derive from the matchgate identities some basic linear algebraic properties of the columns of pseudo-signatures. By Observation 2, these also apply for the rows.

Firstly, we have the following immediate consequence of Theorem 2.3.1. We specifically consider the case where $\mathrm{wt}(\zeta \oplus \eta)$ is even and, by Remark 2.3.2, $\mathrm{wt}(\sigma \oplus \tau)$ is even. So write $\zeta \oplus \eta = e_{q_1} \oplus \cdots \oplus e_{q_{2d'}}$ and $\sigma \oplus \tau = e_{p_1} \oplus \cdots \oplus e_{p_{2d}}$.

Reverse the roles of ζ and η in (2.3). Subtract the resulting equation from (2.3) to find

$$\sum_{i=1}^{2d}(-1)^{i+1}\left(\Gamma_{\eta}^{(\sigma\oplus e_{p_1}\oplus e_{p_i})}\Gamma_{\zeta}^{(\tau\oplus e_{p_1}\oplus e_{p_i})} - \Gamma_{\zeta}^{(\sigma\oplus e_{p_1}\oplus e_{p_i})}\Gamma_{\eta}^{(\tau\oplus e_{p_1}\oplus e_{p_i})}\right) =$$

$$\epsilon_{\zeta,\eta}\sum_{j=1}^{2d'}(-1)^{j+1}\left(\Gamma_{(\zeta\oplus e_{q_j})}^{(\sigma\oplus e_{p_1})}\Gamma_{(\eta\oplus e_{q_j})}^{(\tau\oplus e_{p_1})} - \Gamma_{(\eta\oplus e_{q_j})}^{(\sigma\oplus e_{p_1})}\Gamma_{(\zeta\oplus e_{q_j})}^{(\tau\oplus e_{p_1})}\right),$$

or equivalently,

$$\sum_{i=1}^{2d}(-1)^{i+1}\begin{vmatrix}\Gamma_{\zeta}^{(\sigma\oplus e_{p_1}\oplus e_{p_i})} & \Gamma_{\eta}^{(\sigma\oplus e_{p_1}\oplus e_{p_i})}\\ \Gamma_{\zeta}^{(\tau\oplus e_{p_1}\oplus e_{p_i})} & \Gamma_{\eta}^{(\tau\oplus e_{p_1}\oplus e_{p_i})}\end{vmatrix} = \epsilon_{\zeta,\eta}\sum_{j=1}^{2d'}(-1)^{j+1}\begin{vmatrix}\Gamma_{(\zeta\oplus e_{q_j})}^{(\sigma\oplus e_{p_1})} & \Gamma_{(\eta\oplus e_{q_j})}^{(\sigma\oplus e_{p_1})}\\ \Gamma_{(\zeta\oplus e_{q_j})}^{(\tau\oplus e_{p_1})} & \Gamma_{(\eta\oplus e_{q_j})}^{(\tau\oplus e_{p_1})}\end{vmatrix} \tag{2.4}$$

Example 2.3.4. Suppose $d = d' = 1$. Then (2.4) becomes

$$2\begin{vmatrix}\Gamma^{\sigma}_{\zeta} & \Gamma^{\sigma}_{\eta}\\ \Gamma^{\tau}_{\zeta} & \Gamma^{\tau}_{\eta}\end{vmatrix} = 2\epsilon_{\zeta,\eta}\begin{vmatrix}\Gamma^{\sigma\oplus e_{p_1}}_{\zeta\oplus e_{q_1}} & \Gamma^{\sigma\oplus e_{p_1}}_{\zeta\oplus e_{q_2}}\\ \Gamma^{\sigma\oplus e_{p_2}}_{\zeta\oplus e_{q_1}} & \Gamma^{\sigma\oplus e_{p_2}}_{\zeta\oplus e_{q_2}}\end{vmatrix},$$

so in particular, the matrix on the left is singular iff the latter is.

More generally, only suppose that $d = 1$. Then (2.4) becomes

$$2\begin{vmatrix}\Gamma^{\sigma}_{\zeta} & \Gamma^{\sigma}_{\eta}\\ \Gamma^{\tau}_{\zeta} & \Gamma^{\tau}_{\eta}\end{vmatrix} = \epsilon_{\zeta,\eta}\sum_{j=1}^{2d'}(-1)^{j+1}\begin{vmatrix}\Gamma^{(\sigma\oplus e_{p_1})}_{(\zeta\oplus e_{q_j})} & \Gamma^{(\sigma\oplus e_{p_1})}_{(\eta\oplus e_{q_j})}\\ \Gamma^{(\sigma\oplus e_{p_2})}_{(\zeta\oplus e_{q_j})} & \Gamma^{(\sigma\oplus e_{p_2})}_{(\eta\oplus e_{q_j})}\end{vmatrix} \tag{2.5}$$

so in particular, the matrix on the left-hand side is singular if all $2d'$ matrices on the right-hand side are singular.

In other words, if $\ell = 2$ so that Γ only has four rows, columns ζ and η as defined above are linearly dependent if all pairs of neighboring columns are linearly dependent. We shall see in the next section (Corollary 2.3.6) that this is true even when Γ has an arbitrary number of rows.

2.3.3 Wedge products of columns

Motivated by Example 2.3.4, we'd like to study the set of all 2×2 determinants $\begin{vmatrix}\Gamma^{\sigma}_{\zeta} & \Gamma^{\sigma}_{\eta}\\ \Gamma^{\tau}_{\zeta} & \Gamma^{\tau}_{\eta}\end{vmatrix}$ given two column vectors Γ_ζ and Γ_η of the same parity. These are merely the coefficients of the wedge product $\Gamma_\zeta \wedge \Gamma_\eta$ under the standard basis $\{v_\sigma \wedge v_\tau\}_{\sigma,\tau\in\{0,1\}^{2^\ell},\ \sigma<\tau}$ (where the relation $\sigma < \tau$ denotes lexicographic ordering) of $\Lambda^2\mathbb{C}^{2^\ell}$, the second exterior power of $\mathbb{C}^{2^\ell}$. The matchgate identities imply the following consequence about the relationships among the wedge products $\Gamma_\zeta \wedge \Gamma_\eta$ as ζ and η vary.

Lemma 2.3.5. *If $\zeta_1, \eta_1, \zeta_2, \eta_2, \ldots, \zeta_m, \eta_m$ are even indices for which*

$$\sum_{\nu=1}^{m} a_\nu\left(\Gamma_{\zeta_\nu} \wedge \Gamma_{\eta_\nu}\right) = 0 \tag{2.6}$$

for some $a_1, \ldots, a_m \in \mathbb{C}$, then

$$\sum_{\nu=1}^{m} \epsilon_{\zeta_\nu,\eta_\nu} a_\nu\left(\sum_{i=1}^{2d_\nu}(-1)^{j+1}\Gamma_{\zeta_\nu\oplus e_{p_i^\nu}} \wedge \Gamma_{\eta_\nu\oplus e_{p_i^\nu}}\right) = 0, \tag{2.7}$$

where for each ν, $\mathrm{wt}(\zeta_\nu \oplus \eta_\nu) = 2d_\nu$ and $\zeta_\nu \oplus \eta_\nu = e_{p_1^\nu} \oplus \cdots \oplus e_{p_{2d_\nu}^\nu}$.

Proof. For convenience, we will denote $\epsilon_{\zeta_\nu,\eta_\nu}$ by ϵ_ν. First, we rewrite (2.7) in terms of coordinates as

$$\sum_{\nu=1}^{m}\epsilon_\nu a_\nu\left(\sum_{i=1}^{2d_\nu}(-1)^{i+1}\sum_{\sigma<\tau}\begin{vmatrix}\Gamma^{\sigma}_{\zeta_\nu\oplus e_{p_i^\nu}} & \Gamma^{\sigma}_{\eta_\nu\oplus e_{p_i^\nu}}\\ \Gamma^{\tau}_{\zeta_\nu\oplus e_{p_i^\nu}} & \Gamma^{\tau}_{\eta_\nu\oplus e_{p_i^\nu}}\end{vmatrix}(v_\sigma\wedge v_\tau)\right) = 0, \tag{2.8}$$

where $\sigma < \tau$ denotes lexicographical ordering. Note that the determinants that appear in the left-hand side of (2.8) are zero when σ and τ are of opposite parity. Moreover, depending on the parity of the signature Γ, either all such determinants are also zero for σ and τ even, or they are all zero for σ and τ odd.

Rearranging the order of summations in (2.8), the desired identity becomes

$$\sum_{\sigma<\tau}(v_\sigma\wedge v_\tau)\cdot\left(\sum_{\nu=1}^{m}\epsilon_\nu a_\nu\sum_{i=1}^{2d_\nu}(-1)^{i+1}\begin{vmatrix}\Gamma^{\sigma}_{\zeta_\nu\oplus e_{p_i^\nu}} & \Gamma^{\sigma}_{\eta_\nu\oplus e_{p_i^\nu}}\\ \Gamma^{\tau}_{\zeta_\nu\oplus e_{p_i^\nu}} & \Gamma^{\tau}_{\eta_\nu\oplus e_{p_i^\nu}}\end{vmatrix}\right) = 0. \tag{2.9}$$

For a fixed pair $\sigma < \tau$, let $\mathrm{wt}(\sigma\oplus\tau) = 2d'$ and $\sigma\oplus\tau = e_{q_1}\oplus\cdots\oplus e_{q_{2d'}}$. If we apply (2.4) and rearrange the order of summation once more, the coefficient of $v_\sigma \wedge v_\tau$ above becomes

$$\sum_{\nu=1}^{m} a_\nu\cdot\sum_{j=1}^{2d'}(-1)^{j+1}\begin{vmatrix}\Gamma^{\sigma\oplus e_{q_j}}_{\zeta_\nu} & \Gamma^{\sigma\oplus e_{q_j}}_{\eta_\nu}\\ \Gamma^{\tau\oplus e_{q_j}}_{\zeta_\nu} & \Gamma^{\tau\oplus e_{q_j}}_{\eta_\nu}\end{vmatrix} = \sum_{j=1}^{2d'}(-1)^{j+1}\sum_{\nu=1}^{m} a_\nu\begin{vmatrix}\Gamma^{\sigma\oplus e_{q_j}}_{\zeta_\nu} & \Gamma^{\sigma\oplus e_{q_j}}_{\eta_\nu}\\ \Gamma^{\tau\oplus e_{q_j}}_{\zeta_\nu} & \Gamma^{\tau\oplus e_{q_j}}_{\eta_\nu}\end{vmatrix}.$$

But note that if we expand (2.6) in terms of coordinates, the term

$$a_\nu \begin{vmatrix} \Gamma^{\sigma\oplus e_{q_j}}_{\zeta_\nu} & \Gamma^{\sigma\oplus e_{q_j}}_{\eta_\nu} \\ \Gamma^{\tau\oplus e_{q_j}}_{\zeta_\nu} & \Gamma^{\tau\oplus e_{q_j}}_{\eta_\nu} \end{vmatrix}$$

is precisely the coefficient of $v_{\sigma\oplus e_{q_j}} \wedge v_{\tau\oplus e_{q_j}}$ in the expansion of (2.6) in terms of coordinates and hence zero by assumption, so (2.9) holds as desired. ☐

Corollary 2.3.6. *Let $\zeta,\eta \in \{0,1\}^{(n-1)\ell}$ be such that $\zeta\oplus\eta = \bigoplus_{j=1}^{2d} e_{p_i}$. If column $\Gamma_{(\zeta\oplus e_{p_i})}$ is linearly dependent with column $\Gamma_{(\eta\oplus e_{p_i})}$ for $1 \le i \le 2d$, then column Γ_ζ is linearly dependent with column Γ_η.*

Corollary 2.3.7. *Let $\zeta,\eta \in \{0,1\}^{(n-1)\ell}$ be such that $\zeta\oplus\eta = \bigoplus_{i=1}^{2d} e_{p_i}$. If there exists $i \in [2d']$ such that column $\Gamma_{\zeta\oplus e_{p_i}}$ is linearly dependent with column $\Gamma_{\eta\oplus e_{p_i}}$ for $i = 1,...,\hat{j},...,2d'$, where $\hat{j}$ denotes omission of index j, and if Γ_ζ is also linearly dependent with column Γ_η, then $\Gamma_{\zeta\oplus e_{p_j}}$ and $\Gamma_{\eta\oplus e_{p_j}}$ are linearly dependent.*

Lemma 2.3.5 says that any linear relation among wedges of even columns yields a linear relation among wedges of odd columns, and vice versa.

Lastly, we need the following elementary result in multilinear algebra.

Lemma 2.3.8. *If $v_1,...,v_n$ are linearly independent in vector space V, then the set of all $v_i \wedge v_j$ for $i < j$ are linearly independent in $\Lambda^2 V$.*

Combining this with Lemma 2.3.5 yields the following key ingredient to the analysis in Section 3.3.

Lemma 2.3.9. *Suppose $\zeta_0,\eta \in \{0,1\}^K$ such that $\zeta_0 \ne \eta$, and the indices in $T = \{\zeta_1,...,\zeta_m\} \subset \{0,1\}^K$ are distinct and have the same parity. Suppose further that $\zeta_0 \ne \zeta_1,...,\zeta_m$. Let $\zeta_i\oplus\eta = e_{p^i_1} \oplus\cdots\oplus e_{p^i_{d_i}}$ for $0 \le i \le m$, where $d_i := \mathrm{wt}(\zeta_i \oplus \eta)$. Define*

$$S := \bigcup_{0\le i\le m, 1\le j\le d_i} \{\eta\oplus e_{p^i_j}, \zeta_i \oplus e_{p^i_j}\} \subset \{0,1\}^K$$

not in the sense of multisets, that is, we throw out duplicates so that the strings in S are all distinct.

Suppose the columns indexed by S are linearly independent. Then

$$\Gamma_{\zeta_0} \notin span(\Gamma_{\zeta_1},...,\Gamma_{\zeta_m}) \tag{2.10}$$

If $\mathrm{wt}(\eta\oplus\zeta_0) \ge 4$ and $j^* \in [d_0]$, then (2.10) holds even if only the columns indexed by $S' := S\backslash\{\zeta_0 \oplus e_{p^0_{j^*}}\}$ are linearly independent.

Proof. We first prove the claim without the assumption that $\mathrm{wt}(\eta\oplus\zeta_0) \ge 4$. Suppose to the contrary that $\Gamma_{\zeta_0} = \sum_{i=1}^m a_i\Gamma_{\zeta_i}$ so that $\Gamma_{\zeta_0}\wedge\Gamma_\eta - \sum_{i=1}^m a_i\Gamma_{\zeta_i}\wedge\Gamma_\eta = 0$. By Lemma 2.3.5, this linear relation implies the following linear relation among wedges of columns of the other parity:

$$\epsilon_{\zeta_0,\eta}\left(\sum_{j=1}^{d_0}(-1)^{j+1}\Gamma_{\zeta_0\oplus e_{p^0_j}}\wedge\Gamma_{\eta\oplus e_{p^0_j}}\right) - \sum_{i=1}^m \epsilon_{\zeta_i,\eta}a_i\left(\sum_{j=1}^{d_i}(-1)^{j+1}\Gamma_{\zeta_i\oplus e_{p^i_j}}\wedge\Gamma_{\eta\oplus e_{p^i_j}}\right) = 0 \tag{2.11}$$

We claim this is a nontrivial linear relation contradicting the linear independence of the columns indexed by S. For each of the $m+1$ sums indexed by $1 \le j \le d_i$ appearing in (2.11), if $d_i = 2$, rewrite $\sum_{j=1}^{d_i}(-1)^{j+1}\Gamma_{\zeta_i\oplus e_{p^i_j}}\wedge\Gamma_{\eta\oplus e_{p^i_j}}$ as $2\Gamma_{\zeta_i\oplus e_{p^i_1}}\wedge\Gamma_{\eta\oplus e_{p^i_1}}$.

After this consolidation, note that the wedge products in (2.11) are now all distinct. Certainly for any $j,j' \in [d_i]$ where $d_i \ge 4$, $\Gamma_{\zeta_i\oplus e_{p^i_j}}\wedge\Gamma_{\eta\oplus e_{p^i_j}}$ and $\Gamma_{\zeta_i\oplus e_{p^i_{j'}}}\wedge\Gamma_{\eta\oplus e_{p^i_{j'}}}$ are linearly independent. For i,i' such that $d_i = 2$ and $d_{i'} > 2$, $2\Gamma_{\zeta_i\oplus e_{p^i_1}}\wedge\Gamma_{\eta\oplus e_{p^i_1}}$ and any $\Gamma_{\zeta_{i'}\oplus e_{p^{i'}_j}}\wedge\Gamma_{\eta\oplus e_{p^{i'}_j}}$ are linearly independent as $\zeta_i\oplus\eta \ne \zeta_{i'}\oplus\eta$, contradicting the assumption that $\{\zeta_1,...,\zeta_m\}$ are distinct. Similarly, for i,i' such that $d_i > 2$ and $d_{i'} > 2$, any $\Gamma_{\zeta_i\oplus e_{p^i_j}}\wedge\Gamma_{\eta\oplus e_{p^i_j}}$ and any $\Gamma_{\zeta_{i'}\oplus e_{p^{i'}_{j'}}}\wedge\Gamma_{\eta\oplus e_{p^{i'}_{j'}}}$ are linearly independent as $\zeta_i\oplus\eta \ne \zeta_{i'}\oplus\eta$.

We conclude that (2.11), after consolidating sums for which $d_i = 2$, consists of a nonzero number of linearly independent wedge products of columns indexed by S, so (2.11) is indeed a nontrivial linear relation among the wedge products $\Gamma_s \wedge \Gamma_{s'}$ for $s, s' \in S$. But all columns indexed by S are linearly independent by assumption, so this linear relation contradicts Lemma 2.3.8 and the linear independence of columns indexed by S.

For the second part of Lemma 2.3.9, we claim that (2.11) is still a nontrivial relation. Pick any $k \neq j^*$ inside $[d_0]$. Because $d_0 \geq 4$,

$$\zeta_0 \oplus e_{p_k^0}, \eta \oplus e_{p_k^0} \neq \eta \oplus e_{p_{j^*}^0}. \tag{2.12}$$

We have already taken care of the case where $\Gamma_{\zeta_0 + e_{p_{j^*}^0}} \notin \operatorname{span}(S')$ above, so suppose instead that

$$\Gamma_{\zeta_0 + e_{p_{j^*}^0}} = \sum_{s \in S'} b_s \Gamma_s. \tag{2.13}$$

If we consolidate sums for which $d_i = 2$ in (2.11) as above and substitute (2.13) into the resulting equation, the wedge products that (2.11) now contains also include ones of the form $\Gamma_s \wedge \Gamma_{\eta + e_{p_{j^*}^0}}$, which cannot be linearly dependent with $\Gamma_{\zeta_0 \oplus e_{p_k^0}} \wedge \Gamma_{\eta \oplus e_{p_k^0}}$ by (2.12).

Every other wedge product in (2.11) is of the form $\Gamma_{\zeta_i \oplus e_{p_j^i}} \wedge \Gamma_{\eta \oplus e_{p_j^i}}$ and also cannot be linearly dependent with $\pm \Gamma_{\zeta_0 \oplus e_{p_k^0}} \wedge \Gamma_{\eta \oplus e_{p_k^0}}$ or else, as before, we'd find that $\zeta_i \oplus \eta = \zeta_k \oplus \eta$, contradicting the assumption that $\{\zeta_1, ..., \zeta_m\}$ are distinct.

It follows that if we rewrite the left-hand side of (2.11) in the form $\sum_{s,s' \in S'} b_{s,s'} \Gamma_s \wedge \Gamma_{s'}$ (uniquely because by Lemma 2.3.8 the $\Gamma_s \wedge \Gamma_{s'}$ are linearly independent), $b_{\zeta_0 \oplus e_{p_k^0}, \eta \oplus e_{p_k^0}} \neq 0$. So (2.11) is still a nontrivial linear relation, contradicting Lemma 2.3.8 and the linear independence of columns indexed by S. □

2.4 Relation to spinor varieties

In [34], Landsberg et al. observed that the Matchgate Identities are the same quadratic relations that cut out the variety of pure spinors. In Appendix A.7, we very briefly develop the basic algebraic theory of spinors, mostly following [16, 64], and we use it here first to get from the definition of pure spinors as the spinors representative of maximal totally isotropic subspaces to a characterization of them as vectors of sub-Pfaffians.

In addition, [34] reformulated the theory of holographic algorithms over the Boolean domain as an SL(2)-orbit containment problem, and we describe some parts of their formulation here, with a view towards extending it to arbitrary domain sizes.

2.4.1 Characterization as sub-Pfaffians

Let V be an n-dimensional vector space equipped with a quadratic form Q of maximal isotropic index and for simplicity assume $n = 2m$ is even. Pick a splitting of V into maximal totally isotropic subspaces $N \oplus P$.

Define the *spinor variety* $\mathbb{S}_N = \mathbb{S}_m$ as the variety of all maximal totally isotropic subspaces of V. $\mathbb{S}_N$ has two isomorphic components $\mathbb{S}_N^+ \subset S_+ := \bigwedge^{even} V$ and $\mathbb{S}_N^- \subset S_- := \bigwedge^{odd} V$ consisting of even and odd pure spinors, respectively. We will denote the corresponding spinor variety for V^* by $\mathbb{S}_{N^*} = \mathbb{S}_{m^*}$.

Lemma 2.4.1. *A generic point on $\mathbb{S}_m^+$ is parametrized by sub-Pfaffians of a generic $m \times m$ skew-symmetric matrix.*

Proof. We first show that in an open neighborhood of N, maximal totally isotropic subspaces Z are parametrized by $m \times m$ skew-symmetric matrices. To such a Z we may associate an $m \times 2m$ matrix M_Z which, after a change of basis, we may assume to be of the form $\begin{pmatrix} I_m & M'_Z \end{pmatrix}$ for some $M'_Z \in \mathcal{M}_{m \times m}(\mathbb{C})$. Because Z is totally isotropic, if $\mathcal{B}$ is the $n \times n$ matrix associated to the polarization of Q, then $M_Z \cdot \mathcal{B} \cdot M_Z^T = 0$, from which we conclude that M'_Z must be skew-symmetric.

Let N and P have bases $\{e_1, ..., e_m\}$, $\{f_1, ..., f_m\}$ for which $B(e_i, e_j) = \delta_{ij}$, and suppose $N \in \mathbb{S}_m^+$. As before, denote by e and f the products $e = \prod e_i$ and $f = \prod f_i$. Then by the above, for any totally isotropic Z of dimension m near N, there exists a skew-symmetric $M'_Z := (a_{ij})$ for which Z has a basis $(e_i + \sum_{j>i} a_{ij} f_j)_{1\le i\le m}$. The pure spinor corresponding to Z is

$$\prod_{i=1}^{m}\left(e_i + \sum_{j>i} a_{ij} f_j\right) = \left(\sum_{S\subset[m]} \mathrm{Pf}_{[m]\setminus S}(M) \prod_{i\notin S} e_i\right)\cdot f,$$

where $\mathrm{Pf}_{[m]\setminus S}(M)$ denotes the Pfaffian of the submatrix of M consisting of rows and columns indexed by $[m]\setminus S$. So we conclude that vectors of sub-Pfaffians parametrize points in $\mathbb{S}^+_m$. □

In [9], the relations among the sub-Pfaffians of a generic $m \times m$ skew-symmetric matrix were shown to be exactly the same as those among the entries of a standard signature.

Henceforth, denote the set of sub-Pfaffians of a matrix $x \in \Lambda^2 N$ by $\mathrm{sPfaff}(x)$, where $(\mathrm{sPfaff}(x))_S = \mathrm{Pf}_S(x)$. Dually, denote the set of sub-Pfaffians of a matrix $x \in \Lambda^2 N^*$ by $\mathrm{sPfaff}^\wedge(x)$, where $(\mathrm{sPfaff}^\wedge(x))_S = \mathrm{Pf}_{[m]\setminus S}(x)$.

2.4.2 Holographic algorithms and orbit containment

Recall that if there are n wires in matchgrid Ω, $\mathrm{Holant}(\Omega)$ is a pairing $\langle u, v\rangle$ for $u \in \mathbb{C}^{2^n}$ the tensor product of all generator signatures and $v \in \mathbb{C}^{*2^n}$ the tensor product of all recognizer signatures. The point of holographic algorithms is that u and v are in "special position" so that that the pairing takes only $\mathsf{poly}(n)$ steps to compute rather than $O(2^n)$ steps if they were general.

The effectiveness of holographic algorithms rests in part in the ability to solve "global" problems by encoding local constraints with short matchgate signatures and tensoring these together to produce u and v.

Lemma 2.4.2. *If maximal isotropic subspace $N \subset V$ decomposes as $N_1 \oplus \cdots \oplus N_g$, then there is a natural inclusion $\iota : \hat{\mathbb{S}}_{N_1} \times \cdots \times \hat{\mathbb{S}}_{N_g} \hookrightarrow \hat{\mathbb{S}}_N$. Specifically, if for each $1 \le j \le g$ we have that $v_j = \mathrm{sPfaff}(x_j)$ for some skew-symmetric matrix x_j, then ι sends $(v_1, \ldots, v_g)$ to $\mathrm{sPfaff}(x)$, where x is a block matrix with blocks $x_1, \ldots, x_g$.*

Remark 2.4.3. Pure spinors $\underline{G_i}$ can also be interpreted as vectors lying in the $\mathrm{Spin}(V)$-orbit of the highest weight vector of the/one of the spin representation(s) of $\mathfrak{so}_{n_i}$. In this case, Lemma 2.4.2 is simply a consequence of the fact that the tensor product of highest weight vectors for different subgroups $G_i \subset G$ with compatible Weyl chambers is itself a highest weight vector for G.

Lemma 2.4.2 implies that as long as the edges incident to generators in Ω are labeled "contiguously" in the following sense, $\bigotimes_i G_i$ is also a pure spinor.

Definition 2.4.4. An *edge order* on matchgrid Ω is a bijective labeling of the w wires with indices from $[w]$. A *generator order* is an edge order such that for each generator G_i in Ω, the indices of the wires incident to external nodes of G_i are adjacent. *Recognizer orders* can be defined analogously.

Holographic algorithms also rely on the fact that if $u \in \mathbb{S}_n$ and $v \in \mathbb{S}_{n*}$, then $\langle u, v\rangle$ can be computed efficiently.

Lemma 2.4.5. *For $x \in \Lambda^2 N$ and $y \in \Lambda^2 N^*$,*

$$\langle \mathrm{sPfaff}(x), \mathrm{sPfaff}^\vee(y)\rangle = \mathrm{Pf}(\tilde{x} + y),$$

where $\tilde{x}$ is the matrix given by $\tilde{x}^i_j = (-1)^{i+j+1} x^i_j$.

Proof. The following identity was observed in [49].

Claim 2.4.6.

$$\mathrm{Pf}(x+y) = \sum_{|S|\ \ even} sgn(S)\,\mathrm{Pf}_S(x)\,\mathrm{Pf}_{[m]\setminus S}(y),$$

where $sgn(S) = (-1)^{\sigma(S)+|S|/2}$ for $\sigma(S) = \sum_{i\in S} i$.

This identity implies the lemma because $\mathrm{Pf}_S(x) = \mathrm{sgn}(S)\,\mathrm{Pf}_S(\tilde{x})$. □

The role of the basis matrix is thus to move signatures $G_1, ..., G_g$ and $R_1, ..., R_r$ simultaneously into spinor varieties. First consider basis size $\ell = 1$ and domain size $k = 2$, in which case we can take $M \in \mathrm{SL}_2(\mathbb{C})$. $\mathrm{SL}_2(\mathbb{C})$ has a natural action on $\mathbb{C}^{\otimes 2^n} = (\mathbb{C}^2)^{\otimes n}$ and $(\mathbb{C}^*)^{\otimes 2^n} = (\mathbb{C}^{2*})^{\otimes n}$, from which we obtain Definition 2.2.5. Specifically, the action of M is given by

$$G_i \mapsto M^{\otimes n_i} G_i \quad R_j \mapsto R_j (M^{-1})^{\otimes n_j}. \tag{2.14}$$

By the above, we can conclude the following:

Proposition 2.4.7. *If there is an edge order on matchgrid Ω which is simultaneously a generator and recognizer order, and if under appropriate base change M we have that each matchgate signature is contained in a spinor variety, i.e. of the form $\mathrm{sPfaff}(u_i)$ or $\mathrm{sPfaff}^\vee(v_j)$ for some skew-symmetric matrix $u \in \Lambda^2 N$ or $v \in \Lambda^2 N^*$, then*

$$\mathrm{Holant}(\Omega) = \mathrm{sPfaff}(\tilde{G} + R),$$

which can be computed in $\mathsf{poly}(n)$ *time.*

Remark 2.4.8. We will not deal with the issue of the existence of such an edge order for a given Ω except to say that it exists, for instance, when the underlying graph of Ω is planar (this is the premise of the FKT algorithm and was shown in [38]).

Now consider the following problem fundamental to understanding the limitations of holographic algorithms and studied extensively by Cai et al. in [4, 5, 6, 7, 8, 9, 10, 11, 12, 13].

Question 2.4.9 (Simultaneous Realizability Problem)**.** Given a generator signature $\mathbf{G}$ and a recognizer signature $\mathbf{R}$, does there exist a basis M over which they are simultaneously realizable?

Here, $\mathbf{G}$ and $\mathbf{R}$ should be thought of as the same as those defined in Observation 1, i.e. tensor products of smaller signatures. Over the Boolean domain, we can now easily reinterpret this geometrically as a question about $\mathrm{SL}_2(\mathbb{C})$-orbits. Denoting the arities of $\mathbf{G}$ and $\mathbf{R}$ by m_1 and m_2 respectively, we know that $\mathrm{SL}_2(\mathbb{C})$ acts naturally on $\mathbb{S}_{m_1} \times \mathbb{S}_{m_2*}$.

Question 2.4.10 (Orbit Containment Problem- Boolean)**.** Does the orbit $\mathrm{SL}_2(\mathbb{C}) \cdot (\mathbb{S}_{m_1} \times \mathbb{S}_{m_2*})$ contain the point $(\mathbf{G}, \mathbf{R})$?

In [10], it is shown that there is a deterministic polynomial time algorithm for deciding simultaneous realizability provided that $\mathbf{G}$ and $\mathbf{R}$ are *symmetric*, that is, each of their entries only depends on the Hamming weight of its index. This implies a deterministic polynomial time algorithm for deciding orbit containment for such points $(\mathbf{G}, \mathbf{R})$.

Over domain size k and basis size ℓ, the geometric interpretation is slightly more subtle. For one, signatures and standard signatures no longer live in the same space. In addition, there is not necessarily a natural action of $\mathcal{M}_{2^\ell \times k}$ on $(\mathbb{C}^{k*})^{\otimes n}$ because basis M might not have a right inverse, let alone a unique one. We will return to this question in the next chapter and conclude that the simultaneous realizability problem over domain k can still be formulated as an orbit containment problem.

Chapter 3

Basis Collapse

Fundamental to understanding the ultimate limitations of Valiant's holographic framework is understanding the power afforded by the change of basis. Indeed, one major direction of study in this field concerns better understanding the relationship between the two key parameters of the basis change: domain and basis size. It is not unreasonable to expect that for a given domain size k, increasing the basis size ℓ extends the range of counting problems that can be solved. But if this were always the case, understanding the holographic framework's limitations would be rather hopeless: there's a potentially unbounded range of dimensions of basis matrix that we need to consider, each one more versatile than the last.

We may ask instead: over a given domain size k, is there a smallest basis size ℓ for which any holographic algorithm over domain size k can be simulated by one with a basis size of ℓ? To prove such a general *basis collapse theorem* has been a long-standing open problem in this area, and in this chapter we resolve this problem.

In Section 3.1, we mention some previous work in this area, further motivate the need for such a collapse theorem, and formally state the conjecture that we will prove. We outline our general approach in Section 3.2 and implement it in the remaining sections. At the very end, we make a few remarks about how this result is relevant to the interpretation of holographic algorithms in terms of spinor varieties.

3.1 History and Other Motivations

The first holographic algorithms studied were on domain size 2, dealing with counting problems involving matchings, 2-colorings, graph bipartitions, and Boolean satisfying assignments, and almost exclusively used bases of size 1. The first holographic algorithm over the Boolean domain to use a basis of size 2 was the one given in [59] for $\#_7$Pl-Rtw-Mon-3CNF. It was suspected at the time that this might be an example where increasing basis size adds power to the holographic approach, but as we saw in Section 2.1.2, [12] subsequently demonstrated that $\#_7$Pl-Rtw-Mon-3CNF can be solved with a basis of size 1.

In [11], Cai and Lu then showed to the contrary that on domain size 2, any nontrivial holographic algorithm with basis of size $\ell \geq 2$ can be simulated by one with basis of size 1. In other words, for all ℓ, the class of domain size 2 problems solvable with basis size ℓ *collapses* to the class solvable with basis size 1.

That this collapse theorem restricted the focus to basis matrices in $\mathrm{GL}_2(\mathbb{C})$ allowed Cai and his collaborators to initiate a systematic study of the structural theory of holographic algorithms over the Boolean domain [4, 5, 6, 7, 8, 9, 10, 11, 12, 13], compiling what amounts to a catalogue of such constructions that turns the process of finding basic holographic reductions into something essentially algorithmic. For example, as a corollary to their results in [10], they noted that the modulus 7 appears in $\#_7$Pl-Rtw-Mon-3CNF because it is a Mersenne prime and that more generally, there is a polynomial-time holographic algorithm for $\#_{2^k-1}$Pl-Rtw-Mon-kCNF.

On the other hand, the theory for holographic algorithms over higher domains is almost wholly underdeveloped because of the absence of a general basis collapse theorem.

In [8], Cai and Fu showed that holographic algorithms on domain sizes 3 and 4 using at least one full-rank signature collapse to basis sizes 1 and 2 respectively. They conjectured that for domain size k, we get a collapse to basis size $\lfloor \log_2 k \rfloor$ and suggested a heuristic explanation that for domain size $k = 2^K$, we only need $\log_2 k$ bits to encode each of the k colors.

3.2 Our Techniques

3.2.1 Our results and techniques

We prove Cai and Fu's conjecture in the affirmative for all domain sizes.

Theorem 3.2.1. *Any holographic algorithm on domain size k using at least one matchgate with signature of rank k can be simulated by a holographic algorithm with basis size $\lfloor \log_2 k \rfloor$.*

As Cai and Fu noted in [8], their "information-theoretic" explanation for the collapse theorems on domain sizes up to 4 is insufficient to explain why holographic algorithms on domain size 3 collapse to basis size 1 and not just 2.

To prove their collapse theorem on domain size 3, Cai and Fu actually showed that the bases of holographic algorithms on domain size 3 which use at least one full-rank signature must be of rank 2 rather than 3. They then invoked the main result of Fu and Yang [21] which reduces holographic algorithms on domain size k with bases of rank 2 to ones on domain size 2.

Our key observation is that this phenomenon occurs at a much larger scale. As a bit of informal background, the *standard signature* of a matchgate G is a vector encoding the perfect matching properties of G. By indexing appropriately, we can regard the standard signature as a matrix Γ. The entries of this matrix are known [9] to satisfy a collection of quadratic polynomial identities called the Matchgate Identities (MGIs), and by using these identities together with some multilinear algebra, we prove the following:

Theorem 3.2.2 (Rank Rigidity Theorem). *The rank of the standard signature matrix Γ is always a power of two.*

We can then conclude that the basis of a nontrivial holographic algorithm on domain size k must be of rank 2^ℓ, where ℓ is the largest integer for which $2^\ell \leq k$. With this step, together with a generalization of Fu and Yang's result to bases of rank k, we show it is enough to prove a collapse theorem for holographic algorithms on domain sizes that are powers of two. Cai and Fu [8] achieved such a collapse theorem for domain size 4 by proving that 1) any standard signature of rank 4 contains a full-rank 4×4 submatrix whose entries have indices are "close" in Hamming distance, 2) full-rank $4 \times 2^{2n-2}$ standard signatures have right inverses that are also standard signatures.

For 1), the proof in [8] used algebraic techniques involving the matchgate identities, but these methods seem to work only for domain size 4. We instead show that the required generalization of 1) to arbitrary domain sizes almost trivially follows from the rank rigidity theorem and the MGIs. Roughly, we prove the following:

Theorem 3.2.3 (Cluster Existence - informal). *If Γ is a $2^\ell \times 2^{(n-1)\ell}$ matrix of rank at least k realizable as the standard signature of some matchgate, then there exists a $2^{\lceil \log_2 k \rceil} \times 2^{\lceil \log_2 k \rceil}$ submatrix of full rank in Γ whose column (resp. row) indices differ in at most $\lceil \log_2 k \rceil$ bits.*

For 2), the proof in [8] nonconstructively verifies that the set of all invertible 4×4 matrices satisfying the matchgate identities up to sign forms a group under multiplication. 4×4 matrices are easy to handle because there is only one nontrivial MGI in this case. Rather than generalizing this approach, we note that Li and Xia [39] proved a very similar but more general result under a different framework of matchgate computation known as character theory, showing that the set of all invertible $2^K \times 2^K$ matrices realizable as matchgate characters forms a group under multiplication for all K. It turns out their technique carries over with minor modifications into the framework of signature theory that we consider, and we use it to show the following:

Theorem 3.2.4 (Group Property - informal). *If G is a generator matchgate with $2^K \times 2^{(n-1)K}$ standard signature $\underline{G}$ of rank 2^K, then there exists a recognizer matchgate with $2^{K(n-1)} \times 2^K$ standard signature $\underline{R}$ such that $\underline{G} \cdot \underline{R} = I_{2^K}$, where I_{2^K} denotes the $2^K \times 2^K$ identity matrix.*

Our general collapse theorem then follows from our generalizations of 1) and 2). This result gives a way forward for the development of the structural theory of holographic algorithms on higher domain sizes in the same vein as Cai et al.'s work on domain size 2. In [62], Valiant gave examples of holographic algorithms on domain size 3, but holographic algorithms on higher domain sizes have yet to be explored. Our result shows that for domain size k, we can focus on understanding changes of basis in $\mathcal{M}_{2^{\lfloor \log_2 k \rfloor} \times k}(\mathbb{C})$ rather than over an infinite set of dimensions, just as the collapse theorem of Cai and Lu [11] showed that on the Boolean domain, they could focus on understanding changes of basis in $\mathrm{GL}_2(\mathbb{C})$.

3.2.2 Organization

In Section 3.3, we prove Theorems 3.2.2 and 3.2.3. In Section 3.4, we then prove Theorem 3.2.4. In Section 3.5, we generalize the main result of Fu and Yang [21] to bases of rank k and reduce proving a collapse theorem on all domain sizes to proving one on domain sizes equal to powers of two. Finally, in Section 3.6, we prove the desired collapse theorem, Theorem 3.2.1, on domain sizes equal to powers of two by invoking the results from Section 3.3.

3.3 Rigidity and Cluster Existence

In this section, we will prove the rigidity theorem and the cluster existence theorem, informally stated as Theorems 3.2.2 and 3.2.3. Now that we have introduced the appropriate terminology, we first state both precisely.

Theorem 3.3.1 (Rank Rigidity - formal)**.** *If* Γ *is a* $2^\ell \times 2^m$ *pseudo-signature, then* $\mathrm{rank}(\Gamma)$ *is a power of 2. Equivalently, for all* $\kappa \geq 1$,

$$\mathrm{rank}(\Gamma) \geq 2^\kappa + 1 \Rightarrow \mathrm{rank}(\Gamma) \geq 2^{\kappa+1} \tag{3.1}$$

Theorem 3.3.2 (Cluster Existence - formal)**.** *If* $\Gamma = \underline{G}(t)$ *for some matchgate* G *of arity* $n\ell$*, and* $\mathrm{rank}(\Gamma) \geq k$*, then there is a* $\lceil \log_2 k \rceil$*-cluster submatrix of full rank.*

To prove Theorem 3.3.2, we claim it is enough to show the following:

Theorem 3.3.3. *If* Γ *is a* $2^\ell \times 2^m$ *pseudo-signature such that* $\mathrm{rank}(\Gamma) \geq k$*, then there exists a* $(\lceil \log_2 k \rceil, m)$*-cluster* Z *for which* Γ_Z *is of full rank.*

Equivalently, for all $\kappa \geq 1$,

$$\mathrm{rank}(\Gamma) \geq 2^\kappa + 1 \Rightarrow \exists\ (\kappa+1, m)\text{-cluster } Z \text{ such that } \mathrm{rank}(\Gamma_Z) = 2^{\kappa+1} \tag{3.2}$$

Proof of Theorem 3.3.2. Apply Theorem 3.3.3 to Γ to obtain Γ_Z. By Observation 2, Γ_Z^T is also a pseudo-signature. Apply Theorem 3.3.3 to Γ^T to get the desired cluster submatrix. ☐

Note that for fixed κ, (3.2) implies (3.1). We will jointly prove Theorem 3.3.1 and Theorem 3.3.3 by induction on k. Cai and Fu have already shown both for $k \leq 4$; we take these results as our base case. We complete the following two inductive steps.

Inductive Step 1. If implication (3.2) holds for $1 \leq \kappa \leq K-1$, then implication (3.1) holds for $1 \leq \kappa \leq K$.

Inductive Step 2. If implication (3.1) holds for $1 \leq \kappa \leq K$ and implication (3.2) holds for $\kappa \leq K-1$, then implication (3.2) also holds for $\kappa = K$.

Note that once we have proven the rigidity and cluster existence theorems, we additionally obtain the following.

Corollary 3.3.4. *If* G *is a full-rank matchgate signature on domain size* k*, it is only realizable over bases* M *of rank at most* $2^{\lfloor \log_2 k \rfloor}$.

Proof. If k is a power of two, the claim is trivial. If not, assume G is a generator (standard signatures of recognizer are also pseudo-signatures, so the argument in that case is analogous). If $\mathrm{rank}(G(t)) = k$ and $\mathrm{rank}(M) \geq 2^{\lfloor \log_2 k \rfloor} + 1$, then by Lemma 2.2.14, $2^{\lfloor \log_2 k \rfloor} + 1 \leq \mathrm{rank}(\underline{G}(t)) \leq k$. But Theorem 3.3.1 would then imply $\mathrm{rank}(\underline{G}(t)) \geq 2^{\lfloor \log_2 k \rfloor + 1} > k$, a contradiction. ☐

3.3.1 Rank rigidity theorem

In this subsection, we complete the former inductive step, and in the next, we complete the latter.

Before we prove Inductive Step 1 in its entirety, we take care of the case where $m = K+1$. While this might appear to be the simplest case because m is minimal, it will turn out that cases where m is greater will reduce to this case. For this reason, the wedge product machinery introduced in Section 2.3 is used exclusively in the proof of this case.

Once we take care of this case, we will essentially show that if standard signature Γ is any wider, i.e. if $m > K+1$, then if Z is a cluster of size greater than 2^K indexing columns of rank at least 2^K+1, we can always find a proper subcluster also indexing columns of rank at least 2^K+1, or else the matchgate identities would erroneously imply that certain columns which are known to be linearly independent are linearly dependent.

We begin with the case of $m = K+1$.

Theorem 3.3.5. *If* Γ *is a* $2^\ell \times 2^{K+1}$ *pseudo-signature such that* $\operatorname{rank}(\Gamma) \geq 2^K+1$*, then* $\operatorname{rank}(\Gamma) = 2^{K+1}$.

Proof. Because we assume implication (3.2) holds for $\kappa = K-1$, Γ contains a (K,m)-cluster $Z = s \oplus \{e_1, ..., \hat{e_j}, ..., e_{K+1}\}$ of linearly independent columns; denote the even indices of Z by Z_0 and the odd ones by Z_1. Because $\operatorname{rank}(\Gamma) > 2^K$, there exists $t \notin Z$ for which $\Gamma_t \notin \operatorname{span}(Z)$. Denote the parity of t by $b \in \{0,1\}$, and denote by $\bar{b}$ the opposite parity.

Select any $t' = t \oplus e_{i^*}$ for $i^* \neq j$ and apply Lemma 2.3.9 to $\zeta_0 = t'$, $\eta = t \oplus e_j$, $T = Z_{\bar{b}}$ to conclude that $\Gamma_{t'} \notin \operatorname{span}(Z_{\bar{b}})$.

Let S_{d_1,d_2} denote the set of column indices $\zeta \notin Z$ for which $\operatorname{wt}(\zeta \oplus t) = d_1$ and $\operatorname{wt}(\zeta \oplus t') = d_2$. Note that because $\operatorname{wt}(t \oplus t') = 1$, S_{d_1,d_2} is empty if $|d_1 - d_2| \neq 1$.

We will show by induction on D that the columns indexed by $\left(\bigcup_{d=0}^{D} S_{d,d+1} \cup S_{d+1,d}\right) \cup Z$ are linearly independent for all $0 \leq D \leq K$. The definition of t and the argument above for t' give the base case of $D = 0$.

For the inductive step, for each d let d_0 and d_1, respectively denote the even and odd value in $\{d, d+1\}$. As columns indexed by S_{d_0,d_1} and columns indexed by S_{d_1,d_0} have opposite parity, it suffices to show that the columns indexed by $T_D^0 := \left(\bigcup_{d=0}^{D} S_{d_0,d_1}\right) \cup Z_b$ are linearly independent, and that the columns indexed by $T_D^1 := \left(\bigcup_{d=0}^{D} S_{d_1,d_0}\right) \cup Z_{\bar{b}}$ are linearly independent.

Within this inductive step, we will further induct on the elements within S_{D_0,D_1} and S_{D_1,D_0}. Specifically, suppose we have already proven that for some subset $S'_{D_0,D_1} \subset S_{D_0,D_1}$, all columns indexed by $T_{D-1}^0 \cup S'_{D_0,D_1}$ are linearly independent, and that for $S'_{D_1,D_0} := \{u \oplus e_{i^*} : u \in S'_{D_0,D_1}\} \subset S_{D_1,D_0}$, all columns indexed by $T_{D-1}^1 \cup S'_{D_1,D_0}$ are linearly independent. Select any $u \notin S'_{D_0,D_1}$ and apply Lemma 2.3.9 to $\zeta_0 = u$, $\eta = t' \oplus e_j$, $T = T_{D-1}^0 \cup S'_{D_0,D_1}$ to see that $\Gamma_u \notin \operatorname{span}(T_{D-1}^0 \cup S'_{D_0,D_1})$. Note that when $\operatorname{wt}(\zeta_0 \oplus \eta) \geq 4$, we do not yet know that $\Gamma_{\zeta_0 \oplus e_{i^*}} = \Gamma_{u \oplus e_{i^*}}$ lies outside $\operatorname{span}(T_{D-1}^1 \cup S'_{D_1,D_0})$, that is, we do not know whether all the columns indexed by the set S defined in Lemma 2.3.9 are linearly independent, but the second part of Lemma 2.3.9 says that we may still conclude that $\Gamma_u \notin \operatorname{span}(T_{D-1}^0 \cup S'_{D_0,D_1})$ because the columns indexed by $S \backslash \{u \oplus e_{i^*}\}$ are linearly independent.

Lastly, apply Lemma 2.3.9 to $\zeta_0 = u \oplus e_{i^*}$, $\eta = t \oplus e_j$, $T = T_{D-1}^1 \cup S'_{D_1,D_0}$ to see that $\Gamma_{u \oplus e_{i^*}} \notin \operatorname{span}(T_{D-1}^1 \cup S'_{D_1,D_0})$. Note that here we only need to invoke the first part of Lemma 2.3.9 we already know that $\Gamma_{\zeta_0 \oplus e^*} = \Gamma_u$ lies outside $\operatorname{span}(T_{D-1}^0 \cup S'_{D_0,D_1})$. ☐

We are now ready to complete Inductive Step 1.

Completion of Inductive Step 1. As we remarked earlier, implication (3.2) for a fixed value of κ implies implication (3.1) for that value of κ, so we just need to show that implication (3.1) also holds for $\kappa = K$.

Suppose to the contrary that there exists pseudo-signature Γ of rank k such that $2^K + 1 \leq k < 2^{K+1}$. Without loss of generality, we may assume that for all clusters $Z \subsetneq \{0,1\}^m$, $\operatorname{rank}(\Gamma_Z) \leq 2^K$; otherwise, replace Γ by Γ_Z for some small enough cluster Z such that $\operatorname{rank}(\Gamma_Z) \geq 2^K + 1$ and $\operatorname{rank}(\Gamma_{Z'}) \leq 2^K$ for all subclusters $Z' \subsetneq Z$. Furthermore, by Theorem 3.3.5, we may assume $m > K+1$.

Lemma 3.3.6. *If* $Z = s + \{e_{p_1}, ..., e_{p_K}\}$ *is a* (K,m)*-cluster of linearly independent columns in* Γ*, then any column* Γ_t *for which* $t_i = s_i$ *for some* $i \neq p_1, ..., p_K$ *lies in the span of the columns indexed by* Z.

Proof. If to the contrary there existed such a Γ_t not lying in the span of Z so that $t_i = s_i$ for some $i \neq p_1, ..., p_K$, then if Z' is the $(m-1,m)$-cluster of column indices ζ for which $\zeta_i = s_i$, $\Gamma_{Z'}$ contains t and all of Z and thus has rank at least $2^K + 1$, contradicting our assumption on the ranks of the proper subclusters of Γ. □

By the inductive hypothesis, Γ contains a (K,m)-cluster Z of linearly independent columns $s+\{e_{p_1}, ..., e_{p_K}\}$. As s is only uniquely defined modulo the bits in positions $p_1, ..., p_K$, we will leave those bits of s unspecified for now.

Because $k > 2^K$, there exists $t \notin Z$ for which all columns indexed by $Z \cup \{t\}$ are linearly independent. Moreover, by Lemma 3.3.6, $t_j = \overline{s_j}$ for all $j \notin \{p_1, ..., p_K\}$. Let Z' denote the cluster $t + \{e_{p_1}, ..., e_{p_K}\}$. Set $s_i = t_i$ for $i \in \{p_2, ..., p_K\}$; we will set s_{p_1} to be 0 or 1 depending on the parity of the number $m - K$ of bits outside of positions $p_1, ..., p_K$.

Case 1. $m - K$ is even.

Set $s_{p_1} = t_{p_1}$ so that s and t have the same parity.

Claim 3.3.7. *If $u \notin Z'$ and $u_i = s_i$ for each $i \in \{p_1, ..., p_K\}$, then Γ_u and Γ_s are linearly dependent.*

Proof. For each $i \in \{p_1, ..., p_K\}$ and $j \notin \{p_1, ..., p_K\}$, let T_i denote the cluster of all column indices u for which $u_i = s_i$, and let T_i^j denote the cluster of all column indices u for which $u_i = s_i$ and $u_j = s_j$. Let $Z_i = Z \cap T_i$; obviously $Z_i \subset T_i^j \subset T_i$.

Because T_i^j is a cluster properly contained in $\{0,1\}^m$, we inductively know that $\text{rank}(\Gamma_{T_i^j}) \leq 2^K$. And because $Z_i \subset T_i^j$, $\text{rank}(\Gamma_{T_i^j}) \geq 2^{K-1}$. But if $\text{rank}(\Gamma_{T_i^j}) \geq 2^{K-1} + 1$, then by inductive hypothesis (3.2) applied to $\Gamma_{T_i^j}$ for $\kappa = K - 1$, $\text{rank}(\Gamma_{T_i^j}) \geq 2^K$. In other words, $\text{rank}(\Gamma_{T_i^j})$ is either 2^{K-1} or 2^K.

We will show that the latter is impossible. Suppose to the contrary that $\text{rank}(\Gamma_{T_i^j}) = 2^K$.

Then because $T_i^j \subset T_i$ and $\text{rank}(\Gamma_{T_i}) = 2^K = \text{rank}(\Gamma_{T_i^j})$, it follows that $\text{span}(T_i) = \text{span}(T_i^j)$. For any $u \in T_i^j$, $\Gamma_u \in \text{span}(Z)$ by Lemma 3.3.6, so

$$\text{span}(Z) \supset \text{span}(T_i^j) = \text{span}(T_i).$$

But T_i contains t, and by definition $\Gamma_t \notin \text{span}(Z)$, a contradiction.

We conclude that $\text{rank}(\Gamma_{T_i^j}) = 2^{K-1}$. Then because $Z_i \subset T_i^j$ and $\text{rank}(\Gamma_{Z_i}) = 2^{K-1} = \text{rank}(\Gamma_{T_i^j})$, it follows that $\text{span}(Z_i) = \text{span}(T_i^j)$.

In particular, all columns indexed by $\bigcap_{k=1}^K T_{p_k}^j$ lie in $\bigcap_{k=1}^K \text{span}(Z_{p_k}) = \text{span}(\{s\})$. Our choice of j was arbitrary, so we get the desired claim. □

From the above claim and the fact that we're assuming $m > K + 1$, we conclude that $\Gamma_{s \oplus e_j}$ for any $j \notin \{p_1, ..., p_K\}$ lies in the span of Γ_s. But s and $s \oplus e_j$ are of opposite parity, so by the parity condition, $\Gamma_{s \oplus e_j} = 0$ for all such j. Applying Corollary 2.3.6 to s and t, it follows that Γ_s and Γ_t are linearly dependent, a contradiction.

Case 2. $m - K$ is odd.

Set $s_{p_1} = \overline{t_{p_1}}$ so that s and t have the same parity.

Claim 3.3.8. *For any $u \in \{0,1\}^m$, if $u \notin Z'$ and $u_i = s_i$ for all $i \in \{p_2, ..., p_K\}$, then:*

1. *If u and s have the same parity, then Γ_u and Γ_s are linearly dependent.*
2. *If u and s have the opposite parity, then Γ_u and $\Gamma_{s \oplus e_{p_1}}$ are linearly dependent.*

Proof. The proof is the same as that of Claim 3.3.7, the only subtlety being that s and t now only necessarily agree on bits $p_2, ..., p_K$. By the argument there, all columns indexed by T_i^j lie in $\text{span}(Z_i)$ for $i = p_2, ..., p_K$. In particular, for all $j \notin \{p_1, ..., p_K\}$, all columns indexed by $\bigcap_{k=2}^K S_{p_k}^j$ lie in $\bigcap_{k=2}^K \text{span}(Z_{p_k}) = \text{span}(\{s, s \oplus e_{p_1}\})$.

So given $u \notin Z'$, write $\Gamma_u = \alpha\Gamma_s + \beta\Gamma_{s \oplus e_{p_1}}$. If u and s have the same parity, $\beta = 0$ by the parity condition, so $\Gamma_u \in \text{span}(\{s\})$. If u and s have the opposite parity, $\alpha = 0$ by the parity condition, so $\Gamma_u \in \text{span}(\{s \oplus e_{p_1}\})$. □

Pick any $j \notin \{p_1, ..., p_K\}$ and define $s^* = s \oplus e_j$ and $t^* = t \oplus e_j$. s^* and t^* both satisfy the hypotheses of Claim 3.3.8 and have parity opposite to that of s, so by the latter case of Claim 3.3.8, they are both linearly dependent with $\Gamma_{s \oplus e_{p_1}}$. But $\Gamma_{s \oplus e_{p_1}} \neq 0$ because $s \oplus e_{p_1} \in Z$ and the columns indexed by Z are linearly independent, so Γ_{s^*} and Γ_{t^*} are linearly dependent with each other.

To show Γ_s and Γ_t are linearly dependent, we wish to apply Corollary 2.3.7 to s^*, t^*, noting that $s^* \oplus t^* = e_j \oplus \sum_{j' \notin \{j, p_2, ..., p_k\}} e_{j'}$.

For any $j' \notin \{j, p_2, ..., p_K\}$, note that $s^* \oplus e_{j'}$ and $t^* \oplus e_{j'}$ both satisfy the hypotheses of Claim 3.3.8 and have the same parity as s, so by the former case of Claim 3.3.8, they are both linearly dependent with Γ_s. But $\Gamma_s \neq 0$ because $s \in Z$ and the columns indexed by Z are linearly independent, so $\Gamma_{s^* \oplus e_{j'}}$ and $\Gamma_{t^* \oplus e_{j'}}$ are linearly dependent with each other.

Applying Corollary 2.3.7 to s^* and t^*, it follows that Γ_s and Γ_t are linearly dependent, a contradiction. □

3.3.2 Existence of cluster submatrix

Completion of Inductive Step 2. As in inductive step 1, we may assume without loss of generality that for all clusters $Z \subsetneq \{0,1\}^m$, $\text{rank}(\Gamma_Z) \leq 2^K$. If $m = K + 1$, then by the inductive hypothesis that (3.1) holds for $\kappa = K$, we're done. So suppose $m > K + 1$.

By the second part of the inductive hypothesis, implication (3.2) holds for $1 \leq \kappa \leq K - 1$, so Γ contains a (K, m)-cluster Z of linearly independent columns $s + \{e_{p_1}, ..., e_{p_K}\}$.

As in inductive step 1, we can apply Lemma 3.3.6 to Z to see that all columns outside the span of the columns indexed by Z must be indexed by $Z' = t + \{e_{p_1}, ..., e_{p_K}\}$, where $t = \left(s \oplus \bigoplus_{i \neq p_1, ..., p_K} e_i\right)$. But $|Z'| = |Z| = 2^K$, and $\text{rank}(\Gamma) \geq 2^{K+1}$ by implication (3.1) for $\kappa = K$, so the columns indexed by $Z \cup Z'$ are linearly independent. Because $m > K + 1$, there exist columns not indexed by either Z or Z', and by Lemma 3.3.6 applied once to Z and once to Z', these columns are in both $\text{span}(Z)$ and $\text{span}(Z')$ and thus must be zero.

If s and t are of the same parity, apply Corollary 2.3.6 to s and t to find that Γ_s and Γ_t are linearly dependent, a contradiction.

If s and t are of opposite parity, apply Corollary 2.3.7 to $s \oplus e_j$ and $t \oplus e_j \oplus e_{p_1}$ for any $j \notin \{p_1, ..., p_K\}$ to find that Γ_s and $\Gamma_{t \oplus e_{p_1}}$ are linearly dependent, a contradiction. □

3.4 Group Property of Standard Signatures

We will now prove the following generalization of the group property result over domain size 4 due to Cai and Fu (Theorem 5.5, [8]):

Theorem 3.4.1. *If G is a generator matchgate of arity Kn with standard signature $\underline{G}$, and* $\text{rank}(\underline{G}(t)) = 2^K$ *for some t, then there exists a recognizer matchgate of arity Kn such that $\underline{G}(t)\underline{R}(t) = I_{2^K}$.*

Roughly, we invoke Theorem 3.3.3 to obtain a full-rank K-cluster submatrix G' of $\underline{G}(t)$ with column indices belonging to cluster $\zeta + \{e_{p_1}, ..., e_{p_K}\}$. Assume without loss of generality that $\zeta_{p_i} = 0$ for all $i \in [K]$. We will show that the matrix obtained by replacing G' in $\underline{G}(t)$ with $(G')^{-1}$ and the remaining entries with zeroes is the standard signature of some arity-Kn recognizer.

We first fix some notation. Denote $\underline{G}(t)$ by Γ. Suppose that nodes $p_1 < \cdots < p_m \in [Kn]$ belong to blocks before the t-th, and nodes $p_{m+1} < \cdots < p_K \in [Kn]$ belong to blocks after the t-th. For expository purposes, we wish to use a particular permutation $(q_1, ..., q_K)$ of $(p_1, ..., p_K)$, so for $i \leq m$, let $q_i = p_{m-i+1}$, and for $i > m$, let $q_i = p_{K+m-i+1}$ (see Figure 3.1). If the column indices of Γ are of the form $i_1 \cdots i_{K(n-1)}$, those of G' are of the form $i_{p_1} \cdots i_{p_K}$.

In [39], Li and Xia gave a constructive proof that in the character theory of matchgates, the $2^K \times 2^K$ *character* matrices of invertible K-input, K-output matchgates form a group under matrix multiplication. One can check that their construction carries over to show that the $2^K \times 2^K$ standard signatures of such matchgates likewise form a group, but unfortunately this is not enough to prove Theorem 3.4.1, as G' alone is merely a pseudo-signature and may not be realizable as the standard signature of a K-input, K-output matchgate. That said, Theorem 3.4.1 can still be proved with minor modifications to Li and Xia's approach.

We begin with a toy example motivating the notation in the previous paragraphs. Suppose that for each $i \in [K]$, there exists an edge of weight 1 such that the i-th external node in block t and external node q_i are both incident only to this edge. Note that in this case, G' is a symmetric permutation matrix and thus equal to its own inverse. We can easily construct a recognizer R out of G for which $\underline{G}(t)\underline{R}(t) = I_{2^K}$ as follows. Remove all non-external nodes of G, as well as all edges incident to nodes outside of block t and nodes $q_1, ..., q_K$. For external node i outside of block t such that $i \notin \{q_1, ..., q_K\}$, if $\zeta_i = 0$, attach a distinct edge of weight 1 to node i and designate the other endpoint of the edge as the ith external node of R; if $\zeta_i = 1$, attach a distinct path graph of length 2 consisting of two edges of weight 1, and denote the other endpoint of the path as the ith external node of R. By construction, in the $2^{K(n-1)} \times 2^K$ matrix $\underline{R}(t)$, the submatrix indexed by rows $q_1, ..., q_K$ is equal to G', and all other entries are zero. Because $G' = (G')^{-1}$, $\underline{G}(t)\underline{R}(t) = I_{2^K}$ as desired.

See Figure 3.1 for an example of this construction.

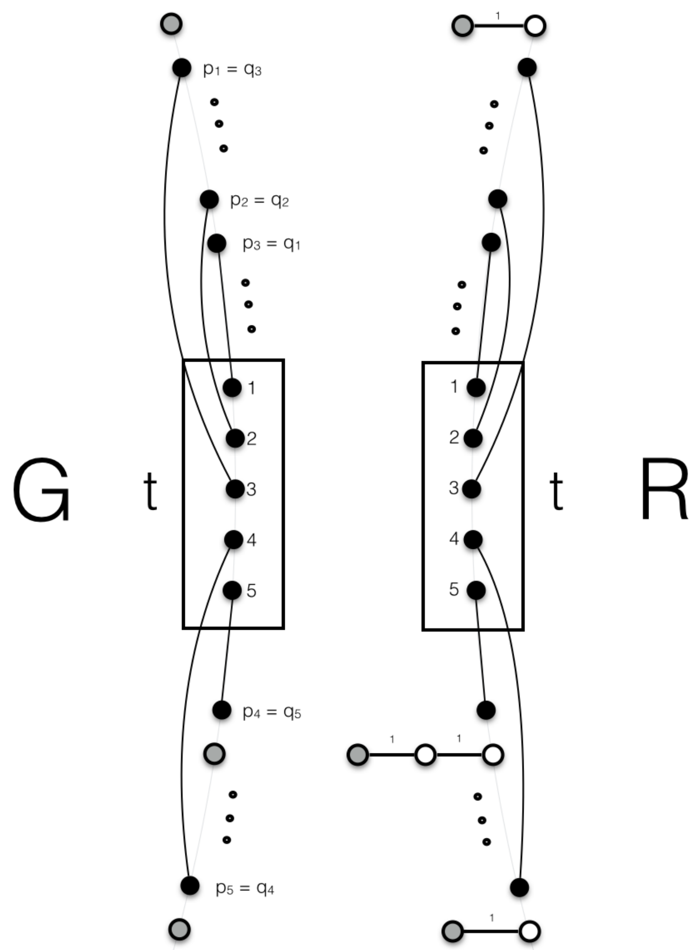


Figure 3.1: Toy example of G reduced at $i = 1, ..., K$. Here, $\ell = 5$, $K = 5$, $m = 3$. Black nodes in G denote nodes $1, ..., K$ in block t and nodes $q_1, ..., q_K$. External nodes of G and R are shown in black/gray.

Definition 3.4.2. For $i \in [K]$, if $i \leq m$ (resp. $i > m$), G is *reduced at* i if there exists an edge of weight 1 in G connecting the i-th external node in block t and external node q_i such that these two nodes are both incident only to that edge.

To prove Theorem 3.4.1, it is enough to reduce to the special case of the toy example above where Γ is realizable by a generator G reduced at all $i \in [K]$. The rest of this section will be dedicated to proving the following:

Lemma 3.4.3. *If Γ is the standard signature of a generator of arity Kn reduced at $1, \ldots, i$, there exist non-singular K-input, K-output transducers $L_1, \ldots, L_r$ and $K(n-1)$-input, $K(n-1)$-output transducers $R_1, \ldots, R_s$ such that $\underline{L_r} \cdots \underline{L_1} \cdot \Gamma \cdot \underline{R_1} \cdots \underline{R_s}$ is the standard signature of a generator reduced at $1, \ldots, i+1$.*

We first give a sufficient characterization of standard signatures of matchgates reduced at $1, \ldots, i$ in terms of the entries of their standard signatures.

The following terminology is borrowed from [39].

Definition 3.4.4. Let M be any $2^r \times 2^c$ matrix whose rows and columns are indexed by $\sigma \in \{0,1\}^r$ and $\tau \in \{0,1\}^c$ respectively. Γ^σ_τ is an *edge entry of M* iff $r + c - 2 \leq \mathrm{wt}(\sigma) + \mathrm{wt}(\tau) < r + c$.

Lemma 3.4.5. *Γ is the standard signature of a generator G that is reduced at i if Γ satisfies the following:*

1. $(G')^{1^K}_{1^K} = (G')^{1^K \oplus e_i}_{1^K \oplus e_{q_i}} = 1$.
2. $(G')^\sigma_\tau = 0$ *for all other edge entries of G' for which $\sigma \in \{1^K, 1^K \oplus e_i\}$ or $\tau \in \{1^K, 1^K \oplus e_{q_i}\}$.*

To show this, it suffices to prove the following useful consequence of the matchgate identities, first observed in [7] and translated below to our setting of standard signatures in matrix form.

Lemma 3.4.6 (Theorem 4.2, [7])**.** *If $(G')^{1^K}_{1^K} \neq 0$, the entries of G' are uniquely determined by $(G')^{1^K}_{1^K}$ and the edge entries of G'.*

Proof. Assume that these entries uniquely determine all entries Γ^σ_τ for which $\mathrm{wt}(\sigma) + \mathrm{wt}(\tau) \geq m$ for some $m \leq n-2$. We proceed by downward induction on m (by the parity condition, if m is even, the case of $m+1$ follows immediately from that of m). Then for $\sigma, \tau \in \{0,1\}^K$ such that $\mathrm{wt}(\sigma) + \mathrm{wt}(\tau) = m - 2$, apply (2.3) from Theorem 2.3.1 to $\sigma := \sigma$, $\zeta := \tau$, $\tau := 1^K$, and $\eta := 1^K$. One can check that the resulting identity consists of $(G')^\sigma_\tau \cdot (G')^{1^K}_{1^K}$ and terms which have already been uniquely determined by the inductive hypothesis, so because $(G')^{1^K}_{1^K} \neq 0$, we conclude that $(G')^\sigma_\tau$ is also uniquely determined. □

Observation 3. If Γ is the standard signature of a generator reduced at i, $(G')^{\sigma \oplus e_i}_{\tau \oplus e_{q_i}} = (G')^\sigma_\tau$, and if $\sigma_i \neq \tau_{q_i}$, $(G')^\sigma_\tau = 0$.

Proof. By hypothesis, external node i of block t and external node q_i are only connected to each other. If $\sigma_i \neq \tau_{q_i}$, $(G')^\sigma_\tau$ counts the number of perfect matchings of Γ where, among other conditions, exactly one of these two nodes is removed, and no such matching exists. On the other hand, if $\sigma_i = \tau_{q_i}$, then $(G')^\sigma_\tau$ and $(G')^{\sigma \oplus e_i}_{\tau \oplus e_{q_i}}$ count the number of perfect matchings in which, among other conditions, both of these two nodes are removed, or neither is. The number of perfect matchings in either scenario is the same. □

Let G'_i be the $2^{K-i} \times 2^{K-i}$ submatrix of G' consisting of entries $(G')^\sigma_\tau$ for which $\sigma_j = 0$ for $j = 1, \ldots, i$ and $\tau_j = 0$ for $j = q_1, \ldots, q_i$. If the row and column indices of G' are of the form $i_1 \cdots i_K$ and $i_{p_1} \cdots i_{p_K}$ respectively, the row and column indices of G'_i are of the form $i_{i+1} \cdots i_K$ and $i_{p_{i+1}} \cdots i_{p_K}$ respectively. When referring to the row (resp. column) of G' containing a row $i_{i+1} \cdots i_K$ (resp. column $i_{p_{i+1}} \cdots i_{p_K}$) of G'_i, we use the notation $0^i \circ i_{i+1} \cdots i_K$ (resp. $0^i \circ i_{p_{i+1}} \cdots i_{p_K}$) to denote its index in G'. For example, column $0^i \circ 1^{K-i}$ of G' is the column of G' indexed by $\sigma \in \{0,1\}^K$ for which $\sigma_{q_1} = \cdots = \sigma_{q_i} = 0$ and $\sigma_{q_{i+1}} = \cdots = \sigma_{q_K} = 1$, and this contains column 1^{K-i} of G'_i.

Corollary 3.4.7. *If Γ is the standard signature of a generator that is reduced at $1, \ldots, i$, then Γ is the standard signature of a generator reduced at $1, \ldots, i+1$ if Γ satisfies the following:*

1. $(G'_i)^{1^{K-i}}_{1^{K-i}} = (G'_i)^{1^{K-i} \oplus e_{i+1}}_{1^{K-i} \oplus e_{q_{i+1}}} = 1$
2. $(G'_i)^\sigma_\tau = 0$ *for all other edge entries of G'_i such that $\sigma \in \{1^{K-i}, 1^{K-i} \oplus e_{i+1}\}$ or $\tau \in \{1^{K-i}, 1^{K-i} \oplus e_{q_{i+1}}\}$.*

Proof. Apply Observation 3 to each of $1,...,i$, and invoke Lemma 3.4.5. □

Proof of Lemma 3.4.3. We execute the transformation $\Gamma = \Gamma^{(0)} \Rightarrow \Gamma^{(1)} \Rightarrow \Gamma^{(2)} \Rightarrow \Gamma^{(3)} \Rightarrow \Gamma^{(4)}$ outlined below.

1. $(\Gamma^{(0)} \Rightarrow \Gamma^{(1)})$: Turn the entry indexed by $(1^{K-i}, 1^{K-i})$ in G'_i to 1.
2. $(\Gamma^{(1)} \Rightarrow \Gamma^{(2)})$: Turn edge entries of G'_i in row or column 1^{K-i} to 0.
3. $(\Gamma^{(2)} \Rightarrow \Gamma^{(3)})$: Turn entry $(1^{K-i} \oplus e_{i+1}, 1^{K-i} \oplus e_{q_{i+1}})$ in G'_i to 1.
4. $(\Gamma^{(3)} \Rightarrow \Gamma^{(4)})$: Turn all other edge entries in G'_i in row $1^{K-i} \oplus e_{i+1}$ or column $1^{K-i} \oplus e_{q_{i+1}}$ to zero.

We need not care what these transformations do to entries outside of G', but we must ensure they preserve the fact that Γ is the standard signature of a generator reduced at $1,...,i$. To do this, for each matrix M by which we left- or right-multiply Γ, if σ does not index a row (resp. column) of G'_i, the only nonzero entry of M in row (resp. column) σ will be 1 in column (resp. row) σ.

In each step j, we will for convenience refer to $\Gamma^{(j-1)}$ as Γ.

Step 1. $(\Gamma^{(0)} \Rightarrow \Gamma^{(1)})$: Turn the entry indexed by $(1^{K-i}, 1^{K-i})$ in G'_i to 1.

We first show how to move a nonzero entry $c := (G'_i)^{\sigma^*}_{\tau^*}$ of G'_i into entry $(1^{K-i}, 1^{K-i})$ of G'_i.

For each j for which $i < j \le K$, we would like a $2^K \times 2^K$ standard signature C_j such that left-multiplication of Γ by C_j interchanges row σ in G' with $\sigma \oplus e_j$ for all $\sigma \in \{0,1\}^K$, and a $2^{K(n-1)} \times 2^{K(n-1)}$ standard signature D_j such that right-multiplication of Γ by D_j interchanges column τ in G' with $\tau \oplus e_{q_j}$. We could then define $L_1 = \prod_{j:\sigma^*_j=0} C_j$ and $R_1 = \prod_{j:\tau^*_j=0} D_j$, and $L_1 \cdot \Gamma R_1$ would have nonzero entry c at index $(1^{K-i}, 1^{K-i})$ of G'_i and still be the standard signature of a matchgate reduced at $1,...,i$.

C_j (resp. D_j) is the permutation matrix whose only nonzero entry in row $\sigma \in \{0,1\}^K$ (resp. $\sigma \in \{0,1\}^{K(n-1)}$) is 1 in column $\sigma \oplus e_j$ (resp. column $\sigma \oplus e_{q_j}$) if G'_i contains entries from Γ^σ (resp. Γ_σ), and 1 in column σ otherwise. C_j and D_j are certainly nonsingular.

To construct the K-input, K-output transducer realizing C_j as a standard signature, begin with a (K,K)-bipartite graph where for every $\nu \ne j$, left node ν and right node ν are connected by an edge of weight 1. Add an extra vertex between left node j and right node j, and draw a path of length two connecting these three vertices, where both edges of the path have weight 1. This construction is shown in Figure 3.2a.

The $K(n-1)$-input, $K(n-1)$-output transducer realizing D_j as a standard signature is similarly constructed, the only difference being that the bipartite graph has left and right vertex sets of size $K(n-1)$, and the path of length two is drawn between left node q_j and right node q_j.

Next, we want to scale all of the entries of $L_1 \Gamma R_1$ by a factor of $1/c$, so define $L_2(c)$ to be the diagonal matrix whose entry at index 1^{K-i} is $1/c$ and whose entries at all other indices are 1. Obviously $L_2(c)$ is nonsingular and satisfies the matchgate identities (2.3). We take $\Gamma^{(1)} = L_2(c) L_1 \Gamma^{(0)} R_1$.

Step 2. $(\Gamma^{(1)} \Rightarrow \Gamma^{(2)})$: Turn edge entries of G'_i in row or column 1^{K-i} to 0.

We demonstrate how to do this for edge entries in column 1^{K-i}. Firstly, edge entries $(1^{K-i} \oplus e_j, 1^{K-i})$ in G'_i are already zero by the parity condition.[1] To set each of the remaining edge entries in this column to zero, we will proceed in reverse lexicographic order over the rows $1^{K-i} \oplus e_j \oplus e_k$ of G'_i and at each step left-multiply Γ by a matrix $L_3^{j,k}$ which corresponds in Γ to subtracting $b := (G'_i)^{1^{K-i} \oplus e_j \oplus e_k}_{1^{K-i}}$ times row $0^i \circ 1^{K-i}$ of G' from row $0^i \circ (1^{K-i} \oplus e_j \oplus e_k)$ of G'.

$L_3^{j,k}$ must be a matrix whose nonzero entries include diagonal entries equal to 1 and entry $(0^i \circ (1^{K-i} \oplus e_j \oplus e_k), 0^i \circ 1^{K-i})$ equal to $-b$. A standard signature satisfying these conditions can be realized by the following matchgate: construct a (K,K)-bipartite graph in which left node ν and right node ν are connected by an edge of weight 1 for all ν, and left nodes j and k are also connected by an edge of weight $-b$. The standard signature of this matchgate is nonsingular. The construction is shown in Figure 3.2b.

$L_3^{j,k}$ additionally has nonzero entries $(\sigma \oplus e_j \oplus e_k, \sigma)$ equal to $-b$ for all $\sigma \in \{0,1\}^K$, i.e. left-multiplication by $L_3^{j,k}$ corresponds to subtracting b times row σ from row $\sigma \oplus e_j \oplus e_k$. These extraneous side effects do

[1] As characters of general matchgates with omittable nodes do not satisfy the parity condition necessarily, the proof of the group property in [39] requires an extra construction to turn edge entries $(1^{K-i} \oplus e_j, 1^{K-i})$ to zero. This is an instance where our proof of the group property for signatures is easier than that for characters.

not however affect any of the progress we've made as the rows $1^{K-i} \oplus e_j \oplus e_k$ of G'_i are taken in reverse lexicographic order.

The only issue is that the matchgate we have constructed is not necessarily planar. But by Lemma A.6.1 in Appendix A.6, there exists a planar matchgate with standard signature equal to $L_3^{j,k}$, except at nonzero off-diagonal entries other than $(0^i \circ (1^{K-i} \oplus e_j \oplus e_k), 0^i \circ 1^{K-i})$, where it may differ from $L_3^{j,k}$ by a factor of -1, but by the reasoning in the previous paragraph, this does not matter. Denote the standard signature of this planar matchgate by $L_3'^{j,k}$.

To achieve step 2 for edge entries in *rows* 1^{K-i} as well, we can define matrices $R_3'^{j,k}$ analogously. We can thus take $\Gamma^{(2)} = \left(\prod_{i+2\le j,k\le K} L_3'^{j,k}\right) \cdot \Gamma^{(1)} \cdot \left(\prod_{i+2\le j,k\le K} R_3'^{j,k}\right)$, where the indexing in the products respects the abovementioned reverse lexicographic order.

Step 3. $(\Gamma^{(2)} \Rightarrow \Gamma^{(3)})$: Turn entry $(1^{K-i} \oplus e_{i+1}, 1^{K-i} \oplus e_{q_{i+1}})$ in G'_i to 1.

Note that $c' := (G'_i)^{1^{K-i}\oplus e_j}_{1^{K-i}\oplus e_{q_k}}$ must be nonzero for some $j,k \in [K]$ or else (G'_i) is singular. As in Step 1, we will first left-multiply Γ by some L_4 to move this nonzero entry to row $1^{K-i} \oplus e_{i+1}$ and then right-multiply by some R_4 to move it to column $1^{K-i} \oplus e_{q_{i+1}}$. Unfortunately, multiplying by C_j or D_j defined in Step 1 would interfere with the progress made so far in Steps 1 and 2.

Instead, L_4 must be a matrix whose only nonzero entry in row (resp. column) $0^i \circ (1^{K-i} \oplus e_j)$ is 1 in column (resp. row) $0^i \circ (1^{K-i} \oplus e_{i+1})$, and whose only nonzero entry in row (resp. column) $0^i \circ (1^{K-i} \oplus e_{i+1})$ is 1 in column (resp. row) $0^i \circ (1^{K-i} \oplus e_j)$. A signature satisfying these conditions can be realized by the following matchgate: construct a (K,K)-bipartite graph in which left node ν and right node ν are connected by an edge of weight 1 for all $\nu \ne j, i+1$. Connect left node j to right node $i+1$ and left node $i+1$ to right node j with edges of weight 1. The standard signature L_4 of this is nonsingular. The construction is shown in Figure 3.2c.

L_4 also has nonzero entries $(i_1\cdots i_j\cdots i_{i+1}\cdots i_K, i_1\cdots i_{i+1}\cdots i_j\cdots i_K)$, so left-multiplication by L_4 corresponds to switching row $0^i \circ (i_1\cdots i_{i+1} i_{i+2}\cdots i_j\cdots i_K)$ with row $0^i \circ (i_1\cdots i_j i_{i+2}\cdots i_{i+1}\cdots i_K)$ for all $i_{i+1},\dots,i_K \in \{0,1\}$. Multiplication by L_4 affects neither the progress we've made on entry $(1^{K-i}, 1^{K-i})$ of G'_i because all bits in the row and column indices are equal, nor the progress on the edge entries in row 1^{K-i} and column 1^{K-i} of G'_i because these get swapped with each other and were already all zero, keeping them equal to zero.

As before, the only issue is that the matchgate constructed is not planar. But by Lemma A.6.2 in Appendix A.6, there exists a planar matchgate whose standard signature agrees with L_4 at every entry up to sign. By the above, that the nonzero entries other than $(0^i \circ 1^{K-i} \oplus e_j, 0^i \circ 1^{K-i} \oplus e_{i+1})$ and $(0^i \circ 1^{K-i} \oplus e_{i+1}, 0^i \circ 1^{K-i} \oplus e_j)$ may be -1 does not matter. Furthermore, if either entry $(0^i \circ 1^{K-i} \oplus e_j, 0^i \circ 1^{K-i} \oplus e_{i+1})$ or $(0^i \circ 1^{K-i} \oplus e_{i+1}, 0^i \circ 1^{K-i} \oplus e_j)$ were -1, at worst we may eventually need to replace c' with $-c'$, but the argument still holds. Denote the standard signature of this planar matchgate by L'_4. We can analogously define R'_4.

Next, we scale entry $(1^{K-i} \oplus e_{i+1}, 1^{K-i} \oplus e_{q_{i+1}})$ of G'_i by a factor of $1/c'$, so we take $\Gamma^{(3)} = L_2(c') L'_4 \Gamma^{(2)} R'_4$.

Step 4. $(\Gamma^{(3)} \Rightarrow \Gamma^{(4)})$: Turn all other edge entries in G'_i in row $1^{K-i} \oplus e_{i+1}$ or column $1^{K-i} \oplus e_{q_{i+1}}$ to zero.

We demonstrate how to do this for edge entries in column $1^{K-i} \oplus e_{q_{i+1}}$ of G'_i. To set each of the edge entries in this column to zero, we will proceed in reverse lexicographic order over the row indices $1^{K-i} \oplus e_j$ of G'_i and at each step left-multiply Γ by a matrix L_5^j which subtracts $b := (G'_i)^{1^{K-i}\oplus e_j}_{1^{K-i}\oplus e_{q_{i+1}}}$ times row $0^i \circ 1^{K-i}$ of G' from row $0^i \circ (1^{K-i} \oplus e_j)$ of G'.

L_5^j is a matrix whose nonzero entries include diagonal entries equal to 1 and entry $(0^i \circ (1^{K-i} \oplus e_j), 0^i \circ (1^{K-i} \oplus e_{i+1}))$ equal to $-b$. A signature satisfying these conditions can be realized by the following matchgate: construct a (K,K)-bipartite graph in which left node ν and right node ν are connected by an edge of weight 1 for all ν, and connect left node j to right node $i+1$ by an edge of weight $-b$. The standard signature of this matchgate is nonsingular. The construction is shown in Figure 3.2d.

L_5^j additionally has nonzero entries $(0^i \circ (\sigma \oplus e_j \oplus e_{i+1}), 0^i \circ \sigma)$ equal to $-b$ for all $\sigma \in \{0,1\}^{K-i}$ such that $\sigma_j = 1$ and $\sigma_{i+1} = 0$, i.e. for all such σ, left-multiplication by L_5^j corresponds to subtracting b times row $0^i \circ \sigma$ of G' from row $0^i \circ \sigma \oplus e_j \oplus e_{i+1}$ of G'. As before, these extraneous side effects do not affect the progress we've made in this step because the edge entries in column $1^{K-i} \oplus e_{q_{i+1}}$ of G'_i are being taken in reverse lexicographic order. They also certainly do not affect row 1^{K-i}, nor do they affect column 1^{K-i} as 1^{K-i} does not satisfy the above criteria for σ.

Again, the issue is that the matchgate constructed is not planar, but by Lemma A.6.3 in Appendix A.6, there exists a planar matchgate with standard signature equal to L_5^j except possibly at the nonzero off-diagonal entries other than $(0^i \circ (1^{K-i} \oplus e_j), 0^i \circ (1^{K-i} \oplus e_{i+1}))$, where it may differ by a factor of -1. Denote the standard signature of this planar matchgate by $L_5'^j$.

To achieve step 4 for edge entries in rows $1^{K-i} \oplus e_{i+1}$ as well, we can define matrices $R_5'^j$ analogously. We can thus take $\Gamma^{(4)} = \left(\prod_{i+2 \le j \le K} L_5'^j\right) \cdot \Gamma^{(3)} \cdot \left(\prod_{i+2 \le j \le K} R_5'^j\right)$, where the products respect the abovementioned reverse lexicographic order. □

3.5 Reducing to Domain Size 2^K

In this section we use Theorem 3.3.1 to reduce proving a basis collapse theorem over all domain sizes to proving one over domain sizes 2^K. The result we will prove is the following generalization of the main result in [21] whose strategy we follow.

Theorem 3.5.1. *Suppose Theorem 3.2.1 has been proven for domain size r. If recognizer signatures $R_1, ..., R_a$ and generator signatures $G_1, ..., G_b$ on domain size $k > r$ belonging to matchgrid Ω are simultaneously realizable on a $2^\ell \times k$ basis M of rank r and R_1 is of full rank, then there exists a basis M' of size at most $\lfloor \log_2 r \rfloor$ on which they are simultaneously realizable.*

We'll need some preliminaries before we can prove this. Express M as $\begin{pmatrix}\alpha_1 & \alpha_2 & \cdots & \alpha_k\end{pmatrix}$ where each α_i is a 2^ℓ-dimensional column vector. Let $i_1, ..., i_r \in [k]$ be column indices of M for which $M^{i_1 \cdots i_r} := \begin{pmatrix}\alpha_{i_1} & \alpha_{i_2} & \cdots & \alpha_{i_r}\end{pmatrix}$ is of full rank. Define *sub-signature* $R^{i_1 \cdots i_r}$ to consist of entries $(R_{j_1 \cdots j_n})$ of R ranging over all $j_1, ..., j_n \in \{i_1, ..., i_r\} \subset [k]$. We can define the sub-signature $G^{i_1 \cdots i_r}$ of a generator analogously. Equivalently,

$$R^{i_1 \cdots i_r} = \underline{R}(M^{i_1 \cdots i_r})^{\otimes n} \tag{3.3}$$

Lemma 3.5.2. *For a recognizer R realizable on basis M, if there exists t for which* $\operatorname{rank}(R(t)) \ge r$, *then* $\operatorname{rank}(R^{i_1 \cdots i_r}(t)) = r$.

Proof. By Lemma 2.2.14, $R(t) = (M^T)^{\otimes(n-1)} \underline{R}(t) M$, so $\operatorname{rank}(\underline{R}(t)) \ge r$. By (3.3) and Lemma 2.2.13, $R^{i_1 \cdots i_r}(t) = ((M^{i_1 \cdots i_r})^T)^{\otimes(n-1)} \underline{R}(t) M^{i_1 \cdots i_r}$, so $\operatorname{rank}(R^{i_1 \cdots i_r}(t)) = r$. □

For such a recognizer R, define for each $w \in [k]$ a nk^{n-1}-dimensional column vector b_w by

$$b_w = \begin{pmatrix} R_{w1\cdots11} & \cdots & R_{wk\cdots kk} & R_{1w\cdots11} & \cdots & R_{kw\cdots kk} & \cdots & R_{11\cdots1w} & \cdots & R_{kk\cdots kw} \end{pmatrix}^T \tag{3.4}$$

and define $A_{i_1 \cdots i_r}$ to be the $nk^{(n-1)} \times r$ matrix whose jth column is b_{i_j}.

Observation 4. $\operatorname{rank}(A_{i_1 \cdots i_r}) = r$.

Proof. $R^{i_1 \cdots i_r}(t)$ is a submatrix of $A_{i_1 \cdots i_r}$ and already has rank r by Lemma 3.5.2. □

Observation 5. As M has rank r, every column α_w can be expressed as a linear combination $\sum_{j=1}^r x_w^{i_j} \alpha_{i_j}$.

Denote the $r \times k$ matrix $\left(x_w^{i_j}\right)$ of these coefficients by $X_{i_1 \cdots i_r}$.

Lemma 3.5.3. *For each $w \in [k]$, $A_{i_1 \cdots i_r} X = b_w$ has the unique solution $X = \begin{pmatrix} x_w^{i_1} & \cdots & x_w^{i_r} \end{pmatrix}^T$.*

Proof. A solution for X exists and is unique because $\operatorname{rank}(A_{i_1 \cdots i_r}) = r$ by Observation 4. To check that the purported solution for X is correct, pick any entry $R_{j_1 \cdots j_{t-1} w j_t \cdots j_n}$ of b_w. By definition of recognizer signatures,

$$\begin{aligned}
R_{j_1\cdots j_{t-1}wj_t\cdots j_n} &= \langle \underline{R}, \alpha_{j_1}\otimes\cdots\otimes\alpha_{j_{t-1}}\otimes\alpha_w\otimes\alpha_{j_{t+1}}\otimes\cdots\otimes\alpha_{j_n}\rangle\\
&= \langle \underline{R}, \alpha_{j_1}\otimes\cdots\otimes\alpha_{j_{t-1}}\otimes\left(\sum_{j=1}^r x_w^{i_j}\alpha_{i_j}\right)\otimes\alpha_{j_{t+1}}\otimes\cdots\otimes\alpha_{j_n}\rangle\\
&= \sum_{j=1}^r x_w^{i_j}\cdot\left\langle \underline{R}, \alpha_{j_1}\otimes\cdots\otimes\alpha_{j_{t-1}}\otimes\alpha_{i_j}\otimes\alpha_{j_{t+1}}\otimes\cdots\otimes\alpha_{j_n}\right\rangle\\
&= \sum_{j=1}^r x_w^{i_j} R_{j_1\cdots j_{t-1}i_j j_{t+1}\cdots j_n}.
\end{aligned}$$

Here $\langle\cdot,\cdot\rangle$ denotes the inner product. □

The content of Lemma 3.5.3 is that to any such R we can get a matrix $X_{i_1\cdots i_r}$ without needing to know the actual basis M realizing R.

Lemma 3.5.4. *If* $\operatorname{rank}(R_1(t)) \geq r$ *for some* t*, then recognizers* $R_1,\ldots,R_a$ *are simultaneously realizable on some basis of rank* r *iff the following conditions hold:*

1. $\operatorname{rank}(R_1^{i_1\cdots i_r}(t)) = r$ *for some* $i_1,\ldots,i_r\in[k]$.
2. *There exists a unique* $r\times k$ *matrix* $X_{i_1\cdots i_r} = (x_w^{i_j})$ *such that* $A_{i_1\cdots i_r}X = b_w$ *has the solution* $X = \begin{pmatrix} x_w^{i_1} & \cdots & x_w^{i_r}\end{pmatrix}^T$ *for each* $w\in[k]$.
3. *There exists a* $2^\ell\times r$ *basis* $M_{(r)}$ *such that the* $R_j^{i_1\cdots i_r}$ *are simultaneously realizable on* $M_{(r)}$ *for all* $j\in[a]$.
4. $R_j = R_j^{i_1\cdots i_r}X_{i_1\cdots i_r}^{\otimes n}$ *for all* $j\in[a]$.

Proof. Suppose $R_1,\ldots,R_a$ are simultaneously realizable are some basis M. Conditions 1 and 2 follow from Lemma 3.5.2 and Lemma 3.5.3 respectively. Take $M_{(r)}$ to be $M^{i_1\cdots i_r}$, and condition 3 follows from the definition of sub-signature. By Observation 5, $X_{i_1\cdots i_r}$ satisfies $M_{(r)}X_{i_1\cdots i_r} = M$, so $R_j = \underline{R}_j M^{\otimes n} = \underline{R}_j M_{(r)}^{\otimes n}X_{i_1\cdots i_r}^{\otimes n} = R_j^{i_1\cdots i_r}X_{i_1\cdots i_r}^{\otimes n}$, and condition 4 follows.

Conversely, suppose conditions 1-4 hold. Condition 3 tells us that there is some $M_{(r)}$ for which $R_j^{i_1\cdots i_r} = \underline{R}_j\left(M_{(r)}\right)^{\otimes n}$ for all $j\in[a]$. Then

$$R_j = R_j^{i_1\cdots i_r}X_{i_1\cdots i_r}^{\otimes n} = \underline{R}_j(M_{(r)})^{\otimes n}X_{i_1\cdots i_r}^{\otimes n} = \underline{R}_j(M_{(r)}X_{i_1\cdots i_r})^{\otimes n},$$

so $R_1,\ldots,R_a$ are simultaneously realizable on $M := M_{(r)}X_{i_1,\ldots,i_r}$. □

Theorem 3.5.5. *If recognizer signatures* $R_1,\ldots,R_a$ *and generator signatures* $G_1,\ldots,G_b$ *in matchgrid* Ω *are simultaneously realizable on a basis of rank* r *and there exists* t *for which* $\operatorname{rank}(R_1(t))\geq r$*, then there exist recognizer signatures* $\check{R}_1,\ldots,\check{R}_a$ *and generator signatures* $\check{G}_1,\ldots,\check{G}_b$ *in matchgrid* Ω' *over domain size* r *that are simultaneously realizable on a* $2^\ell\times r$ *basis* $M_{(r)}$*. Furthermore,*

$$\operatorname{Holant}(\Omega) = \operatorname{Holant}(\Omega'). \tag{3.5}$$

Proof. We first construct $\check{R}_1,\ldots,\check{R}_a,\check{G}_1,\ldots,\check{G}_b$. $X_{i_1,\ldots,i_r}$ obtained from R_1 via Lemma 3.5.3 has rank r, so let $X'_{i_1\cdots i_r}$ be the $k\times k$ invertible matrix for which $X_{i_1\cdots i_r}X'_{i_1\cdots i_r} = \left(I_r \mid \mathbf{0}_{r\times(k-r)}\right)$, where I_r is the $r\times r$ identity matrix and $\mathbf{0}_{r\times(k-r)}$ denotes the $r\times(k-r)$ matrix consisting solely of zeroes. For each $j\in[a]$, let $R'_j = R_j(X'_{i_1,\ldots,i_r})$, and let $\check{R}_j$ be the sub-signature $(R'_j)^{1\cdots r}$. Likewise, for each $j\in[b]$, let G'_j be defined by $G_j = (X'_{i_1\cdots i_r})^{\otimes n}G'_j$, and let $\check{G}_j$ be the sub-signature $(G'_j)^{1\cdots r}$.

Claim 3.5.6. *For all* j*,* $\check{R}_j = R_j^{i_1\cdots i_r}$ *and* $\check{G}_j = G_j^{i_1\cdots i_r}$.

Proof. We need to check that

$$\check{R}_j = \underline{R}_j(M^{i_1\cdots i_r})^{\otimes n} \tag{3.6}$$

$$(M^{i_1\cdots i_r})^{\otimes n}\check{G}_j = \underline{G}_j. \tag{3.7}$$

Indeed,

$$\begin{aligned} R'_j &= \underline{R}_j M^{\otimes n}(X'_{i_1\cdots i_r})^{\otimes n} \\ &= \underline{R}_j(M^{i_1\cdots i_r})^{\otimes n} X^{\otimes n}_{i_1\cdots i_r}(X'_{i_1\cdots i_r})^{\otimes n} \\ &= \underline{R}_j\left(M^{i_1\cdots i_r} \mid \mathbf{0}_{r\times(k-r)}\right)^{\otimes n}, \end{aligned}$$

proving (3.6). Similarly,

$$\begin{aligned} \underline{G}_j &= M^{\otimes n} G_j \\ &= M^{\otimes n}(X'_{i_1\cdots i_r})^{\otimes n} G'_j \\ &= \left(M^{i_1\cdots i_r} \mid \mathbf{0}_{r\times(k-r)}\right)^{\otimes n} G'_j, \end{aligned}$$

proving (3.7). □

We conclude that $\check{R}_1, ..., \check{R}_a, \check{G}_1, ..., \check{G}_b$ are simultaneously realizable on the basis $M_{(r)} := M^{i_1\cdots i_r}$.

To check that the Holants agree, first note that if $R'_1, ..., R'_a, G'_1, ..., G'_b$ lie in a corresponding matchgrid Ω'', $\text{Holant}(\Omega) = \text{Holant}(\Omega'')$ because we're just applying a basis change from M to $MX'_{i_1\cdots i_r}$. And $\text{Holant}(\Omega') = \text{Holant}(\Omega'')$ because the operation of taking sub-signatures does not lose any information in this case, i.e. $(R'_j)_\sigma = 0$ for all $\sigma \in [k]^n \backslash [r]^n$. □

For the next two results, suppose recognizer signatures $R_1, ..., R_a$ and generator signatures $G_1, ..., G_b$ in matchgrid Ω are simultaneously realizable on a basis of rank r and there exists t for which $\text{rank}(R_1(t)) \geq r$.

Theorem 3.5.7. *If the recognizer signatures $\check{R}_1, ..., \check{R}_a$ and generator signatures $\check{G}_1, ..., \check{G}_b$ constructed in Theorem 3.5.5 are also simultaneously realizable on a $2^{\ell'} \times r$ basis $M'_{(r)}$ of rank r, then recognizer signatures $R_1, ..., R_a$ and generator signatures $G_1, ..., G_r$ are simultaneously realizable on the $2^{\ell'} \times k$ basis $M'_{(r)}X_{i_1\cdots i_r}$, where $X_{i_1\cdots i_r}$ is obtained from R_1 by Lemma 3.5.3.*

Proof.

$$R_j = R_j^{i_1\cdots i_r} X^{\otimes n}_{i_1\cdots i_r} = \check{R}_j X^{\otimes n}_{i_1\cdots i_r} = \underline{R}(M'_{(r)}X_{i_1\cdots i_r})^{\otimes n},$$

where the first equality holds by condition 4 of Lemma 3.5.4, the second by Claim 3.5.6, the third by definition of $M'_{(r)}$. Likewise, because

$$\check{G}_j = \left(I_r \mid \mathbf{0}_{r\times(k-r)}\right) G'_j = X_{i_1\cdots i_r} X'_{i_1\cdots i_r} G'_j,$$

we have that

$$\underline{G}_j = M^{\otimes n}_{(r)}\check{G}_j = (M_{(r)}X_{i_1\cdots i_r})^{\otimes n} X'^{\otimes n}_{i_1\cdots i_r} G'_j = (M_{(r)}X_{i_1\cdots i_r})^{\otimes n} G_j,$$

so we conclude that $R_1, ..., R_a, G_1, ..., G_b$ are indeed simultaneously realizable on $M_{(r)}X_{i_1\cdots i_r}$. □

We are now ready to prove Theorem 3.5.1.

Proof of Theorem 3.5.1. By Theorem 3.5.5, signatures $\check{R}_1, ..., \check{R}_a$ and $\check{G}_1, ..., \check{G}_b$ on domain size r are simultaneously realizable on a $2^\ell \times r$ basis.

By definition, $\check{R}_1 = R_1^{i_1\cdots i_r}$, and because R_1 was assumed to be full-rank, Lemma 3.5.2 tells us that $\check{R}_1$ is full-rank. Then by the hypothesis that Theorem 3.2.1 has already been proven for domain size r, there exists a $2^{\lfloor \log_2 r \rfloor} \times r$ basis $M'_{(r)}$ on which $\check{R}_1, ..., \check{R}_a$ and $\check{G}_1, ..., \check{G}_b$ are simultaneously realizable. By Theorem 3.5.7, $R_1, ..., R_a$ and $G_1, ..., G_b$ are simultaneously realizable on $2^{\lfloor \log_2 r \rfloor} \times k$ basis $M' := M'_{(r)}X_{i_1\cdots i_r}$. □

By Corollary 3.3.4 and Theorem 3.5.1, it remains to prove collapse theorems for holographic algorithms on domain sizes $k = 2^K$ and over bases of full rank, after which we get the following corollary.

Corollary 3.5.8. *Any holographic algorithm on a basis of size ℓ and domain size k not a power of 2 which uses at least one generator signature of full rank can be simulated on a basis of size at most $2^{\lfloor \log_2 k \rfloor}$.*

3.6 Collapse Theorem For Domain Size 2^K

The following is a direct generalization of the argument from Section 5.3 of [8], but we include it for completeness. We will take G to be a generator signature of full rank on domain size $k = 2^K$, basis M to be a $2^\ell \times 2^K$ matrix of rank 2^K, and $\underline{G} = M^{\otimes n} G$ to be the corresponding standard signature of arity $n\ell$. By Theorem 3.3.2 applied to the transpose of $\underline{G}(t)$, there exists a cluster $Z = s + \{e_{p_1}, ..., e_{p_K}\}$ of rows of full rank in $\underline{G}(t)$. Denote by M^Z the submatrix of M consisting of rows with indices in Z.

Lemma 3.6.1. *M^Z is invertible.*

Proof. The $(k, n\ell)$ cluster submatrix of $\underline{G}(t)$ of full rank whose existence is guaranteed by Theorem 3.3.2 is a submatrix of $M^Z G(t)(M^T)^{\otimes(n-1)}$, so M^Z has rank at least 2^K. But M^Z is a $2^K \times 2^K$ matrix, so M^Z is invertible. □

Following the notation of [8], now denote the column vector $(M^Z)^{\otimes n} G$ of dimension 2^{Kn} by $\underline{G}^{*\leftarrow Z}$ and the column vector $(M^Z)^{\otimes(t-1)} \otimes M \otimes (M^Z)^{\otimes(n-t)} \cdot G$ of dimension $2^{Kn+\ell-K}$ by $\underline{G}^{t^c\leftarrow Z}$. Because M^Z and $G(t)$ both have rank 2^K, $\underline{G}^{*\leftarrow Z}$ and $\underline{G}^{t^c\leftarrow Z}$ also have rank 2^K. We check that these can be realized as standard signatures.

Lemma 3.6.2. *$\underline{G}^{*\leftarrow Z}$ is the standard signature of a generator matchgate of arity Kn.*

Proof. Take the matchgate G, and in each block, attach an edge of weight 1 to external node i $(1 \le i \le \ell)$ if $s_i = 1$. In the matchgate G' we get from these operations, designate external nodes $p_1, ..., p_K$ in each block as the new external nodes of G'. The resulting matchgate realizes $\underline{G}^{*\leftarrow Z}$. □

Lemma 3.6.3. *$\underline{G}^{t^c\leftarrow Z}$ is the standard signature of a generator matchgate of arity $Kn - K + \ell$.*

Proof. The proof of Lemma 3.6.3 is almost identical to that of Lemma 3.6.2, except block t is treated differently. Take the matchgate G, and in each block except the t-th one, attach an edge of weight 1 to external node i $(1 \le i \le \ell)$ if $s_i = 1$. In the matchgate G' we get from these operations, take the external nodes to be all ℓ external nodes in block t, as well as nodes $p_1, ..., p_K$ in every other block. The resulting matchgate realizes $\underline{G}^{t^c\leftarrow Z}$. □

Now define $T = M(M^Z)^{-1}$. Here is the key step of the collapse theorem, making use of Theorem 3.4.1.

Lemma 3.6.4. *T is the standard signature of a K-input, ℓ-output transducer.*

Proof. We first express T in terms of $\underline{G}^{*\leftarrow Z}$ and $\underline{G}^{t^c\leftarrow Z}$. If the entries of $\underline{G}^{t^c\leftarrow Z}$ are indexed by $(i_{1,1}\cdots i_{1,K})\cdots(i_{t-1,1}\cdots i_{t-1,K})(i'_1\cdots i$ denote by $\underline{G}^{t^c\leftarrow Z}(t)$ the matrix form of $\underline{G}^{t^c\leftarrow Z}$ in which the rows are indexed by $i'_1\cdots i'_\ell$ and the columns are indexed by $(i_{1,1}\cdots i_{1,K})\cdots(i_{t-1,1}\cdots i_{t-1,K})(i_{t+1,1}\cdots i_{t+1,\ell})\cdots(i_{n,1}\cdots i_{n,K})$.

Observe that

$$\underline{G} = M^{\otimes n} G = T^{\otimes n}(M^Z)^{\otimes n} G = T^{\otimes n}\underline{G}^{*\leftarrow Z}$$

so that

$$\underline{G}^{t^c\leftarrow Z} = (T^Z)^{\otimes(t-1)} \otimes T \otimes (T^Z)^{\otimes(n-t)}\underline{G}^{*\leftarrow Z}. \tag{3.8}$$

Putting both sides of (3.8) in matrix form, we conclude that

$$\underline{G}^{t^c\leftarrow Z}(t) = T\underline{G}^{*\leftarrow Z}(t). \tag{3.9}$$

Applying Theorem 3.4.1 to the arity-Kn standard signature $\underline{G}^{*\leftarrow Z}$, we have a recognizer whose standard signature $\underline{R}$ satisfies $\underline{G}^{*\leftarrow Z}(t)\underline{R}(t) = I_{2^K}$. Right-multiplying both sides of (3.9) by $\underline{R}(t)$, we find that

$$\underline{G}^{t^c\leftarrow Z}(t)\underline{R}(t) = T.$$

Say that the generator realizing $\underline{G}^{t^c\leftarrow Z}$ as a standard signature has external nodes $X_{i,1}$, $X_{i,2}$, ..., $X_{i,K}$ in block i for each $i \ne t$, and external nodes $Y_{t,1}$,..., $Y_{t,\ell}$ in block t. Say that the generator realizing $\underline{R}$ as a standard signature has external nodes $Z_{i,1}, ..., Z_{i,K}$ in each block i.

Construct the transducer Γ realizing T as a standard signature by connecting $X_{i,j}$ with $Z_{i,j}$ for all $i \ne t$, $j \in [K]$. Designate $Y_{t,1}, ..., Y_{t,\ell}$ to be the output nodes of Γ and $Z_{t,1}$, ..., $Z_{t,K}$ to be the input nodes of Γ. □

From Theorem 3.6.4 we obtain the collapse theorem for domain size 2^K.

Theorem 3.6.5. *Any holographic algorithm on a basis of size ℓ and domain size 2^K which uses at least one generator signature of full rank can be simulated on a basis of size K.*

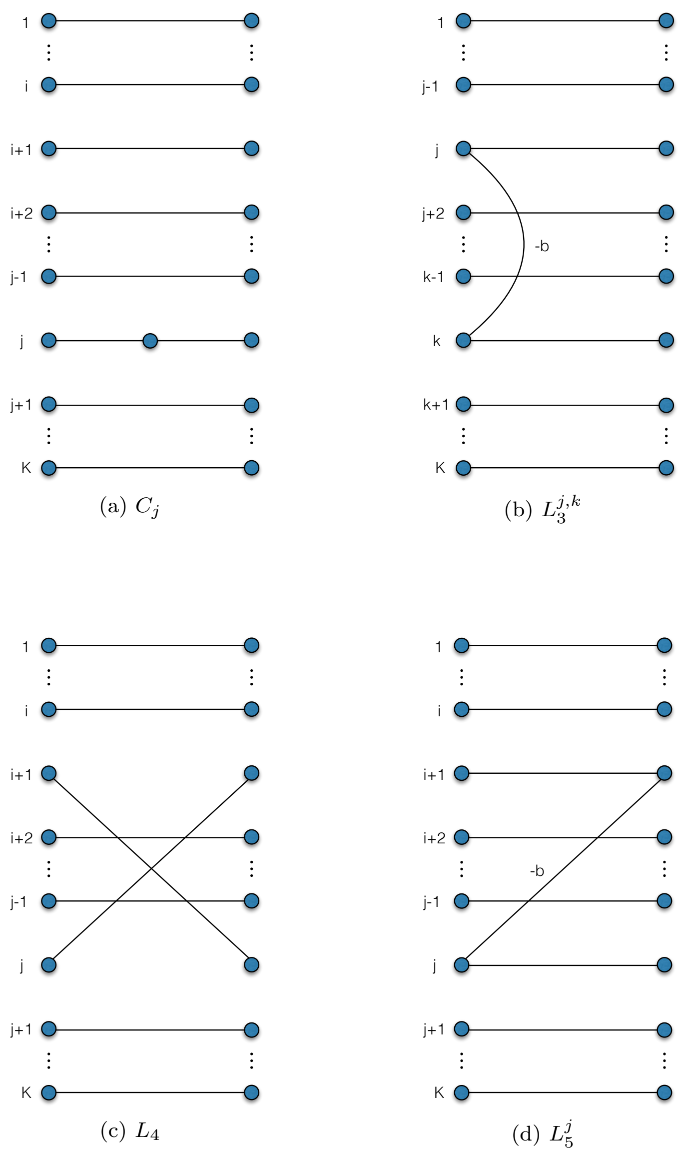
Proof. Suppose the holographic algorithm in question uses signatures R_i, G_j $(1 \le i \le r,\ 1 \le j \le g)$ defined by $\underline{R}_i M^{\otimes m_i} = R_i$ and $\underline{G}_j = M^{\otimes n_j} G_j$ over basis M. Say that G_1 has full rank, and let $Z = s + \{e_{p_1}, ..., e_{p_K}\}$ denote the full-rank (K, ℓ)-cluster of rows in G_1 which must exist by Theorem 3.3.3. By Lemma 3.6.4, $T := M(M^Z)^{-1}$ is the standard signature of some transducer matchgate Γ. Let $\underline{R}'_i = \underline{R}_i T^{\otimes m_i}$ and $\underline{G}'_j = \underline{G}_j^{*\leftarrow Z}$; by Lemma 2.2.4, $\underline{R}'_i$ is the standard signature of some recognizer, and by Lemma 3.6.3, $\underline{G}'_j$ is the standard signature of some generator. We conclude that the R_i, G_j can be simultaneously realized on the basis M^Z of size K. ☐

3.6.1 A final word on spinor varieties

Let $k = 2^K$. Recall that in Section 2.4, we noted that a geometric interpretation of the simultaneous realizability problem over higher domains was tricky because there is no longer a natural action of $\mathcal{M}_{2^\ell \times k}$ on $(\mathbb{C}^{k^\ell *})^{\otimes n}$. Theorem 3.2.1 tells us we can at least restrict to the case of $\ell = \lfloor \log_2 k \rfloor = K$, while Theorem 3.5.1 tells us we can assume M has maximal (row) rank.

It follows that, analogous to Question 2.4.10, we can interpret the simultaneous realizability problem over domain size k as a question of orbit containment. Denote the arities of $\mathbf{G}$ and $\mathbf{R}$ by m_1 and m_2 respectively.

Question 3.6.6 (Orbit Containment Problem- Domain Sizes $k = 2^K$)**.** Does the orbit $\mathrm{GL}(k) \cdot (\mathbb{S}_{m_1} \times \mathbb{S}_{m_2 *})$ contain the point (G, R)?



(a) C_j

(b) $L_3^{j,k}$

(c) L_4

(d) L_5^j

Figure 3.2: Transducers realizing row/column operations in steps 1-4

Part II

Complexity Theory

Chapter 4

Lower Bounds with Geometry

4.1 P versus NP and Determinant vs. Permanent

In this section we present Valiant's model for algebraic complexity and show that it is equivalent to a problem about determinantal complexity of the permanent, which will pave the way for the geometric approach to proving hardness results in this chapter. The mathematics here will be fairly formal and elementary, and the proofs of the results in this section are unrelated enough to the rest of this thesis that they can be skipped on a first reading.

Throughout we will be working over $\mathbb{C}$.

4.1.1 Algebraic Complexity Classes

In algebraic complexity, the analogues to Boolean circuits and Boolean functions will be *arithmetic circuits* and polynomials.

Definition 4.1.1. An *arithmetic circuit* C is a labeled directed acyclic graph which consists of vertices of indegree 0, called *input gates*, and vertices of indegree 2, called *computation gates*, among which is a single vertex of outdegree 0, called the *output gate*. Input gates are labeled with constants in $\mathbb{C}$ or formal variables, and computation gates are labeled by $\times$ or $+$. A circuit is inductively said to *compute* a polynomial as follows: an input gate computes the polynomial given by its label, and a computation gate $f \in \{\times, +\}$ with incoming neighbors that compute polynomials p and q respectively computes $f(p,q)$. We then say that C computes the polynomial computed by its output gate.

The number of gates is the *size* of C. The length of the longest path from an input gate to the output gate is the *depth* of C. The *degree* of C is the degree of the polynomial computed by C.

Whereas in the Boolean setting, complexity of a function is determined by the minimal resources needed for a Turing machine to compute it, in the algebraic setting, complexity of a polynomial is determined by the size and degree of the smallest circuit which can compute it. Algebraic complexity classes are merely sets of sequences of polynomials.

Definition 4.1.2. A sequence of polynomials (p_n) is said to lie in VP if there exists a corresponding sequence of circuits (C_n) computing (p_n) for which $\text{size}(C_n), \deg(C_n) \le \mathsf{poly}(n)$.

(p_n) is said to lie in VNP if there exists a polynomial ℓ and a corresponding sequence $(q_n) \in \mathsf{VP}$ for which $p_n(x) = \sum_{s \in \{0,1\}^{\ell(|x|)}} q_n(x,s)$.

We note the obvious inclusion $\mathsf{VP} \subset \mathsf{VNP}$. Valiant conjectured in [54] that the reverse inclusion does not hold:

Conjecture 4.1.3. $\mathsf{VNP} \not\subset \mathsf{VP}$.

Central to the classical theory of NP-hardness is the notion of *reductions* in which computation of one Boolean function is shown to reduce to that of another. In the algebraic setting, reductions will be modeled by affine linear transformations on the variables of polynomials.

Definition 4.1.4. Let $f \in \mathbb{C}[V], g \in \mathbb{C}[W]$. If there exists an affine linear map $\tilde{A} : V \to W$ for which $f = g \circ \tilde{A}$, then we say that f is a *projection* of g. We call $\tilde{A}$ the corresponding *g-representation of f*. For $g = \det_n$, we will call $\tilde{A}$ a *determinantal representation* of f.

A sequence (p_n) is a *p-projection* of (q_n) if there exists some $t(n) \le \mathsf{poly}(n)$ for which p_n is a projection of $q_{t(n)}$ for all n.

It is straightforward to check that VP and VNP are closed under p-projections. This allows us to define an algebraic notion of hardness with respect to these classes.

Definition 4.1.5. Let $\mathcal{C}$ be an algebraic complexity class which is closed under p-projections. A sequence (p_n) is *$\mathcal{C}$-hard* if any $(q_n) \in \mathcal{C}$ is a p-projection of (p_n). (p_n) is *$\mathcal{C}$-complete* if it both lies in $\mathcal{C}$ and is $\mathcal{C}$-hard.

As we alluded to in the introductory chapter, the key players in the geometric approach to complexity will be the determinant and permanent polynomials. Roughly speaking, the determinant is easy to compute, e.g. by Gaussian elimination, whereas the permanent seems not to be, and as we will show, they are suitable algebraic analogues of P and NP.

To motivate the former, let us first check that $\det_n$ and perm_n respectively lie in VP and VNP.

Lemma 4.1.6. $(det_n) \in \mathsf{VP}$.

Proof. We cannot simply invoke Gaussian elimination because arithmetic circuits do not support division by formal variables. Instead, the strategy is to exploit the following facts: 1) the determinant of a matrix M is the product of its eigenvalues λ_j, 2) the power sums $p_j := \sum_j \lambda_j^i$ can be computed efficiently, 3) elementary symmetric polynomials e_k in the eigenvalues can be recovered efficiently from these power sums.

For 2), note that $p_j = \mathrm{trace}(M^i)$. In the circuit that we construct for $\det_n$, we will construct gates computing $p_1(\lambda), ..., p_n(\lambda)$ by first inductively constructing gates computing all n^3 entries in $M, M^2, ..., M^n$. If the gates for M^{i-1} have already been constructed, it takes $O(n^3)$ additional gates to construct those for M^i, for a total of $O(n^4)$ intermediate gates. For each M^i, we include an extra addition gate, with incoming neighbors the diagonal entries of M^i, that computes $p_j(\lambda)$. It thus takes $\sum_{i=1}^n O(n^4) = O(n^5)$ gates to compute $p_1(\lambda), ..., p_n(\lambda)$.

For 3), the Newton-Girard identities say that

$$e_k = \frac{1}{k} \sum_{i=1}^{k} (-1)^{i-1} e_{k-i} \cdot p_i,$$

so to compute $e_n(\lambda) = \det_n$, we inductively construct gates for all $e_k(\lambda)$. $e_1(\lambda)$ is merely $p_1(\lambda)$. If the gates for $e_1(\lambda), ..., e_{k-1}(\lambda)$ have been constructed, then k multiplication gates above the gates for e_{k-i} and p_i, $O(k)$ addition gates summing over these multiplication gates, and a final multiplication gate between this sum and the scalar $1/k$ suffice to compute $e_k(\lambda)$, for a total of $O(k)$ gates.

To conclude, the circuit we've constructed requires $O(n^5) + \sum_{i=1}^n O(k) = O(n^5)$ gates, and $\det_n \in \mathsf{VP}$ as desired. □

Lemma 4.1.7 ([55])**.** $(\mathrm{perm}_n) \in \mathsf{VNP}$.

Proof of Lemma 4.1.7. We wish to define a sequence of polynomials (q_n) taking as inputs $M \in \mathcal{M}_{n\times n}(\mathbb{C})$ and $S \in \mathcal{M}_{n\times n}(\{0,1\})$ such that $q_n(M, S) = \prod_{i=1}^n M^i_{\sigma(i)}$ if S is a permutation matrix corresponding to permutation σ, or $q_n(M, S) = 0$ otherwise. If we can show that $(q_n) \in \mathsf{VP}$, we are done because $\mathrm{perm}_n(M) = \sum_S q_n(M, S)$.

A matrix S is a permutation matrix if and only if 1) no row or column has more than a single 1, 2) no row or column has all zeroes. S satisfies 1) if and only if $q^1_n(S) = \prod_{i,j,k,\ell}(1 - S^i_j S^k_\ell)$ does not vanish, where the product is taken over all $(i, j) \ne (k, \ell)$ for which $i = k$ or $j = \ell$. S satisfies 2) if and only if $q^2_n(S) = \prod_i \sum_j S^i_j$ does not vanish. So $(q^1_n \cdot q^2_n)(S)$ equals 1 if S is a permutation matrix and vanishes otherwise. So

$$q_n(M, S) = (q^1_n \cdot q^2_n)(S) \cdot \prod_{i=1}^{n} \sum_{j=1}^{n} M^i_j S^i_j$$

clearly satisfies the desired properties. Moreover, it is clear that each factor requires at most $O(n^3)$ gates, so (q_n) thus defined lies in VP. □

Unfortunately, it is unknown whether $\det_n$ is VP-hard, and in fact there are no known natural polynomials that are complete for VP. The determinant is an important enough polynomial that people decided to modify the definition of VP to get a class for which $\det_n$ was complete. The definition of this alternative class VP_{ws} is somewhat artificial, and the proof of completeness, due to Malod and Portier [40], is fairly irrelevant to the rest of this thesis. For this reason, we relegate it to Section A.2 of the appendix and merely state their result and show $(\mathrm{perm}_n) \in \mathsf{VNP}$.

Theorem 4.1.8 ([40])**.** *$(\det_n)$ is VP_{ws} complete.*

We are led to conjecture that $\mathsf{VNP} \not\subseteq \mathsf{VP}_{ws}$. By Theorem 4.1.8, it is enough to show that no sequence in VNP is a p-projection of $(\det_n)$. To this end, we introduce the following important definitions.

Definition 4.1.9. The *determinantal complexity* of $p \in S^d V^*$, denoted $\mathrm{dc}(p)$, is the smallest n for which p is a projection of $\det_n$.

On the other hand, Valiant [55] showed that the permanent polynomial perfectly captures the class VNP. We will omit the proof of this result.

Theorem 4.1.10 ([55])**.** *(perm_n) is VNP-complete.*

By Theorem 4.1.10, $\mathsf{VNP} \not\subseteq \mathsf{VP}_{ws}$ is therefore equivalent to the following conjecture.

Conjecture 4.1.11 (Valiant's conjecture)**.** $\mathrm{dc}(\mathrm{perm}_m)$ grows faster than any polynomial in m.

4.1.2 Valiant's conjecture as an orbit containment problem

We can frame this as an orbit containment problem as follows. Let $V = \mathbb{C}^{m^2}$ and $W = \mathbb{C}^{n^2}$.

Definition 4.1.12. Take ℓ a linear coordinate on $\mathbb{C}^1$ and take any linear inclusion $\mathbb{C}^1 \oplus V \hookrightarrow W$ so that $\ell^{n-m}\,\mathrm{perm}_m \in S^n W^*$. Then $\mathrm{perm}_m = \det_n \circ \tilde{A}$ if and only if

$$\mathrm{End}(W) \cdot \ell^{n-m}\,\mathrm{perm}_m \subset \mathrm{GL}(W) \cdot \det_n. \tag{4.1}$$

Proof. Suppose $\mathrm{perm}_m = \det_n \circ \tilde{A}$. Pick any basis $\{x_{ij}\}$ of W^* and regard $\tilde{A}$ as an $n \times n$ matrix with entries affine linear in $\{x_{ij}\}$. Homogenizing using ℓ, we find that $\ell^{n-m}\,\mathrm{perm}_m = \det_n \circ B$, where B is a matrix with entries homogeneous linear forms in $\{x_{ij}\}$ and ℓ, so (4.1) follows.

In the other direction, (4.1) implies that there exists some B as above. But then

$$\ell^{-m}\,\mathrm{perm}_m = \ell^{-n}\det_n \circ B = \det_n \circ (\ell^{-1} \cdot B),$$

so we take $\tilde{A}$ to be the $n \times n$ matrix where each x_{ij}/ℓ_j in $\ell^{-1} \cdot B$ is replaced by x_{ij}. Then $\mathrm{perm}_m = \det_n \circ \tilde{A}$ as desired. □

This perspective will motivate the GCT approach to Valiant's conjecture which we present in Section 4.4.

4.1.3 Symmetries of $\det_n$ and perm_m

Let G be a reductive group acting on a vector space W. G has an induced action on $S^d W^*$ given by $g \cdot p(v) = p(g^{-1} \cdot v)$.

Definition 4.1.13. The *symmetry group* of a polynomial $p \in S^d V^*$ is defined by

$$G_p := \{g \in G \mid g \cdot p = p\}.$$

The *projective symmetry group* of p is defined by

$$\mathbb{G}_p := \{g \in G \mid [g \cdot p] \in [p]\}.$$

Define the *character* of p to be the group homomorphism $\chi_p : \mathbb{G}_p \to \mathbb{C}^*$ for which $g \cdot p(y) = \chi_p(g)p(y)$.

It is a well-known theorem of Frobenius that transpose and left- and right-multiplication by elements of $\mathrm{GL}(n)$ are the only projective symmetries of the determinant.

Theorem 4.1.14 ([20])**.** *Let $\mathbb{C}^{n^2} = E \otimes F$. Then*

$$G_{det_n} \simeq \left(\left(SL(E) \times SL(F) \right) / \mu_n \right) \ltimes \mathbb{Z}_2$$
$$\mathbb{G}_{det_n} \simeq \left(\left(GL(E) \times GL(F) \right) / \mathbb{C}^* \right) \ltimes \mathbb{Z}_2.$$

On the other hand, it is also known that transpose and left- and right-multiplication by permutation matrices with entries scaled arbitrarily are the only projective symmetries of the permanent.

Theorem 4.1.15 ([41])**.** *Let $\mathbb{C}^{m^2} = E \otimes F$, and let T_E denote the maximal torus of diagonal matrices and $N(T_E)$ its normalizer, which we know to be $T_E \ltimes \mathcal{S}_m$. Then*

$$G_{\mathrm{perm}_m} \simeq \left(\left(N(T_E) \times N(T_F) \right) / \mu_m \right) \ltimes \mathbb{Z}_2$$
$$\mathbb{G}_{\mathrm{perm}_m} \simeq \left(\left(N(T_E) \times N(T_F) \right) / \mathbb{C}^* \right) \ltimes \mathbb{Z}_2.$$

One useful property of the determinant and permanent is that they are characterized by their symmetries.

Definition 4.1.16. A polynomial $p \in S^n W^*$ is said to be *characterized by its symmetries* if for any $q \in S^n W^*$ for which $G_p \subseteq G_q$, there exists some $\lambda \in \mathbb{C}^*$ for which $p = \lambda \cdot q$.

We will show that this is the case for $p = \det_n$ and $p = \mathrm{perm}_m$; the key step is the *Cauchy formula*, which just follows by Schur-Weyl duality (see Theorem A.4.1 in Appendix A.4).

Lemma 4.1.17 (Cauchy formula)**.** *Let A, B be vector spaces. As a $GL(A) \times GL(B)$-module, $S^d(A \otimes B)$ decomposes into irreducibles as*

$$S^n(A \otimes B) = \bigoplus_{\pi} S_\pi A \otimes S_\pi B,$$

where the sum is taken over all partitions π of d.

Proof. Using the definition of the symmetric power, rearranging, and applying Schur-Weyl duality, we see that

$$\begin{aligned} S^d(A \otimes B) &= \bigoplus_{\pi,\mu} \left([\pi] \otimes S_\pi A \otimes [\mu] \otimes S_\mu B \right)^{S_d} \\ &= \bigoplus_{\pi,\mu} \left([\pi] \otimes [\mu] \right)^{S_d} \otimes S_\pi A \otimes S_\mu B. \end{aligned}$$

S_d-modules are self-dual so that $([\pi] \otimes [\mu])^{S_d} = \mathrm{Hom}_{S_d}([\pi], \mu])$, so by irreducibility of the S_d-modules $[\pi], [\mu]$ and Schur's lemma, $\mathrm{Hom}_{S_d}([\pi], [\mu])$ is a copy of $\mathbb{C}$ if $\pi = \mu$ and 0 otherwise. ☐

Lemma 4.1.18. *det_n and $perm_n$ are characterized by their symmetries.*

Proof. Take A, B in the Cauchy formula to both be copies of $\mathbb{C}^n$. Then a highest weight vector of $S_{1^n}(A) \otimes S_{1^n}(B) = \Lambda^n A \otimes \Lambda^n B$ is the unique vector, up to a scalar factor, invariant under $\mathrm{SL}(A) \times \mathrm{SL}(B)$. So in fact $\det_n$ is characterized just by $G^o_{\det_n}$. Likewise, the permanent lies in the unique line of weight $(1, ..., 1)_A \times (1, ..., 1)_B$ inside $S_n(A) \otimes S_n(B) = S^n(A) \otimes S^n(B)$ which is invariant under $(T_A \ltimes W_A) \times (T_B \ltimes W_B)$. ☐

4.2 A First Lower Bound with Differential Geometry

Before introducing the machinery of geometric complexity theory, we first explore the geometry that can arise from the problem of determining determinantal complexity of polynomials. In this section, we present Mignon-Ressayre's quadratic simple lower bound on the determinantal complexity of the permanent using classical differential geometry. Roughly speaking, we show that the rank of the second fundamental form is $2n$ at any smooth point of $\{\det_n = 0\}$, but m^2 at a carefully chosen point of $\{\mathrm{perm}_n = 0\}$, implying a quadratic lower bound for determinantal complexity of the permanent.

4.2.1 Second fundamental form

We briefly recall the setup for the second fundamental form of a hypersurface $Z \subset \mathbb{P}^n$. Define the *Gauss map* $\mathcal{G}_Z : Z \to \mathbb{P}^{n*}$ to send point $x \in Z$ to $T_x Z$. If $Z = V(f)$ for some polynomial $f \in \mathbb{C}[z_0, ..., z_n]$, then in terms of coordinates, the Gauss map is given by

$$\mathcal{G}_Z : x \mapsto \left[\frac{\partial F}{\partial z_0}(x), ..., \frac{\partial F}{\partial z_n}(x)\right].$$

Identifying tangent planes in $\mathbb{P}^{n*}$ with normal vectors in S^n and noting that $T_{\mathcal{G}_Z(x)} S^n$ is parallel to $T_x Z$, we can regard the differential of the Gauss map as an endomorphism

$$d(\mathcal{G}_Z)_x : T_x Z \to T_x Z.$$

This endomorphism gives rise to the quadratic form $Q(v) = -d(\mathcal{G}_Z)_x(v) \cdot v$; its corresponding bilinear form $(v, w) \mapsto -d(\mathcal{G}_Z)_x(v) \cdot w$ corresponds to a map which we denote by $T_x^2 f \in \mathrm{Hom}(T_x Z, T_x^* Z)$. This is the *second fundamental form* of Z at x. In terms of coordinates, the second fundamental form is given by the Hessian of f, namely

$$T_x^2 f = \left\{\frac{\partial f}{\partial x_i \partial x_j}\right\}_{1 \le i,j \le n}.$$

Lemma 4.2.1. *If $\iota : V \hookrightarrow \mathbb{P}^n$ is an inclusion of some subspace V, then* $\mathrm{rank}(T_x^2 f) \ge \mathrm{rank}(T_x^2(f \circ \iota))$.

Proof. The dual map ι^* corresponding to restriction of linear functionals is clearly linear. $T_x^2(f \circ \iota)$ is the restriction to V of $\iota^* \circ T_x^2 f$, so the rank of the former cannot exceed that of the latter. ☐

We will borrow the notation of Mignon-Ressayre and denote the second fundamental form of a hypersurface $V(f)$ at a point p by $T_p^2 f$.

4.2.2 Mignon-Ressayre's bound

It remains just to compute the second fundamental forms of the determinantal and permanental hypersurfaces. Fix a generic $n \times n$ matrix of indeterminates A; for row indices i, i' and column indices j, j', denote by $A_{ii',jj'}$ the matrix obtained from removing rows i, j and columns i', j'. By construction, we have the following.

Lemma 4.2.2. *For row indices i, i' and column indices j, j',*

$$(T_A^2 \mathrm{perm}_n)^{ii',jj'} = \begin{cases} 0, & \text{for } i = i' \text{ or } j = j' \\ \mathrm{perm}_{n-2}(A_{ii',jj'}), & \text{otherwise} \end{cases}$$

$$(T_A^2 det_n)_{ii',jj'} = \begin{cases} 0, & \text{for } i = i' \text{ or } j = j' \\ (-1)^{i+i'+j+j'} det_{n-2}(A_{ii',jj'}), & \text{otherwise} \end{cases}$$

Lemma 4.2.3. *For any singular $n \times n$ matrix M,* $\mathrm{rank}(T_M^2 det_n) \le 2n$.

Proof. Up to a scalar factor, translating M by $\mathbb{G}_{\det_n}$ preserves $T_M^2 \det_n$, so we may assume that M is the diagonal matrix with k 0's and $n-k$ 1's, in that order.

If $k > 2$, then $T_M^2 \det_n$ is the zero matrix. If $k = 2$, then the only nonzero entries are $(11, 22)$, $(22, 11)$, $(12, 21)$, and $(21, 12)$. If $k = 1$, then the nonzero entries are $(11, ii)$, $(ii, 11)$, $(1i, i1)$, and $(i1, 1i)$ for all $i \ge 2$. This matrix's rows are spanned by rows 11, nn, and $1i$ and $i1$ for all $i > 1$, so $\mathrm{rank}(T_M^2 \det_n) \le 2n$ as claimed. ☐

Lemma 4.2.4. *There exists a point M for which* $\mathrm{perm}_n(M) = 0$ *such that* $T_M^2 \mathrm{perm}_n$ *is invertible, so* $\mathrm{rank}(T_M^2 \mathrm{perm}_n) = n^2$.

To prove this, we will use the following convenient fact.

Proposition 4.2.5. *An $MN \times MN$ matrix T with the block decomposition*

$$\begin{pmatrix} 0 & U & U & \cdots & U \\ U & 0 & V & \cdots & V \\ U & V & 0 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & V \\ U & V & \cdots & V & 0 \end{pmatrix}$$

for invertible $N \times N$ matrices U, V is invertible.

Proof. We can left-multiply T by the diagonal $(U^{-1}, I, ..., I)$ and thereby assume that $U = I$. Suppose to the contrary that $\ker(T)$ has a nonzero element $(\mathbf{v_1}\ \mathbf{v_2} \cdots \mathbf{v_M})$ so that

$$\mathbf{v_2} + \cdots + \mathbf{v_M} = 0 \tag{4.2}$$

$$\mathbf{v_1} + V\mathbf{v_2} + \cdots + V\hat{\mathbf{v_i}} + \cdots + V\mathbf{v_M} = 0 \tag{4.3}$$

for each $2 \le i \le M$. (4.2) applied to (4.3) implies that for each i, $\mathbf{v_1} = V\mathbf{v_i}$, and this applied to (4.2) implies $(M-1)\mathbf{v_1} = 0$. So $\mathbf{v_1} = \cdots = \mathbf{v_M} = 0$ as desired. □

Proof of Lemma 4.2.4. The particular point that we will evaluate $T^2 \operatorname{perm}_n$ on is the matrix

$$M = \begin{pmatrix} 1-d & 1 & \cdots & 1 \\ 1 & 1 & \cdots & 1 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 1 & \cdots & 1 \end{pmatrix}.$$

One readily checks that M indeed lies on the permanental hypersurface; denoting $T^2_M \operatorname{perm}_n$ by T, we can also check that

$$T_{ii',jj'} = \begin{cases} (d-2)!, & \text{if } 1 \in \{i, j, i', j'\} \\ -2(d-3)! & \text{otherwise} \end{cases}$$

Finally, we apply Proposition 4.2.5 to

$$U = (d-2)! \begin{pmatrix} 0 & 1 & \cdots & 1 \\ 1 & 0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 1 \\ 1 & \cdots & 1 & 0 \end{pmatrix}, \quad V = (d-3)! \begin{pmatrix} 0 & d-2 & d-2 & \cdots & d-2 \\ d-2 & 0 & -2 & \cdots & -2 \\ d-2 & -2 & 0 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & -2 \\ d-2 & -2 & \cdots & -2 & 0 \end{pmatrix},$$

which are themselves invertible by Proposition 4.2.5, to conclude that $T^2_M \operatorname{perm}_n$ is invertible. □

Having exhibited this discrepancy between second fundamental forms, we can now prove a quadratic separation between permanent and determinant.

Theorem 4.2.6. $dc(\operatorname{perm}_m) \ge m^2/2$.

Proof. Suppose there were some affine $\tilde{A} : \mathcal{M}_m(\mathbb{C}) \to \mathcal{M}_n(\mathbb{C})$ for which $\operatorname{perm}_m = \det_n \circ \tilde{A}$. We claim that $\tilde{A}$ is injective. If to the contrary there were some nonzero ϕ in its kernel, then perm_m remains invariant under translation by ϕ so that for any $M \in \mathcal{M}_m(\mathbb{C})$, the functional corresponding to $T^2_M \operatorname{perm}_m$ lies in the hyperplane $\operatorname{Ann}(\phi) \subset \mathcal{M}_m(\mathbb{C})^*$. This implies that the image of $T^2_M \operatorname{perm}_m$ lies in $\operatorname{Ann}(\phi)$ so that $T^2_M \operatorname{perm} m$ is not of full rank, a contradiction.

Denote by p the restriction of $\det_n$ to the image of $\tilde{A}$. Because $\tilde{A}$ must be injective, there is an isomorphism between $V(p)$ and $V(\operatorname{perm}_m)$. So apply Lemma 4.2.4 to get a point M on the former for which $\operatorname{rank}(T^2_M p) = m^2$. By Lemmas 4.2.1 and 4.2.3, $m^2 \le 2n$, giving the desired lower bound on $\operatorname{dc}(\operatorname{perm}_m)$. □

4.3 A Better Lower Bound Assuming Symmetries

First fix some notation. We will write determinantal representations $\tilde{A}$ as $\Lambda + A$ for fixed matrix $\Lambda \in M_n(\mathbb{C})$ and linear map $A : \mathcal{M}_m(\mathbb{C}) \to \mathcal{M}_n(\mathbb{C})$. We will be interested in the projective symmetries of the determinant polynomial which also fix the determinantal representation.

Definition 4.3.1. For $\tilde{A} = \Lambda + A$ a determinantal representation of $P \in S^m V^*$, define the *symmetric group of the determinantal representation* $\tilde{A}$ to be $\mathbb{G}_A = \{g \in \mathbb{G}_{\det_n} \mid g \cdot A(V) = A(V),\ g \cdot \Lambda = \Lambda\}$.

4.3.1 Respecting symmetry

Roughly speaking, we say that a determinantal representation of P *respects the symmetries* of P if the symmetries of the P can be recovered from those of the determinant polynomial, that is, from applying row/column operations and transposes to $\tilde{A}$.

By definition the action of $\mathbb{G}_A$ on $M_n(\mathbb{C})$ fixes $A(V)$, so formally define the representation $\rho_A : \mathbb{G}_A \to \mathrm{GL}(A(V))$ given by restricting the action of $\mathbb{G}_A$ to $A(V)$. If P is nondegenerate, then $A : V \to A(V)$ is bijective, so we get a representation $\bar{\rho}_A : \mathbb{G}_A \to \mathrm{GL}(V)$ which sends g to the linear transformation $v \mapsto A^{-1} \circ \rho_A(g) \circ A(v)$.

Definition 4.3.2. $\tilde{A}$ *respects the symmetries* of P if $\mathrm{Im}(\bar{\rho}_A) = \mathbb{G}_P$. More generally, $\tilde{A}$ *respects* $G \subset \mathbb{G}_P$ if $G \subset \mathrm{Im}(\bar{\rho}_A)$.

Example 4.3.3. Consider the nondegenerate quadratic form $Q = \sum_{j=1}^{M} z_j^2 \in S^2\mathbb{C}^{M*}$. It is well-known that $\mathbb{G}_Q = \mathbb{C}^* \times O(M)$, where $O(M)$ is the orthogonal group on $\mathbb{C}^M$. We claim that the determinantal representation $\tilde{A}$ given by

$$Q = \det_{M+1}\begin{pmatrix} 0 & -Z^T \\ Z & I \end{pmatrix}$$

respects the symmetries of Q, where $Z = (x_1, ..., x_M)^T$ and I is the $M \times M$ identity matrix. Pick one such symmetry $(\lambda, B) \in \mathbb{G}_Q$ and consider the corresponding symmetry of $\det_{M+1}$ given by $g : M \mapsto \begin{pmatrix} \lambda & 0 \\ 0 & B \end{pmatrix} Z \begin{pmatrix} \lambda^{-1} & 0 \\ 0 & B \end{pmatrix}^{-1}$. We claim that $\bar{\rho}_A$ sends this to (λ, B). Indeed, $\rho_A(g) \in \mathrm{GL}(A(\mathbb{C}^M))$ is the linear transformation

$$\begin{pmatrix} 0 & -Z^T \\ Z & 0 \end{pmatrix} \mapsto \begin{pmatrix} \lambda & 0 \\ 0 & B \end{pmatrix}\begin{pmatrix} 0 & -Z^T \\ Z & 0 \end{pmatrix}\begin{pmatrix} \lambda^{-1} & 0 \\ 0 & B \end{pmatrix}^{-1} = \begin{pmatrix} 0 & -(\lambda BZ)^T \\ \lambda BZ & I \end{pmatrix},$$

and $\rho_A(g) = A^{-1} \circ \rho_A(g) \circ A \in \mathrm{GL}(V)$ is thus the map $Z \mapsto \lambda BZ$, i.e. the symmetry of Q given by (λ, B). So ρ_A is indeed surjective.

We now append the condition of respecting symmetries to the definition of determinantal complexity.

Definition 4.3.4. The *equivariant determinantal complexity* of a polynomial $P \in S^m V^*$, denoted $\mathrm{edc}(P)$, is the smallest n for which there is a determinantal representation $\tilde{A} : V \to \mathbb{C}^{n^2}$ of P that respects the symmetries of P.

4.3.2 Preliminaries

Whereas the best lower bound known for the determinantal complexity of perm_n is merely quadratic, Landsberg and Ressayre obtain an exponential lower bound for the equivariant determinantal complexity of perm_n.

First, we need some preliminaries. Let $\Lambda_{n-1} \in \mathcal{M}_n(\mathbb{C})$ denote the matrix with 1's in the last $n-1$ diagonal entries and 0's elsewhere. Denote the image and kernel of Λ_{n-1} by $\mathbb{H} \subset \mathbb{C}^n$ and $\ell_1 \in \mathbb{C}^n$ respectively. To distinguish the domain and target of matrices in $\mathcal{M}_n(\mathbb{C})$, let ℓ_2 be a copy of ℓ_1 in the target so that $\mathcal{M}_n(\mathbb{C})$ can be parametrized as

$$\mathcal{M}_n(\mathbb{C}) = \begin{pmatrix} \ell_1^* \otimes \ell_2 & \mathbb{H}^* \otimes \ell_2 \\ \ell_1^* \otimes \mathbb{H} & \mathbb{H}^* \otimes \mathbb{H} \end{pmatrix}.$$

We additionally need the following basic results in the theory of complex algebraic groups. Recall that a complex algebraic group is *unipotent* if there is a subgroup H of the group U_n of upper triangular matrices with 1's on the diagonal for which $G \simeq H$.

Lemma 4.3.5. *For any complex algebraic group G, there exists a* unipotent radical $R^u(G)$*, i.e. a maximal normal unipotent subgroup, and a* Levi factor $L \subset G$ *for which $G = R^u(G)L$. Moreover, L and $G/R^u(G)$ are reductive.*

Theorem 4.3.6 (Malcev's Theorem, [47], Theorem 5, Chapter 6)**.** *For any Levi factor $L \subset G$ and reductive $H \subset G$, there exists $g \in R^u(G)$ for which $gHg^{-1} \subseteq L$.*

We will also use the following result on the rank of Λ in determinantal representations.

Lemma 4.3.7 ([63])**.** *Let $V(P)_{sing} \subset V$ denote the singular locus of the hypersurface cut out by $P \in S^m V^*$. If* $\operatorname{codim}(V(P)_{sing}) \geq 5$ *and $\tilde{A} = \Lambda + A$ is a determinantal representation of P, then* $\operatorname{rank}(\Lambda) = n - 1$.

Proof. By the chain rule, if $\tilde{A}(v) \in V(\det_n)_{sing}$, then $v \in V(P)_{sing}$. Note that $V(\det_n)_{sing}$ consists of matrices of rank at most $n-2$, which have codimension 4 in $\mathcal{M}_n(\mathbb{C})$. So the set of v for which $\operatorname{rank}(\tilde{A}(v)) \leq n-2$ is either empty or of codimension is at most 4.

Obviously if this set is empty, we're done: $\operatorname{rank}(\tilde{A}(v)) \geq n-1$, so taking $v = 0$ gives the desired claim. If instead this set is of codimension at most 4, then by our hypothesis that $\operatorname{codim}(V(P)_{sing}) \geq 5$, the claim still follows. □

For this reason, we may assume, after transforming Λ by symmetries of $\det_n$, that $\Lambda = \Lambda_{n-1}$. With this in mind, we define one more symmetry group:

$$\mathbb{G}_{\det_n, \Lambda_{n-1}} = \{g \in \mathbb{G}_{\det_n} \mid g \cdot \Lambda_{n-1} = \Lambda_{n-1}\}.$$

Lemma 4.3.8. $\mathbb{G}_{det_n,\Lambda_{n-1}} \simeq [GL(\ell_2) \times GL(\mathbb{H}) \ltimes (\mathbb{H} \oplus \mathbb{H}^* \otimes \ell_2)] \ltimes \mathbb{Z}_2$*. Specifically, every element in this symmetry group is of the form*

$$M \mapsto \begin{pmatrix} \lambda & 0 \\ v & g \end{pmatrix} M \begin{pmatrix} 1 & \phi \\ 0 & g \end{pmatrix}^{-1}$$

(possibly post-composed with a transpose) for some $g \in GL(\mathbb{H})$, $\lambda \in GL(\ell_2) = \mathbb{C}^*$, $v \in \mathbb{H}$, $\phi \in \mathbb{H}^*$.

Proof. Pick some element of $\mathbb{G}^o_{\det_n}$ and denote it by $M \mapsto AMB^{-1}$; because it fixes Λ_{n-1}, A and B stabilize $\mathbb{H}$ and ℓ_2 respectively and are thus of the form

$$A = \begin{pmatrix} \lambda & 0 \\ v & g \end{pmatrix}, \quad B = \begin{pmatrix} \lambda' & \phi \\ 0 & g' \end{pmatrix}$$

In fact, $A\Lambda_{n-1}B = \Lambda_{n-1}$ also forces $g = g'$, so the claim follows. □

This in fact yields a Levi decomposition $\mathbb{G}_{det_n,\Lambda_{n-1}}$ as

$$L = (\mathrm{GL}(\ell_2) \times \mathrm{GL}(\mathbb{H})) \ltimes \mathbb{Z}_2, \qquad R^u(\mathbb{G}_{\det_n,\Lambda_{n-1}}) = (\mathbb{H} \oplus \mathbb{H}^* \otimes \ell_2) \tag{4.4}$$

4.3.3 Exponential lower bound on Q

We work through a toy example to illustrate the general argument.

Theorem 4.3.9. *Let Q be the nondegenerate quadric defined in Example 4.3.3. $edc(Q) \geq M + 1$.*

Proof. Roughly, we will obtain a reductive subgroup L of $\mathbb{G}_{\det_n,\Lambda_{n-1}}$ mapping onto $\mathbb{G}_Q$, after which we can use the fact that $A : V \to A(V)$ is an L^o-equivariant embedding to argue that $\mathbb{H}$ must contain at least one copy of V, giving our lower bound.

Specifically, take a Levi decomposition of $\bar{\rho}_A^{-1}(\mathbb{G}_Q) = R^u(H)L$. $\bar{\rho}_A(R^u(H))$ is normal unipotent inside $\mathbb{G}_Q$, but because $\mathbb{G}_Q \simeq \mathbb{C}^* \times O(M)$ is reductive and thus has trivial unipotent radical, $\bar{\rho}_A(R^u(H))$ is trivial. But $\bar{\rho}_A$ is still surjective, so $\bar{\rho}_A(L) = \mathbb{G}_Q$. In fact, we even have that $\bar{\rho}_A(L^o) = \mathbb{G}_Q$ because $\mathbb{G}_Q$ is connected.

By Theorem 4.3.6, we may assume without loss of generality that L lies in the Levi factor in Lemma so that $L^o \subset \mathrm{GL}(\ell_2)\times\mathrm{GL}(\mathbb{H})$. $A : V \to A(V)$ is an L_0-equivariant embedding of V into irreducible L^o-submodule $A(V) \subset \mathcal{M}_n(\mathbb{C})$, so the component of $A(V)$ in $\ell_1^* \otimes \ell_2$ must be zero. The component in $\ell_1^* \otimes \mathbb{H}$ cannot also be zero, or else $Q = \det_n \circ \tilde{A}$ is zero, so $\ell_1^* \otimes \mathbb{H}$ must contain a copy of V. We conclude that $\dim(\mathbb{H}) \geq \dim(V)$, yielding the lower bound of $\mathrm{edc}(Q) \geq M + 1$. □

4.3.4 Exponential lower bound on perm_n

Landsberg and Ressayre [37] showed that in fact one can get an exponential lower bound on determinantal complexity of the permanent provided the determinantal representation respects just half of the permanent's symmetries.

First, recall from Theorem 4.1.15 that if $\mathbb{C}^{m^2} = E \otimes F$, then

$$\mathbb{G}_{\mathrm{perm}_m} \simeq ((N(T_E) \times N(T_F))/\mathbb{C}^*) \ltimes \mathbb{Z}_2.$$

Henceforth, denote T_E and $N(T_E) = T_E \ltimes \mathcal{S}_m$ by T and N respectively. For $A, B \in T$ and $M_\sigma, M_\tau \in \mathcal{S}_m$, $(AM_\sigma, BM_\tau) \in \left(\left(N(T^{\mathrm{GL}(E)}) \times N(T^{\mathrm{GL}(F)}\right)/\mathbb{C}^*\right)$ corresponds to the action $M \mapsto (AM_\sigma)\cdot M \cdot (BM_\tau)^{-1}$, and $\mathbb{Z}_2$ corresponds to the transpose as usual.

Theorem 4.3.10. *If $m \geq 3$ and $\tilde{A}_m : \mathcal{M}_m(\mathbb{C}) \to \mathcal{M}_n(\mathbb{C})$ is a determinantal representation of perm_m that respects N, then $n \geq 2^m - 1$.*

Proof. Analogous to the first half of the argument for Theorem 4.3.9, we could try lifting the symmetries we're preserving, in this case N, to some reductive subgroup $\mathbb{G} \subset \mathbb{G}_A$ for which the restriction of $\bar{\rho}_A$ to $\mathbb{G}$ is finite and surjects onto N. We could try using the trick above of taking a Levi decomposition of $\bar{\rho}_A^{-1}(N)$ and conjugating using Theorem 4.3.6 to produce a reductive subgroup of $(\mathrm{GL}(\ell_2)\times\mathrm{GL}(\mathbb{H}))\ltimes\mathbb{Z}_2$ mapping onto N. The first obstacle we encounter is that unlike in the case of Q, $\mathbb{G}_{\mathrm{perm}_n}$ is not connected, so we can't just take our lift $\mathbb{G}$ to be L^o.

We can remedy this by passing to an index 2 subgroup of L. The quotient of $\mathbb{G}_{\det_n,\Lambda_{n-1}}$ by its identity component is $\mathbb{Z}_2$. Define L' to be $L \cap \mathbb{G}^o_{\det_n,\Lambda_{n-1}}$ so that L/L' embeds in $\mathbb{Z}_2$, implying L' is either all of L or an index 2 subgroup of L. Recalling that $N \simeq T \ltimes \mathcal{S}_m$ and noting that the alternating group $\mathcal{A}_m$ is the unique index 2 subgroup of $\mathcal{S}_m$, we conclude that $\bar{\rho}_A(L') \supseteq T \ltimes \mathcal{A}_m$; replacing L' by a subgroup if necessary, we may assume that $\bar{\rho}_A(L') = T \ltimes \mathcal{A}_m$.

Unfortunately, this is not enough for us to mimic the proof of Theorem 4.3.9, because in fact the lift $\mathbb{G}$ we are looking for cannot exist: there is no finite universal cover of $\mathcal{S}_m$.

For the time being, let us ignore this fact and proceed to examine the decomposition of $\mathcal{M}_n(\mathbb{C})$ as $(\ell_1 \oplus \mathbb{H})^* \otimes (\ell_2 \oplus \mathbb{H})$ as before. Denote by E the standard representation of L' via $\bar{\rho}_A$; $A(V)$ is an L'-module isomorphic to $E^{\oplus m}$. As before, that A equivariantly embeds V into $\mathcal{M}_n(\mathbb{C})$ forces the $\ell_1^* \otimes \ell_2$ component of $A(V)$ to be zero, while the nonvanishing of the determinant of the matrix of indeterminates given by $A(V)$ implies the $\ell_1^* \otimes \mathbb{H}$ component must be nonzero and contain an $\mathbb{H}_1 \simeq E$.

We can quickly rule out the case that $\mathbb{H}_1^* \otimes \ell_2 \simeq E$, which only occurs when $m = 2$.

In the other case, pick a complement $\mathbb{S}_1$ of $\mathbb{H}_1$ in $\mathbb{H}$ which is stable under the action of L'. We can assume the $\mathbb{H}_1^* \otimes \mathbb{S}_1$ component of $A(V)$ does not vanish, because otherwise the determinant of the $(\ell_1 \oplus \mathbb{S}_1)^* \otimes (\ell_2 \otimes \mathbb{S}_1)$ component of $A(V)$ is the same as that of the $(\ell_1 \oplus \mathbb{H})^* \otimes (\ell_2 \otimes \mathbb{H})$ component. So inside $\mathbb{S}_1$ there is some irreducible L'-module $\mathbb{H}_2$ for which $H_1^* \otimes H_2$ contains a copy of E. Proceeding in this fashion, we obtain a sequence of irreducible L'-submodules $\mathbb{H}_1, ..., \mathbb{H}_k$ which terminates at $k \geq 2$ for which $E \simeq \mathbb{H}_k^* \otimes \ell_2$.

It would be convenient if the list of blocks $\ell_1^* \otimes \mathbb{H}_1, \mathbb{H}_1^* \otimes \mathbb{H}_{i+1}$ $(1 \leq i \leq k)$, and $\mathbb{H}_k^* \otimes \ell_2$ that we've found so far containing copies of E is the complete list of nonzero blocks in $A(V)$. We will show this is indeed true, but to do so and also better understand the L'-modules $\mathbb{H}_i$, we need to return to the question of lifting L' further.

While no finite universal cover of the $\mathcal{S}_m$ component of N exists, we can nevertheless construct a finite cover of the T component of N.

We know $\bar{\rho}_A$ maps L'^o onto T, so in fact the identity component of the center of L'^o, does as well. Denoting this by Z^o, we get an injection of the group of characters $X(T)$ of T into $X(Z^o)$. On the one hand, we have a map $\pi_A : L'/L'^o \to \mathcal{A}_m$ given by quotienting out L'^o and its image in $\bar{\rho}_A$, and on the other, we have an action of L'/L'^o on Z^o just given by conjugation, again quotiented out by L'^o. These actions satisfy

$$\bar{\rho}_A(g \cdot z) = \pi_A(g) \cdot \bar{\rho}_A(z)$$

for any $g \in L'/L'^o$, implying that $X(T)$ and $\ker(\bar{\rho}_A) \cap Z^o$ are stable under the respective actions of L'/L'^o.

We can now produce a subtorus $\tilde{T} \subset Z^o$ that is a finite cover of T via $\bar{\rho}_A$. Define $\Gamma_{\mathbb{Q}}$ to be the complement of $X(T) \otimes \mathbb{Q}$ inside $X(Z^o) \otimes \mathbb{Q}$ that is stable under the action of L'/L^o, and define $\Gamma = \Gamma_{\mathbb{Q}} \cap X(Z_0)$. We take our lift of T to be $\tilde{T} := \{t \in Z^o : \chi(t) = 1 \ \forall \chi \in \Gamma\}$, and because $\bar{\rho}_A$ restricted to $\tilde{T}$ is a finite morphism onto T, we can identify $X(T)$ with an index-k_0 subgroup of $X(\tilde{T})$ for some finite k_0, i.e. we get embeddings

$$k_0 X(\tilde{T}) \subset X(T) \subset X(\tilde{T}).$$

Equivariance and $\tilde{T}$ now give a straightforward proof that $\ell_1^* \otimes \mathbb{H}_1, \mathbb{H}_1^* \otimes \mathbb{H}_{i+1}$ $(1 \le i \le k)$, and $\mathbb{H}_k^* \otimes \ell_2$ are the only nonzero blocks of $A(V)$. Indeed, we have a one-parameter subgroup $\gamma : \mathbb{C}^* \to \tilde{T}$ given by $\bar{\rho}_A \circ \gamma(t) = t^{k_0} \operatorname{Id}_E$ which acts with weight k_0 on E and acts trivially elsewhere. It remains to bound the dimensions of the $\mathbb{H}_i$; we do this by studying the their weights as $\tilde{T}$-modules. More generally, consider any irreducible L'-module W and take its isotypic decomposition as a $\tilde{T}$-module:

$$W = \bigoplus_{\chi \in \mathrm{Wt}(\tilde{T}, W)} W_\chi.$$

Recall that L' acts on $\tilde{T}$ by conjugation and thus also on $X(\tilde{T})$. By definition, L'^o acts trivially on the torus Z^o, and because $\tilde{T}$ is a central subtorus of L'^o, by rigidity of tori we conclude that L'^o also acts trivially on $\tilde{T}$ and thus on $X(\tilde{T})$.

So now we have an (L'/L'^o)-action on $X(\tilde{T})$, under which $\mathrm{Wt}(\tilde{T}, W)$ is stable. But for any $\chi \in \mathrm{Wt}(\tilde{T}, W)$, $\bigoplus_{\sigma \in L'/L'^o} W_{\sigma \cdot \chi}$ is L'-stable and thus must be all of W by irreducibility of W. We conclude that $\mathrm{Wt}(\tilde{T}, W) = (L'/L'^o) \cdot \chi_W$ and thus that $k_0 \mathrm{Wt}(\tilde{T}, W) = \mathcal{A}_m \cdot \chi_W$.

From the set of all such χ_W, we pick a distinguished weight as follows. Define $\epsilon_i \in \chi(T)$ to send an element (t_k^j) to its i-th diagonal entry t_i^i. The collection of $\mathcal{A}_m$-dominant weights $\{a_1\epsilon_1 + \cdots + a_m\epsilon_m : a_1 \ge \cdots \ge a_{m-1}, a_{m-2} \ge a_m\}$ is a fundamental domain for the action of $\mathcal{A}_m$ on $X(T)$, so take χ_W to be the unique $\mathcal{A}_m$-dominant weight for which $k_0 \mathrm{Wt}(\tilde{T}, W) = \mathcal{A}_m \cdot \chi_W$.

Now apply the above to the case of $W = \mathbb{H}_i$. We find that $\dim(\mathbb{H}_i) \ge |\mathcal{A}_m \cdot \chi_{H_i}|$. In the basis $\{\epsilon_i\}$, we call the number of nonzero coordinates of $\chi \in X(T)$ the *length* $\ell(\chi)$. Then if $\ell = \ell(\chi_{\mathbb{H}_i})$, the $\mathcal{A}_m$-orbit of $\chi_{\mathbb{H}_i}$ is of size $\binom{n}{\ell}$. To obtain the desired exponential lower bound, we show that there are weights $\chi_{\mathbb{H}_i}$ of almost every possible length.

To this end, note that

$$k_0 \mathrm{Wt}(\tilde{T}, \mathbb{H}_i^* \otimes \mathbb{H}_{i+1}) = \{-\sigma_1 \chi_{\mathbb{H}_i} + \sigma_2 \chi_{\mathbb{H}_{i+1}} \mid \sigma_1, \sigma_2 \in \mathcal{A}_m\}.$$

Because $E \subseteq \mathbb{H}_i^* \otimes \mathbb{H}_{i+1}$, $k_0 \mathrm{Wt}(\tilde{T}, E) \subseteq k_0 \mathrm{Wt}(\tilde{T}, \mathbb{H}_i^* \otimes \mathbb{H}_{i+1})$, and the weights of the standard representation are simply the characters ϵ_i, we conclude that $\chi_{\mathbb{H}_{i+1}} = \sigma \cdot \chi_{\mathbb{H}_i} + k_0 \epsilon_i$ for some $\sigma \in \mathcal{A}_m$ and $i \in [m]$. In other words,

$$\ell(\chi_{\mathbb{H}_{i+1}}) \le \ell(\chi_{\mathbb{H}_i}) + 1.$$

Furthermore, $\ell(\chi_{\mathbb{H}_k}) \ge m - 1$ because 1) $\mathbb{H}_k^* \otimes \ell_2 \simeq E$ and 2) χ_{ℓ_2} is $\mathcal{A}_m$-invariant so that χ_{ℓ_2} is some scalar multiple of $\epsilon_1 + \cdots + \epsilon_m$.

In conclusion, we have a subsequence $\mathbb{H}_{j_1}, \ldots, \mathbb{H}_{j_{m-1}}$ for which $\ell(\mathbb{H}_{j_i}) = i$, so

$$n = 1 + \dim \mathbb{H} \ge 1 + \sum_{i=1}^{m-1} \dim \mathbb{H}_{j_i} \ge 1 + \sum_{i=1}^{m-1} \binom{m}{i} = 2^m - 1,$$

as desired. □

Landsberg and Ressayre showed that Grenet's determinantal representation is equivariant, making this exponential lower bound tight.

4.4 Orbit Closures and Obstructions

We sketch here the program put forth by Mulmuley and Sohoni in [44] in anticipation of the material discussed in the next chapter.

4.4.1 Orbit closures

The issue with Definition 4.4.1 is that the relevant geometric objects turn out not to be cut out by polynomial equations and are thus not amenable to study via the techniques of algebraic geometry. Motivated by this, Mulmuley and Sohoni considered a stronger version of Valiant's conjecture involving the Zariski closures of the orbit mentioned in Section 4.1.2.

Define the *projective orbit closures*

$$Det_n := \overline{\mathrm{GL}(W)\cdot[\det_n]}, \qquad Perm_n^m := \overline{\mathrm{GL}(W)\cdot([\ell^{n-m}\,\mathrm{perm}_m])}$$

and modify Definition 4.4.1 as follows.

Definition 4.4.1. Let $V \subset W$ and $P \in S^m(V^*)$, and take ℓ a linear coordinate on $\mathbb{C}^1$ with a linear inclusion $\mathbb{C}^1 \oplus V \hookrightarrow W$ so that $\ell^{n-m}\mathbb{P} \in S^n W^*$. The *border determinantal complexity* $\overline{\mathrm{dc}}(P)$ is the smallest n for which

$$\overline{\mathrm{GL}(W)\cdot[\ell^{n-m}P]} \subset Det_n.$$

Conjecture 4.4.2 ([44])**.** $\overline{\mathrm{dc}}(\mathrm{perm}_m)$ grows faster than any polynomial in m.

Mulmuley and Sohoni's approach, roughly speaking, is to study the coordinate rings of Det_n and $Perm_n^m$ and look for $\mathrm{GL}(W)$-modules that for $m = \mathsf{poly}(n)$ appear in the former but not the latter. We will make this notion more precise in subsection 4.4.3, but we first need to set up some preliminaries in classical representation theory.

4.4.2 Coordinate rings of orbits

Throughout this subsection, let G be a reductive group. The algebraic Peter-Weyl theorem gives a direct-sum decomposition of $\mathbb{C}[G]$ into irreducible $(G \times G)$-modules indexed by all irreducible G-modules:

Theorem 4.4.3 (Algebraic Peter-Weyl)**.** *If G is a reductive algebraic group, the $G \times G$-module $\mathbb{C}[G]$ decomposes as*

$$\mathbb{C}[G] = \bigoplus_\lambda V_\lambda \otimes V_\lambda^*,$$

where the direct sum is over all isomorphism classes of irreducible G-modules V_λ.

We prove this in Appendix A.3.

Example 4.4.4. From Theorem 4.4.3, we get a decomposition of the coordinate ring of $\mathcal{S}_d$. It is an elementary fact in representation theory that the discrete Fourier transform gives a one-to-one correspondence between irreducible representations of a finite group G and conjugacy classes of G. The conjugacy classes of $\mathcal{S}_d$ correspond to the distinct partitions of d, i.e. tuples $(p_1, ..., p_r) \in \mathbb{N}^r$ for which $p_1 + \cdots + p_r = d$ and $p_1 \geq \cdots \geq p_r$. We conclude that $\mathbb{C}[\mathcal{S}_d] = \bigoplus_\pi [\pi]^* \otimes [\pi]$. But $[\pi]^* \simeq [\pi]$ because the dual of an irreducible representation ρ has corresponding character equal to the conjugate of that of ρ, so in fact

$$\mathbb{C}[\mathcal{S}_d] = \bigoplus_\pi [\pi] \otimes [\pi] \tag{4.5}$$

as an $\mathcal{S}_d \times \mathcal{S}_d$-module, where π ranges over all partitions of d.

Noting that the ring of regular functions on the quotient G/H can be identified with the ring $\mathbb{C}[G]^H$ of regular functions on G which fix H, we also get the following corollary to Theorem 4.4.3.

Corollary 4.4.5. *If $H \subset G$ is a closed subgroup, then as a G-module, $\mathbb{C}[G/H]$ decomposes as*

$$\mathbb{C}[G/H] = \bigoplus_{\lambda} V_{\lambda}^{\oplus \dim(V_{\lambda}^{*})^{H}}, \tag{4.6}$$

where the sum is taken over the irreducible G-modules.

Definition 4.4.6. Let G act on some set X. The *stabilizer* of $P \in X$ in G is defined by

$$G_P := \{g \in G \mid g \cdot P = P\}. \tag{4.7}$$

In particular, if G acts on some vector space V, then for a given $p \in V$ we can apply Corollary 4.4.5 to G_p to get an isotypic decomposition of the coordinate ring $G \cdot p$, namely

$$\mathbb{C}[G \cdot p] = \bigoplus_{\lambda} V_{\lambda}^{\oplus \dim(V_{\lambda}^{*})^{G_p}}. \tag{4.8}$$

In the setting of GCT, for $W = \mathbb{C}^{n^2}$, take $G = \mathrm{GL}(W)$, $V = \mathbb{P}(S^n W^*)$, and $p = [\det_n]$ or $[\ell^{n-m} \operatorname{perm}_m]$. As we saw in Lemma 4.1.18, $\det_n$ and perm_m are characterized by their symmetries, so by (4.8), the orbit of any polynomial p that is characterized by its symmetries is the unique one with its coordinate ring.

Remark 4.4.7. It is believed that the characterization of $\det_n$ and perm_m by their symmetries should play a big role in the resolution of the Mulmuley/Sohoni conjecture, and because this property does not hold for general polynomials in $S^n W^*$, this should allow the GCT program to evade Razborov/Rudich's natural proofs barrier [46] in a way that other attempts to separate P vs. NP have not been able to.

4.4.3 Representation-theoretic obstructions

One way to separate Det_n and $Perm_n^m$ is to look for modules in $\mathbb{C}[Perm_n^m]$ that do not occur in $\mathbb{C}[Det_n]$.

We would like first to find explicit modules in $I(Det_n)$ with the ultimate hope that $\ell^{n-m} \operatorname{perm}_m$ does not vanish on them. We may decompose $\mathrm{Sym}(V)$ into $I(Det_n) \oplus \mathbb{C}[Det_n]$ so that any module lying in $\mathrm{Sym}(V)$ but not in $\mathbb{C}[Det_n]$ must lie in $I(Det_n)$.

This leads us to the first kind of obstruction to look for.

Definition 4.4.8. An irreducible G-module V_λ is an *orbit occurrence obstruction* if V_λ occurs in $\mathrm{Sym}(V)$ but not in the coordinate ring of the orbit, i.e. $V_{\lambda}^{*G_{\det_n}} = 0$.

The advantage of such obstructions is that we already have a nice characterization of the coordinate ring of $G \cdot [\det_n]$ by Lemma 4.4.5, but the drawback is that potentially useful orbit occurrence obstructions are very difficult to find. One way to relax this is to look for modules that occur with differing multiplicities in the coordinate rings being compared. In the above discussion, to find an irreducible G-module appearing in $I(Det_n)$, it is enough to find an irreducible G-module of higher multiplicity inside $\mathrm{Sym}(V)$ than in $\mathbb{C}[G/G_{\det_n}]$.

Definition 4.4.9. An irreducible G-module V_λ is an *orbit representation-theoretic obstruction* if the multiplicity of V_λ in $\mathrm{Sym}(V)$ exceeds that of V_λ in the coordinate ring of the orbit, which is equal to $\dim(V_\lambda)^{*G_{\det_n}}$.

The issue is that to test $\ell^{n-m} \operatorname{perm}_m$ on any such obstruction $S_\pi(W)$, we would need an explicit realization of $S_\pi(W)$ in $I(Det_n)$.

In any case, we can relax our search for modules further by asking only for those appearing in $\mathrm{Sym}(V)$ with higher multiplicity than in $\mathbb{C}[Det_n]$.

Definition 4.4.10. An irreducible G-module V_λ is an *occurrence obstruction* if V_λ occurs in $\mathrm{Sym}(V)$ but not in $\mathbb{C}[Det_n]$

More generally, V_λ is a *representation-theoretic obstruction* if the multiplicity of V_λ in $\mathrm{Sym}(V)$ exceeds that of V_λ in the coordinate ring of $\mathbb{C}[Det_n]$.

That said, (non-orbit) occurrence or representation-theoretic obstructions, while more useful, are harder to determine because the general problem of deciding which functions in $\mathbb{C}[G/G_{\det_n}]$ extend to $\mathbb{C}[Det_n]$, the *extension problem*, is especially difficult. More generally, it is still not entirely clear how Mulmuley/Sohoni's conjecture differs from Valiant's in passing to the closure, and we will address this in the next chapter.

4.5 A First Lower Bound with GCT

In this section, we upgrade Mignon-Ressayre's lower bound on $\mathrm{dc}(\mathrm{perm}_m)$ to one on $\overline{\mathrm{dc}}(\mathrm{perm}_m)$.

Theorem 4.5.1 ([36], Theorem 1.0.1). $\overline{dc}(\mathrm{perm}_m) \geq m^2/2$.

The discrepancy in ranks between the second fundamental forms of the determinantal and permanental hypersurfaces still lies at the core of this result, but to account for passing to the Zariski closure, we will interpret this discrepancy as one between dimensions of dual varieties. For a hypersurface $V(P) \subset \mathbb{P}W$ cut out by $P \in S^dW^*$, we define its *dual variety* $V(P)^*$ to be the Zariski closure of the set of all hyperplanes tangent to $V(P)$ at smooth points of $V(P)$. For a vector space W of dimension N, define the variety $Dual_{k,d,N} \subset \mathbb{P}(S^dW^*)$ to be the Zariski closure of the set of degree-d irreducible hypersurfaces in $\mathbb{P}W$ for which $\dim V(P)^* \leq k$.

To show Theorem 4.5.1, we will show that whereas $Det_n \subset Dual_{2n-2,n,n^2}$, $Perm_n^m \notin Dual_{2n-2,n,n^2}$ unless $n \geq m^2/2$.

4.5.1 Hessians and dual varieties

The key connection between Mignon-Ressayre's and Landsberg et al.'s approaches is the following dimension formula due to B. Segre. The proof we give follows [22].

Lemma 4.5.2 (B. Segre). $\dim V(P)^* = \mathrm{rank}(T_x^2P) - 2$ *for x a general point on $V(P)$.*

Proof. The smooth points on $V(P)$ are dense, so we can pick x to be smooth. Denote by $[h]$ the point in $V(P)^* \subset W^*$ corresponding to $T_xV(P)$. It is enough to show that $\dim \hat{T}_xV(P)^* = \mathrm{rank}(T_x^2P) - 1$. Take some curve $h_t \subset V(P)^*$ for which $h_0 = h$. Recall that $T_xV(P)$ is the image of $V(P)$ under the Gauss map $\mathcal{G}(x) := \left(\frac{\partial P}{\partial z_i}(x), ..., \frac{\partial P}{\partial z_n}(x)\right) \in W^*$. Equivalently, we can regard the Gauss map as sending x to the polarization $\partial_{n-1,1}P(x,\cdot)$. We can then write $h_t = \partial_{n-1,1}P(x_t,\cdot)$ for some curve $x_t \subset V(P)$. It follows that $h_0' = \partial_{n-2,1,1}P(x,x_0',\cdot)$. This is nonzero so long as $x_0' \neq \lambda x$, so the dimension of the span of all such velocity vectors h_0' is one less than the rank of the Hessian $\partial_{n-2,2}P(x,\cdot) = T_x^2V(P)$, as desired. ☐

Lemma 4.2.3 thus implies that $\dim V(\det_n)^* = 2n-2$, i.e. that the dual variety of the determinantal hypersurface is highly degenerate. Lemma 4.5.2 implies that for an irreducible $P \in S^dW^*$ to have a dual variety of dimension at most k, $\det(T^2P)$ restricted to any $(k+3)$-dimensional subspace $F \subset W$ must vanish on points of $V(P)$. Equivalently, for any such F, P must divide $Q := \det(T^2P|_F) \in S^eW^*$, where $e := (k+3)(d-2)$.

4.5.2 Conditions for P to divide Q

We wish to find conditions on their coefficients that must hold if P divides Q. In [36], the authors were motivated not just by the problem of lower-bounding $\overline{\mathrm{dc}}(\mathrm{perm}_m)$, but by that of finding defining equations for $Dual_{k,d,N}$. Our focus is only on the former, so for the sake of brevity we can afford to be a bit crude with the equations we obtain, as in Section 6.6.5 of [32]. For P to divide Q, there must exist some $R \in S^{e-d}W^*$ for which $Q = PR$. Picking some basis $\{F_1, ..., F_D\}$ of $S^{e-d}W^*$, where $D = \binom{e-d+n-1}{e-d}$, we note that this R exists iff

$$F_1P \wedge F_2P \wedge \cdots \wedge F_DP \wedge Q = 0.$$

this gives rises to equations in P which cut out some variety, call it $D_{k,d,N} \supset Dual_{k,d,N}$.

4.5.3 Landsberg et al.'s bound

By Lemmas 4.2.3, 4.2.4, and 4.5.2, we know that $Det_n \subset D_{2n-2,n,n^2}$ while $[\mathrm{perm}_m] \notin D_{2n-2,n,n^2}$ for $n < m^2/2$. However, we cannot say anything at this point about $[\ell^{n-m}\,\mathrm{perm}_m]$ because it is not irreducible. The good news is that the (non)degeneracy of the dual variety of an irreducible P implies the (non)degeneracy of the dual variety of any "padded" version of P:

Lemma 4.5.3. *Let* $U = \mathbb{C}^M$ *and* $L = \mathbb{C}$ *and suppose we have a linear inclusion* $U^* \oplus L^* \subset W^*$*. For* $\ell \in L^*$ *nonzero and* $P \in S^m U^*$ *irreducible, if* $[P] \in D_{k,m,M}$ *and* $[P] \notin D_{k-1,m,M}$*, then* $[\ell^{n-m}P] \in D_{k,n,N}$ *and* $[\ell^{n-m}P] \notin D_{k-1,n,N}$*.*

Proof. Picking an appropriate basis for W, we can write $T^2(\ell^{n-m}P)$ in $(M, 1, N-M-1) \times (M, 1, N-M-1)$ block matrix form as

$$T^2(\ell^{n-m}P) = \begin{pmatrix} \ell^{n-m}T^2P & 0 & 0 \\ 0 & (n-m)(n-m-1)\ell^{n-m-2}P & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

By assumption, $\dim V(P)^* = k$. For a general $F \in G(k+2, W)$, $\det(T^2P|_F) \neq 0$ because $[P] \notin D_{k-1,m,M}$, so $\ell^{n-m}P$ does not divide $\det(T^2(\ell^{n-m}P)|_F)$ and thus $[\ell^{n-m}P] \notin D_{k-1,n,N}$. Likewise, for any $F \in G(k+3, W)$, $\det(T^2P|_F) = 0$ because $[P] \in D_{k-1,m,M}$, so $\det(T^2(\ell^{n-m}P)|_F)$ is identically zero or divisible by $\ell^{n-m}P$. We conclude that $[\ell^{n-m}P] \in D_{k,n,N}$. □

By Lemma 4.5.3, $Perm_n^m \notin D_{2n-2,n,n^2}$ for $n < m^2/2$, and Theorem 4.5.1 follows.

Chapter 5

Boundary Components of Det_n

In this chapter, we ask the following fundamental problem in geometry:

Question 5.0.4 (Boundary Component Problem)**.** Classify the irreducible components of ∂Det_n.

As we will see in Section 5.1, because of a negative result of Kumar [31] on the non-normality of Det_n, this question is necessary for better understanding what it takes to find representation-theoretic obstructions. The boundary component problem turns out to be closely related to the classical question of finding maximal linear subspaces on the determinantal hypersurface, so in Section 5.2, we present two theorems of Atkinson [2] and Eisenbud and Harris [19] on what is known about this problem in low dimensions, with a view towards applying these in our own results. The proof we give of these two theorems will follow that of Eisenbud and Harris.

In Section 5.3, we present a recent result of Huttenhain giving a complete classification of boundary components for Det_3 based on resolution of singularities. We then provide our own more elementary, combinatorial proof of their result based on examining sums of polarizations of $\det_3$.

Beyond considering the low-dimensional case, one can also ask for families of boundary components over an infinite set of dimensions. In Section 5.5, we produce a new such family for even n.

5.1 Motivation: the extension problem

If we are to have any hope of finding representation-theoretic obstructions, we need to understand exactly what changes when we pass from $\mathbb{C}[\mathrm{GL}(W)\cdot\det_n]$ to $\mathbb{C}[\overline{\mathrm{GL}(W)\cdot\det_n}]$, i.e. which functions on the $\mathrm{GL}(W)$-orbit extend to the boundary. Indeed, we might initially hope that all such functions extend, by Hartog's phenomenon:

Lemma 5.1.1 (Algebraic Hartog's)**.** *If X is a normal variety and $V\subset X$ is a closed subset for which $\mathrm{codim}(V)\geq 2$, then any function $f: X\backslash V\to\mathbb{C}$ extends uniquely to some $\tilde{f}: X\to\mathbb{C}$, i.e. the restriction map $\Gamma(X,\mathcal{O}_X)\to\Gamma(X\backslash V,\mathcal{O}_X)$ is an isomorphism.*

Proof. Consider the rational map $X\to\mathbb{P}^1$ corresponding to f and denote its graph by $\Gamma\subset X\times\mathbb{P}^1$. Its fiber over ∞ lies inside $(X\backslash V)\times\{\infty\}$ and therefore has codimension at least 2. If the fiber is nonempty, i.e. if $\Gamma\cap(X\times\{\infty\})\neq\varnothing$, then because $X\times\{\infty\}$ is cut out of $X\times\mathbb{P}^1$ locally by a single equation, the fiber would have codimension at most 1, a contradiction. So Γ and $X\times\{\infty\}$ do not intersect, and we conclude that the projection $\Gamma\to X$ is finite. It is also birational, so by normality of X, $\Gamma\to X$ is an isomorphism. We therefore obtain the unique extension $\tilde{f}$ of f as desired. □

Unfortunately, as we shall now see, Kumar [31] showed that Det_n is not normal, so Hartog's does not apply (even if Det_n were normal, we would still need to deal with extending functions defined off a codimension 1 subvariety). Kumar's result necessitates a better understanding of the geometry of the boundary of Det_n.

We begin by fixing some notation. Let E be an n-dimensional vector space and $W = E \otimes E^*$. Let $G = \mathrm{GL}(W)$ and $G' = \mathrm{SL}(W)$, and denote by $\overline{\Omega}$, Ω, and Ω' respectively the varieties $\overline{G \cdot \det_n}$, $G \cdot \det_n$, and $G' \cdot \det_n$ (there is an implicit dependence of these varieties on n).

Theorem 5.1.2. *For $n \geq 3$, $\overline{\Omega}$ is not normal.*

The general outline of Kumar's argument is as follows. First, to show $\overline{\Omega}$ is not normal, it's enough to show that $Z := \overline{\Omega}//G'$ is not normal. We then show that the ideal of the boundary of $\overline{\Omega}$ is the radical of the ideal generated by some nonzero G'-invariant $f_0 \in I$. Now suppose Z is normal. We find in the isotypic decomposition of $\mathbb{C}[\Omega]^{G'}$ a root (in the sense of radicals) of f_0. Normality of Z implies that f_0 lies in $\mathbb{C}[Z]$, and extending this function to all of $S^n(W^*)$, we get a nonzero G'-invariant in $S^d(S^n(W^*))$ of degree d smaller than is possible by a result of Howe [26] (see Theorem 5.1.9 below), a contradiction.

We begin with the following basic fact.

Lemma 5.1.3. *$G' \cdot det_n$ is closed in $S^n W^*$.*

Proof. We will make use of the following result proved in [30].

Claim 5.1.4 ([30], Corollary 5.1)**.** *Let H be a reductive group acting on some vector space V, and let $x \in V$. If H has no nontrivial central one-parameter subgroup and the stabilizer of the line $[x]$ is not contained in any proper parabolic subgroup of H, then $H \cdot x$ is closed.*

We wish to take H in the claim to be G' and x to be $\det_n$. Obviously G' is reductive and has no nontrivial central one-parameter subgroup. For the second hypothesis, we apply Theorem 4.1.14, by which we know the identity component $G^0_{\det_n} \simeq (\mathrm{SL}(E) \times SL(E^*))/\mu_n$, where μ_n is the group of nth roots of unity acting diagonally on $\mathrm{SL}(E) \times \mathrm{SL}(E^*)$. It is evident that $G'_{\det_n}$ contains $G^0_{\det_n}$, and the latter certainly does not stabilize any proper subspace of W. We conclude that $G'_{\det_n}$ is not contained in any proper parabolic subgroup of G', so applying Kempf's result, we see that $G' \cdot \det_n$ is indeed closed. □

Corollary 5.1.5. *An irreducible G'-module M occurs in the isotypic decomposition of $\mathbb{C}[\Omega']$ iff it occurs in that of $\mathbb{C}[\overline{\Omega}]$ as well, i.e. any such M occurs in $\mathbb{C}[\overline{\Omega}]$ iff M contains nonzero G'_{det_n}-invariants.*

Proof. By algebraic Peter-Weyl, it suffices to show that $\mathbb{C}[\overline{\Omega}]_d$ injects into $\mathbb{C}[\Omega']$ for each $d \geq 0$. Indeed, suppose we have some f in the kernel of the restriction map $\phi_d : \mathbb{C}[\overline{\Omega}]_d \to \mathbb{C}[\Omega']$, i.e. f vanishes over Ω'. Because f is homogeneous of degree d, we also know f vanishes over $\mathbb{C} \cdot \Omega'$ and thus over $\overline{\mathbb{C} \cdot \Omega'} = \overline{\Omega}$. But for f to vanish over all of $\overline{\Omega}$, it would have to be the zero function. □

Lemma 5.1.3 will now allow us to give the key step of showing the existence of the G'-invariant f_0. Let $I \subset \mathbb{C}[\overline{\Omega}]$ denote the ideal of $\partial\overline{\Omega}$.

Lemma 5.1.6. *I contains a nonzero G'-invariant f_0.*

Note that if we can show that the closure of any G'-orbit Y in $\partial\overline{\Omega}\backslash\{0\}$ contains 0, then we are done. This would imply that $\partial\overline{\Omega}//G' \simeq \{0\}$, in which case we could take f_0 to be any nonzero homogeneous polynomial in $\mathbb{C}[Z]$, and it obviously vanishes at $\{0\}$, implying that $f_0 \in I$.

To prove this claim about G'-orbits, we will prove that every closed G'-orbit $Y' \subset \overline{Y}$ which avoids zero is the G'-orbit of a scalar multiple of $\det_n$. Then $Y' = G' \cdot (\lambda \cdot \det_n) \subset G \cdot \det_n = \Omega$ for some $\lambda \in \mathbb{C}^*$, contradicting the fact that $Y' \subset Y$ lies on the boundary of $\overline{\Omega}$.

Define the $\mathbb{C}^*$-equivariant map $\sigma : \mathbb{C} \to \overline{\Omega}$ by $\lambda \mapsto \lambda^n \det_n$, where $\mathbb{C}^*$ acts on $\mathbb{C}$ by multiplication and on $\overline{\Omega}$ via the action induced by the action on $S^n W^*$, namely $\lambda \cdot p = (\lambda^{-1} I_{n^2}) \cdot p$. Post-compose σ with the GIT quotient map $\pi : \overline{\Omega} \to Z$ to get a map $\overline{\sigma} : \mathbb{C} \to Z$. $\overline{\sigma}$ is dominant because Ω is obviously dense in $\overline{\Omega}$. We claim moreover that $\overline{\sigma}$ pulls $\{0\} \in Z$ back to $\{0\} \in \mathbb{C}$. But this follows from Lemma 5.1.3.

Next, we can reformulate the conditions that $\overline{\sigma}$ respectively is dominant and pulls $\{0\}$ back to $\{0\}$ algebraically as conditions that the corresponding map on coordinate rings is injective and pulls the augmentation ideal of $\mathbb{C}[Z]$ back to that of $\mathbb{C}[\mathbb{C}]$. We are done given the following result from commutative algebra:

Lemma 5.1.7. *Let R, S be nonnegatively graded, finitely generated integral domains over $\mathbb{C}$ with zeroeth component $\mathbb{C}$ and augmentation ideals m_R, m_S respectively. If graded algebra homomorphism $f : R \to S$ is injective, and the corresponding map $f^* : \operatorname{Spec} S \to \operatorname{Spec} R$ satisfies $(f^*)^{-1}(m_R) = \{m_S\}$, then S is integral over R.*

Proof. The hypothesis on f^* implies that m_S is the only maximal ideal in S containing $f(m_R)$, so in particular, $\sqrt{f(m_R)} = m_S$ and thus $m_S^d \subset f(m_R)$ for some $d > 0$. Thus, $S/f(m_R)$ is a finite-dimensional $\mathbb{C}$-vector space, implying S is a finitely generated R-module. It is a basic fact of commutative algebra that this implies integrality of S over R. ☐

Proof of Lemma 5.1.6. Because the map on coordinate rings corresponding to $\bar{\sigma}$ is integral, $\bar{\sigma}$ is finite and thus surjective. ☐

5.1.1 Examining the isotypic decomposition of $\mathbb{C}[\Omega]^{G'}$

In this section, we finish off the proof of Theorem 5.1.2. First, note that in the argument for Lemma 5.1.6, any member $f_0 \in \mathbb{C}[\overline{\Omega}]^{G'}$ suffices. We will show that any such f_0 is set-theoretically a defining equation for all of $\partial\overline{\Omega}$, i.e. $V(f_0) = \partial\overline{\Omega}$. Indeed, more generally:

Lemma 5.1.8. *If M is a nonzero G-submodule of I, then $V(M) = \partial\overline{\Omega}$.*

Proof. By hypothesis, $\partial\overline{\Omega}$ is contained in $V(M)$. Suppose to the contrary that the containment is proper. Then because M is a G-submodule, $V(M)$ is G-stable and therefore must be all of $\overline{\Omega}$, meaning M is the zero module, a contradiction. ☐

Taking $M = \langle f_0 \rangle$ the ideal generated by f_0, we conclude that for any $f_0 \in \mathbb{C}[\overline{\Omega}]^{G'}$,

$$\sqrt{\langle f_0 \rangle} = I. \tag{5.1}$$

We are now ready to prove Theorem 5.1.2.

Proof of Theorem 5.1.2. It suffices to show that $Z = \overline{\Omega}//G'$ is not normal. Suppose to the contrary.

G is reductive, and $G_{\det_n}$ is also reductive by Theorem 4.1.14, so by Matsushima's theorem (see Theorem A.5.5 in Appendix A.5) Ω is affine, and we may invoke algebraic Peter-Weyl to obtain the isotypic decomposition

$$\mathbb{C}[\Omega]^{G'} \simeq \bigoplus_{a \in \mathbb{Z}} S_{a^{n^2}}(E) \otimes \left(S_{a^{n^2}}(E)^*\right)^{G_{\det_n}}.$$

The Weyl modules $S_{a^{n^2}}$ appearing in the decompostion of $\mathbb{C}[\Omega]^{G'}$ correspond to the characters $g \mapsto (\det g)^a$. So by Theorem 4.1.14, if $n(n-1)/2$ is even, all maps in $S^*_{a^{n^2}}(E)$ are fixed by members of $G_{\det_n}$ because these all have determinant 1; we conclude that $\dim(S^*_{a^{n^2}}(E))^{G_{\det_n}} = 1$. If $n(n-1)/2$ is odd but a is even, this remains true, but if a is odd in this case, no maps in $S^*_{a^{n^2}}(E)$ are fixed, and $\dim(S^*_{a^{n^2}}(E))^{G_{\det_n}} = 0$.

Endow $\mathbb{C}[\Omega]^{G'}$ with the natural grading d based on the power z^{md} by which the matrix $z \cdot I_{n^2}$ acts on different members of $\mathbb{C}[\Omega]^{G'}$. Then we can pick a generator $\hat{f}_0$ in $\mathbb{C}[\Omega]^{G'}_{p_n n}$, where $p_n = 1$ if $n(n-1)/2$ is even and $p_n = 2$ if $n(n-1)/2$ is odd, so that

$$\mathbb{C}[\Omega]^{G'}_{\geq 0} \simeq \bigoplus_{a > 0} \mathbb{C} \cdot \hat{f}_0^a.$$

We can now invoke (5.1): pick any $f_0 \in \mathbb{C}[\overline{\Omega}]^{G'}$, and take the smallest $d > 0$ for which $\hat{f}_0^d = f_0$.

Here is where assumption of the normality of Z breaks down: it implies that $\hat{f}_0$ is actually in $\mathbb{C}[\overline{\Omega}]^{G'}_{p_n n}$ and thus pulls back via the surjective restriction map $\mathbb{C}[S^n(W^*)] \to \mathbb{C}[\overline{\Omega}]$ to a nonzero element of $\mathbb{C}[S^n(W^*)]^{G'}_{p_n n}$. This gives a nonzero polynomial in the plethysm $S^{p_n n}(S^n(W^*))^{G'}$. Finally, we invoke the following result from [26]:

Theorem 5.1.9 ([26], Section 4.3)**.** *$S^\ell(S^p(\mathbb{C}^n))^{G'} = 0$ if $\ell < n$.*

Because we assume $n \geq 3$, $p_n n < n^2$, so we get a contradiction. ☐

5.2 Linear subspaces of $V(\det_n)$

Before we examine the problem of classifying boundary components more closely, we will take a detour by looking at a related classical problem in algebraic geometry. Given a linear subspace of matrices $M \subset \mathrm{Mat}_{m,n}(\mathbb{C})$, define its *rank* to be the highest rank of any member of M; denote this by $\mathrm{rank}(M)$.

Question 5.2.1. For a given n, what are the maximal linear spaces $M \subset \mathrm{Mat}_{n,n}(\mathbb{C})$ for which $\mathrm{rank}(M) < n$?

Put differently, this question asks whether we can classify all non-extendable linear spaces on the determinantal hypersurface.

The connection between Question 5.2.1 and Question 5.0.4 is as follows. Denote the space of $n \times n$ matrices by W. Define the $\mathbb{G}_{\det_n}$-invariant rational map $\phi : \mathbb{P}\mathrm{End}(W) \dashrightarrow Det_n$ sending $[a] \in \mathbb{P}\mathrm{End}(W)$ to $[\det_n \circ a]$. If one could resolve the indeterminacies of ϕ, then the image under the resolved map would be Det_n. The connection to finding maximal linear subspaces of $V(\det_n)$ is that the indeterminacy locus of ϕ is the set of all $[a]$ for which $a(W)$ consists solely of singular matrices, and such an $a(W)$ would be a linear subspace of $V(\det_n)$. So while Question 5.2.1 is beautiful in its own right, it is also essential to solving Question 5.0.4, and in Sections 5.3 and 5.4, we will see this carried out in low dimensions.

In any case, this section will focus solely on solving Question 5.2.1 in low dimensions.

5.2.1 First examples

Throughout, we will alternately regard linear spaces of $n \times n$ matrices as sitting inside $\mathcal{M}_{n,n}(\mathbb{C})$, or as sitting inside $\mathrm{Hom}(V, W) = V^* \otimes W$ for n-dimensional vector spaces V, W. We will also often denote spaces of matrices using matrices of indeterminates. For instance, we can denote the space of all 3×3 skew-symmetric matrices by the parametrization

$$W_{skew} = \begin{pmatrix} 0 & a & b \\ -a & 0 & c \\ -b & -c & 0 \end{pmatrix}.$$

We can describe W_{skew} in the following coordinate-free manner: given an arbitrary vector space V, we have a map $V \hookrightarrow \mathrm{Hom}(V, \Lambda^2 V)$ sending any $v \in V$ to the multiplication map in the exterior algebra, namely the map $\cdot \wedge v : V \to \Lambda^2 V$. When $\dim(V) = 3$, we recover W_{skew} as the image of V under this identification. When $\dim(V) = 4$, the image of V is the space of matrices

$$\begin{pmatrix} b & c & d & 0 & 0 & 0 \\ -a & 0 & 0 & c & d & 0 \\ 0 & -a & 0 & -b & 0 & d \\ 0 & 0 & -a & 0 & -b & -c \end{pmatrix}. \tag{5.2}$$

As we shall see, this matrix will be of particular significance to answering Question 5.2.1.

Other obvious spaces of singular $n \times n$ matrices come to mind, like those with at least one row/column of zeros or more generally a block of zeros of height m and width at least $n - m + 1$. These form another important class of matrix spaces.

Definition 5.2.2. $M \subset \mathrm{Hom}(V, W)$ is a *compression space* if there exist subspaces $V' \subset V$, $W' \subset W$ such that

1. $\mathrm{codim}(V') + \dim(W') = k$
2. M "compresses" V' into W', i.e. for every $A \in M$, $A(V') \subset W'$.

5.2.2 Equivalence and primitivity

In any case, question 5.2.1 seems rather unreasonable as stated; given any such space M of $n \times n$ matrices, there are infinitely many other such spaces (in fact an entire $G_{\det_n}$-orbit) that we could get out of it. So as a first step, we will quotient out by this action:

Definition 5.2.3. $M, N \in \mathcal{M}_{n \times n}(\mathbb{C})$ are *equivalent* if they differ by an action of $G_{\det_n}$.

Modding out by this equivalence does not however stop us from producing spaces of $n' \times n'$ matrices for $n' \neq n$. For $n' > n$, we have the following example.

Example 5.2.4. If $M \in \mathcal{M}_{n,n}(\mathbb{C})$ is the space of skew-symmetric matrices, then the $(n+1) \times (n+1)$ block matrix $M' = \begin{pmatrix} M & * \\ \mathbf{0} & * \end{pmatrix}$ is a space of singular matrices, and $\mathrm{rank}(M') = \mathrm{rank}(M) + 1$. Here, the asterisks denote that the corresponding entries are distinct indeterminates. We say that M' is *imprimitive* and has *primitive part* M.

We define primitivity in a coordinate-free manner. Intuitively, M is primitive if its low rank comes from that of some submatrix, under an appropriate choice of basis.

Definition 5.2.5. For $V' \subset V$, $W' \subset W$, let $\pi_{V',W'}$ be the projection from $\mathrm{Hom}(V, W)$ to $\mathrm{Hom}(V', W/W')$ which sends any $\phi : V \to W$ to the composition

$$V' \hookrightarrow V \xrightarrow{\phi} W \to W/W'.$$

We say $M \subset \mathrm{Hom}(V, W)$ is *imprimitive* if there exist V', W' for which $\mathrm{rank}(M) = \mathrm{rank}(\pi^{-1}_{V',W'}(\pi_{V',W'}(M)))$, in which case $\pi_{V',W'}(M)$ is called the *primitive part*; if no such V', W' exist, we say M is *primitive*.

Example 5.2.6. For any n, the space of all skew-symmetric $n \times n$ matrices is primitive. Compression spaces are imprimitive and in fact have primitive part zero.

We can now try asking Question 9 differently having circumscribed it somewhat.

Question 5.2.7. For a given n, up to equivalence, what are the primitive maximal subspaces of singular $n \times n$ matrices?

In [2], Atkinson solved this for $n = 3, 4$. We follow the bundle-theoretic proof due to Eisenbud and Harris in [19], which holds more generally for all spaces of matrices M for which a certain associated vector bundle has first Chern class equal to 1.

Theorem 5.2.8 ([19], Theorem 1.1)**.** *The primitive maximal subspaces of singular* 3×3 *matrices are all equivalent to the space of skew-symmetric matrices.*

Theorem 5.2.9 ([19], Theorem 1.2)**.** *The primitive maximal subspaces of singular* 4×4 *matrices are all equivalent to exactly one of*

$$\begin{pmatrix} a & b & 0 & 0 \\ 0 & 0 & a & b \\ c & 0 & d & 0 \\ 0 & c & 0 & d \end{pmatrix} \qquad \begin{pmatrix} -a & -b & 0 & 0 \\ c & 0 & -d & -b \\ -b & 0 & a & 0 \\ d & c & 0 & a \end{pmatrix}.$$

5.2.3 Two important sheaves

Fix vector spaces V and W of dimension v and w respectively. Let $M \subset \mathrm{Hom}(V, W)$ be a (rank k) space of matrices and let $\mathbb{P} = \mathbb{P}M$; we would like to encode all the information about M in terms of sheaves on $\mathbb{P}$. Consider the map of sheaves $\phi'_M : V \otimes \mathcal{O}_{\mathbb{P}}(-1) \to W \otimes \mathcal{O}_{\mathbb{P}}$ which over $A \in \mathbb{P}$ sends $v \otimes \lambda A$ to $\lambda A(v)$; one can recover M from this: on the level of global sections, ϕ'_M gives the map $V \otimes M \to W$ adjoint to the inclusion $M \hookrightarrow V^* \otimes W$.

Consider also the dual $\phi'^*_M : W^* \otimes \mathcal{O}_{\mathbb{P}} \to V^* \otimes \mathcal{O}_{\mathbb{P}}(1)$; we want the targets of these two maps to have the same amount of twisting, so twist the former map by $\mathcal{O}_{\mathbb{P}}(1)$ to get $\phi_M : V \otimes \mathcal{O}_{\mathbb{P}} \to W \otimes \mathcal{O}_{\mathbb{P}}(1)$. Define $\mathcal{E}_M$ and $\mathcal{F}_M$ to be the image sheaves $\mathrm{Im}(\phi_M)$ and $\mathrm{Im}(\phi'^*_M)$, both torsion-free of rank k. We can still almost recover M from $\mathcal{E}_M$ and $\mathcal{F}_M$ alone. Indeed, V and W^* sit inside $H^0(\mathcal{E}_M)$ and $H^0(\mathcal{F}_M)$ so that M arises as a projection of the space of linear maps associated to $H^0(\mathcal{E}_M) \otimes \mathcal{O}_{\mathbb{P}} \to H^0(\mathcal{F}_M)^* \otimes \mathcal{O}_{\mathbb{P}}(1)$. Indeed, if $V = H^0(\mathcal{E}_M)$ and $W^* = H^0(\mathcal{F}_M)$, then out of the map on global sections induced by this, we obtain the same map $V \to W \otimes M^*$ as before.

Observation 6. $\mathcal{E}_M$ and $\mathcal{F}_M$ satisfy the following properties:

1. $\mathcal{E}$ and $\mathcal{F}$ are the subsheaves of $\mathcal{E}_M^{**}$ and $\mathcal{F}_M^{**}$ generated by global sections.

2. $\mathcal{E}_M^{**} = \mathcal{F}_M^{*}(1)$ and $\mathcal{F}_M^{**} = \mathcal{E}_M^{*}(1)$.

Note that the point of taking the double duals of these two sheaves is to exploit the fact that rank one reflexive sheaves on $\mathbb{P}$ are locally free.

5.2.4 Criteria for compression spaces/primitivity

Lemma 5.2.10. *The following are equivalent:*

i) M is a compression space.

ii) $\mathcal{E}_M$ and $\mathcal{F}_M$ respectively contain trivial rank k_1 and k_2 vector bundles for which $k_1 + k_2 = \mathrm{rank}(M)$.

iii) $\mathcal{E}_M^{**}$ is a direct sum of rank 1 sheaves, specifically copies of $\mathcal{O}_{\mathbb{P}}$ and $\mathcal{O}_{\mathbb{P}}(1)$.

Proof. i) $\Rightarrow$ ii): Suppose V and W admit direct-sum decompositions $V' \oplus V''$ and $W' \oplus W''$ for which M sends V' and V'' into W' and W'' respectively. Because we are assuming that $\dim V'' + \dim W' = k$, ϕ_M injectively sends $V'' \otimes \mathcal{O}_{\mathbb{P}}$ to $W'' \otimes \mathcal{O}_{\mathbb{P}}(1)$; hence, $V'' \otimes \mathcal{O}_{\mathbb{P}}$ is trivial, so its image under ϕ_M is as well, and this is the desired trivial summand in $\mathcal{E}_M$, of dimension $\dim(V'')$. Taking the transpose of M, we similarly get a summand in $\mathcal{F}_M$, of dimension $\dim(W')$.

ii) $\Rightarrow$ iii): Because $\mathcal{E}_M$ has a trivial summand of rank k_1, so too does $\mathcal{E}_M^{**} = \mathcal{F}_M^{*}(1)$ which thus also has a trivial summand of rank k_2 given by a direct sum of k_2 copies of $\mathcal{O}_{\mathbb{P}}(1)$. Because $k_1 + k_2 = k$, iii) follows.

iii) $\Rightarrow$ i): Consider what happens in the simplest case when $\mathrm{Im}(\phi_M)$ is just a direct sum of k_1 copies of $\mathcal{O}_{\mathbb{P}}$ and k_2 copies of $\mathcal{O}_{\mathbb{P}}(1)$, in which case $H^0(\mathcal{E}_M) = V$, $H^0(\mathcal{F}_M) = W^*$, and $\mathcal{E}$ is reflexive. To produce the requisite $V' \subset V$ and $W' \subset W$ for M to be a compression space, denote by $\mathcal{E}'$ and $\mathcal{F}'$ the maximal summands respectively in $\mathcal{E}_M$ and $\mathcal{F}_M^*$ that are direct sums of copies of $\mathcal{O}_{\mathbb{P}}$. On global sections, the compositions $H^0(\mathcal{E}_M) \otimes \mathcal{O}_{\mathbb{P}} \to \mathcal{E}_M \to \mathcal{E}'$ and $\mathcal{F}' \to \mathcal{E}_M \to H^0(\mathcal{F}_M)^* \otimes \mathcal{O}_{\mathbb{P}}(1)$ respectively induce the maps $\phi_1 : V \to H^0(\mathcal{E}')$ and $\phi_2 : H^0(\mathcal{F}') \to W$.

Let $V' = \ker(\phi_1)$ and $W' = \mathrm{im}(\phi_2)$. Noting that reflexivity of $\mathcal{E}$ implies $\mathcal{E} = \mathcal{F}^*(1)$, we thus see that in the composition $H^0(\mathcal{E}_M \otimes \mathcal{O}_{\mathbb{P}} \to \mathcal{E}_M = \mathcal{F}^*(1) \to H^0(\mathcal{F}_M)^* \otimes \mathcal{O}_{\mathbb{P}}(1)$, $V' \otimes \mathcal{O}_{\mathbb{P}}$ is sent to $W' \otimes \mathcal{O}_{\mathbb{P}}(1)$, so M indeed carries V' into W'. Moreover, $\mathrm{codim}\, V' + \dim W' = k_1 + k_2 = k$, so M is indeed a compression space.

Lastly, we can reduce the general case in which $\mathcal{E}_M^{**}$ is a direct sum of rank 1 sheaves to the above case. $\mathcal{E}_M^{**}$ is reflexive so that these rank 1 sheaves are line bundles; in particular they must be copies of $\mathcal{O}_{\mathbb{P}}$ and $\mathcal{O}_{\mathbb{P}}(1)$ because $\mathcal{E}^{**} = \mathcal{F}^*(1)$, so $\mathcal{E}_M^{**}$ and $\mathcal{F}_M^{**}$ are generated by their global sections just like their subsheaves $\mathcal{E}_M$ and $\mathcal{F}_M$. So if we replace $\mathcal{E}_M$ and $\mathcal{F}_M$ in the above discussion by $\mathcal{E}_M^{**}$ and $\mathcal{F}_M^{**}$, we would conclude that $H^0(\mathcal{E}_M^{**}) \otimes \mathcal{O}_{\mathbb{P}} \to H^0(\mathcal{F}_M^{**})^* \otimes \mathcal{O}_{\mathbb{P}}(1)$ induces a map on global sections corresponding to a compression space M'. M is the projection of M', and $\mathrm{rank}(M) = \mathrm{rank}(M')$, so M is a compression space as desired. □

The following is a neat consequence of this that we will use in Section 5.4.

Corollary 5.2.11. *If* $\dim(M) = 2$*, then M is a compression space.*

Proof. This follows from the Birkhoff-Grothendieck theorem characterizing torsion-free sheaves on $\mathbb{P}^1$:

Theorem 5.2.12 ([24])**.** *Every torsion-free sheaf over $\mathbb{P}^1$ is the direct sum of line bundles.*

$\mathbb{P} = \mathbb{P}M$ is just the projective line, so by Theorem 5.2.12, $\mathcal{E}_M$ is the direct sum of line bundles. Lemma 5.2.10 tells us that M is a compression space. □

Similar arguments give us the following characterization of primitivity.

Lemma 5.2.13. *The following are equivalent:*

i) M is imprimitive.

ii) $\mathcal{E}_M$ *or* $\mathcal{F}_M$ *has* $\mathcal{O}_{\mathbb{P}}$ *as a summand.*

iii) $\mathcal{E}_M^{**}$ *or* $\mathcal{F}_M^{**}$ *has a summand of rank 1.*

Proof. i) ⇒ ii): Assume without loss of generality that $\dim V \leq \dim W$. If $\operatorname{rank}(M) = \dim V$, then $\mathcal{E}_M \equiv V \otimes \mathcal{O}_{\mathbb{P}}$ and we'd be done, so assume $\operatorname{rank}(M) < \dim V$. If M is imprimitive, there exists a hyperplane $H \subset V$ for which

$$\operatorname{rank}(M) = \operatorname{rank}(\pi^{-1}(\pi(M))), \tag{5.3}$$

where $\pi := \pi_{H,W}$. $\operatorname{rank}(\pi(M))$ is simply $\operatorname{rank}(\phi_M(H \otimes \mathcal{O}_{\mathbb{P}}))$, and $\operatorname{rank}(\pi^{-1}(\pi(M))) = \pi(M) + 1$ from adding back a column of indeterminates to $\pi(M)$. Equality 5.3 therefore implies that if L is a complement of H in V, then $\phi_M(L \otimes \mathcal{O}_{\mathbb{P}}) \simeq \mathcal{O}_{\mathbb{P}}$, and the desired result follows.

ii) ⇒ iii): $\mathcal{E}_M$ is a subsheaf of $\mathcal{E}_M^{**}$, so this implication is trivial.

iii) ⇒ i): Suppose $\mathcal{E}_M^{**}$ has a rank 1 summand. Because $\mathcal{E}_M^{**}$ and $\mathcal{F}_M^{**}$ are reflexive and $\mathcal{E}_M^{**} = \mathcal{F}_M^{*}(1)$, the rank 1 summand has to be either $\mathcal{O}_{\mathbb{P}}$ or $\mathcal{O}_{\mathbb{P}}(1)$. If the latter, then $\mathcal{F}_M^{*}$ and thus $\mathcal{F}_M^{**}$ has $\mathcal{O}_{\mathbb{P}}$ as a summand, so assume $\mathcal{E}_M^{**}$ has $\mathcal{O}_{\mathbb{P}}$ as a summand. Consider the map on global sections induced by $V \otimes \mathcal{O}_{\mathbb{P}} \to \mathcal{E}_M \to \mathcal{E}_M^{**} \to \mathcal{O}_{\mathbb{P}}$. Because $\mathcal{E}_M$ and $\mathcal{E}_M^{**}$ have the same rank, this map $V \to \mathbb{C}$ cannot be identically zero. Take V' to be its kernel and note that (5.3) holds for $\pi = \pi_{V',W}$, proving imprimitivity of M. □

5.2.5 Effective criterion for primitivity

While these criteria will be useful for proving the technical crux of Harris and Eisenbud's result, at the end of the day we will have produced spaces of matrices whose primitivity we need to verify in an effective manner, e.g. with *Macaulay2*. To do this, we pass from the category of quasicoherent sheaves to that of graded modules.

Define the polynomial ring $S = \mathbb{C}[M^*]$ and let f_M be the map on graded free S-modules corresponding to ϕ_M. Define E_M to be the graded S-module corresponding to $\mathcal{E}_M$, equivalently the image of f_M, and F_M to be the graded S-module corresponding to $\mathcal{F}_M$, equivalently the image of $f_M^*(1)$. Our criteria for compression spaces and primitivity pass over easily to this setting by the following.

Lemma 5.2.14. *Let E be a graded S-module generated in degree zero and $\mathcal{E}$ be the corresponding sheaf on $\mathbb{P}$. The following are equivalent:*

i) S *is a direct summand of* E

ii) S *is a direct summand of* E^{**}

iii) $\mathcal{O}_{\mathbb{P}}$ *is a direct summand of* $\mathcal{E}$

iv) $\mathcal{O}_{\mathbb{P}}$ *is a direct summand of* $\mathcal{E}^{**}$

Proof. i) ⇔ iii) and ii) ⇔ iv) follow from the fact that $\mathcal{E}$ and $\mathcal{E}^{**}$ are the quasicoherent sheaves over $\mathbb{P}$ associated to the graded S-modules E and E^{**}. i) ⇒ ii) and iii) ⇒ iv) are obvious. It remains to prove ii) ⇒ i), from which the corresponding implication iii) ⇒ iv) will follow. E and E^{**} have the same rank, so the map $E \to S$ induced by the projection map $E^{**} \twoheadrightarrow S$ given by assumption ii) is nonzero. In particular, because S is generated in degree zero, $E \to S$ must be nonzero in degree zero, implying that it is also a projection map. □

Denoting the kernel of f_M by K_M^E and that of $f_{M^*}(1)$ by K_M^F, we note that the presence or absence of the direct summand S in E_M or F_M is not detected by K_M^E or K_M^F, giving the following effective test for primitivity.

Lemma 5.2.15. *If B (resp. B') is the largest submodule of the minimal first syzygy of $(K_M^E)^*(1)$ (resp $(K_M^F)^*(1)$) that is generated in degree zero, then M is primitive if and only if* $\operatorname{rank}(B) = \operatorname{rank}(B') = \operatorname{rank}(M)$.

5.2.6 Characterization of $\mathcal{E}_M$ with low first Chern class

The Chern classes capture the extent to which a given vector bundle is trivial; we will use the first Chern class throughout this subsection. While technically $c_1(\mathcal{E}_M) = d \cdot c_1(\mathcal{O}_{\mathbb{P}}(1))$, we will write $c_1(\mathcal{E}_M) = d$ for convenience. If $d = 0$, i.e. if $\mathcal{E}_M$ were trivial, then by Lemma 5.2.10 M would be a compression space, so we assume that $c_1(\mathcal{E}_M), c_1(\mathcal{F}_M) > 0$.

Proposition 5.2.16. *If $k \le 3$ and $\mathcal{E}_M$ is nontrivial, then $c_1(\mathcal{E}_M) = c_1(\mathcal{O}_{\mathbb{P}}(1))$.*

Proof. Note that

$$c_1(\mathcal{F}_M) \le c_1(\mathcal{F}_M^{**}) = c_1(\mathcal{E}_M^*(1)) = k \cdot c_1(\mathcal{O}_{\mathbb{P}}(1)) - c_1(\mathcal{E}_M),$$

from which we conclude that $c_1(\mathcal{E}_M) + c_1(\mathcal{F}_M) \le k \cdot c_1(\mathcal{O}_{\mathbb{P}}(1))$. We can thus assume without loss of generality that $1 \le d \le \lfloor k/2 \rfloor$, so in particular, $d = 1$ when $k \le 3$. □

This is enough to give the following characterization of $\mathcal{E}_M$, which forms the technical crux of Eisenbud/Harris' proof of Theorems 5.2.8 and 5.2.9.

Theorem 5.2.17. *If $c_1(\mathcal{E}_M) = 1$ and $\mathcal{O}_P$ is not a summand of $\mathcal{E}_M$, then $\mathcal{E}_M$ is either $\mathcal{O}_{\mathbb{P}}(1)$ or the universal quotient bundle $Q_{\mathbb{P}}$ over $\mathbb{P}$.*

Proof. Define $\Psi : \mathbb{P} \to G := G(v-k, V)$ as the map sending any point $A \in \mathbb{P}$ to $\ker(A) \in G(v-k, V)$.

Lemma 5.2.18. *Ψ is a projection of $\mathbb{P}$ onto a smaller projective space $\mathbb{P}'$ embedded in $G(v-k, V)$ via the Plücker embedding.*

Proof. We know that outside some codimension ≥ 2 subvariety $\Sigma \subset \mathbb{P}$, $\mathcal{E}_M$ is locally free and thus the pullback of the universal quotient bundle over G along Ψ. The fact that $c_1(\mathcal{E}_M) = 1$ then implies that outside of Σ, Ψ sends lines in $\mathbb{P}$ to lines in G (with respect to the Plücker embedding). □

The following is a standard fact from Schubert calculus.

Lemma 5.2.19. *If $\mathbb{P}'$ is a linear space on G, then $\mathbb{P}'$ is either a subspace of the set of $(v-k)$-planes containing a particular $(v-k-1)$-plane, or a subspace of the set of $(v-k)$-planes lying in a fixed $(v-k+1)$-plane.*

We can now proceed by casework on the particular linear space $\mathbb{P}'$ that Ψ projects $\mathbb{P}$ onto.

Case 1. $\mathbb{P}'$ is a subspace of the set of $(v-k)$-planes containing a particular $(v-k-1)$-plane.

First, we may and shall quotient V out by the $(v-k-1)$-dimensional subspace common to all planes in $\mathbb{P}'$ in order to take $\dim(V) = k+1$. We claim that in this case, $\mathcal{E}_M$ is the universal quotient bundle over $\mathbb{P}$. The strategy is to consider the exact sequence

$$0 \to \mathcal{K} \to \mathcal{O}_{\mathbb{P}} \otimes V \to \mathcal{E}_M \to 0, \tag{5.4}$$

where $\mathcal{K} := \ker(\phi_M)$, and argue that (1) $\ker(\phi_M) \simeq \mathcal{O}_{\mathbb{P}}(-1)$ and 2) $M = V$.

Because $\operatorname{rank}(\mathcal{K}) = (k+1) - k = 1$, (1) is a straightforward Chern class computation. The fact that $\mathcal{K}$ is a second syzygy implies it is reflexive and thus a line bundle. Furthermore, by (5.4), $c_1(\mathcal{K}) = c_1(\mathcal{O}_{\mathbb{P}} \otimes V) - c_1(\mathcal{E}_M) = -1$, so we conclude that $\mathcal{K} \simeq \mathcal{O}_{\mathbb{P}}(-1)$.

For 2), we show that the inclusion $M \hookrightarrow \operatorname{Hom}(V, W)$ factors through a surjection $M \to V$. First, because $M^* = H^0(\mathcal{O}_{\mathbb{P}}(1))$, $\mathcal{K} \to \mathcal{O}_{\mathbb{P}} \otimes V$ factors through $\mathcal{O}_{\mathbb{P}} \otimes M$. The map $T : M \to V$ induced by $\mathcal{O}_{\mathbb{P}} \otimes M \to \mathcal{O}_{\mathbb{P}} \otimes V$ is clearly surjective because $\mathcal{K}$ does not lie in any trivial summand of the target.

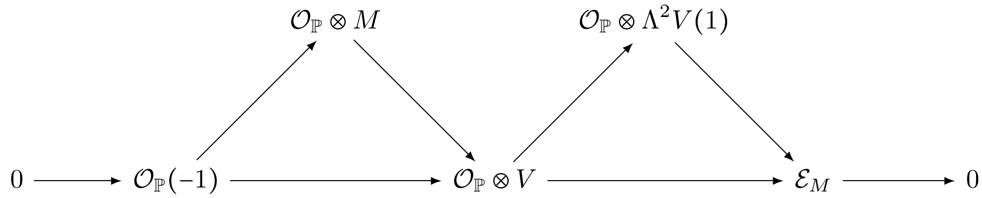
Next, if we dualize (5.4), twist once, and look at global sections, we have

$$0 \to H^0(\mathcal{E}_M^* \otimes \mathcal{O}_{\mathbb{P}}(1)) \to H^0(\mathcal{O}_{\mathbb{P}} \otimes V^*(1)) \to H^0(\mathcal{O}_{\mathbb{P}}(2)),$$

we see that $\Lambda^2 V^*$ fits in as the kernel of $H^0(\mathcal{O}_{\mathbb{P}} \otimes V^*(1)) \to H^0(\mathcal{O}_{\mathbb{P}}(2))$. In particular, we can factor ϕ_{M^*} through the second Koszul complex map to get

$$\phi_{M^*} : \mathcal{E}_M^* \to \mathcal{O}_{\mathbb{P}} \otimes \Lambda^2 V^*(-1) \to \mathcal{O}_{\mathbb{P}} \otimes V^*.$$

To conclude, we have the diagram



which exhibits the inclusion $M \to \mathrm{Hom}(V, W)$ as the composition $M \twoheadrightarrow V$ followed by an inclusion $V \hookrightarrow \mathrm{Hom}(V, \Lambda^2 V)$ and a projection $\mathrm{Hom}(V, \Lambda^2 V) \twoheadrightarrow \mathrm{Hom}(V, W)$. So because T is surjective, $M = V$, completing the proof for this case.

Case 2. $\mathbb{P}'$ is a subspace of the set of $(v-k)$-planes lying in a fixed $(v-k+1)$-plane.

We claim that in this case, $\mathcal{E}_M$ is $\mathcal{O}_{\mathbb{P}}(1)$. We first show that $\mathrm{rank}(\mathcal{E}_M) = 1$. Let $V = V' \oplus V''$ where V' is the $(v-k+1)$-plane containing $\mathbb{P}'$ and V'' is some $(k-1)$-dimensional complement. Because $\mathrm{rank}(\phi_M(\mathcal{O}_{\mathbb{P}} \otimes V')) = 1$, $\mathrm{rank}(\phi_M(\mathcal{O}_{\mathbb{P}} \otimes V'')) = \mathrm{rank}(\mathcal{O}_{\mathbb{P}} \otimes V'')$, so $\phi_M(\mathcal{O}_{\mathbb{P}} \otimes V'') \simeq \mathcal{O}_{\mathbb{P}} \otimes V''$. In particular, $\mathcal{E}_M = \phi_M(\mathcal{O}_{\mathbb{P}} \otimes V') \oplus (\mathcal{O}_{\mathbb{P}} \otimes V'')$, and the latter summand must vanish by the assumption that $\mathcal{E}_M$ does not contain a copy of $\mathcal{O}_{\mathbb{P}}$ as a summand, so $\mathrm{rank}(\mathcal{E}_M) = 1$ as claimed.

$\mathcal{E}_M^{**}$ is therefore some line bundle, say $\mathcal{O}_{\mathbb{P}}(1) \otimes L$ for $L \subset W$. Pick a complement $W' \subset W$ so that the map $\mathcal{E}_M \to \mathcal{O}_{\mathbb{P}}(1) \otimes W'$ induced by $\mathcal{E}_M \to \mathcal{E}_M^{**} \otimes W$ is the zero map. Ignoring W', we see that of the family of maps corresponding to the map $\mathcal{O}_{\mathbb{P}} \otimes H^0(\mathcal{O}_{\mathbb{P}}(1)) \to \mathcal{O}_{\mathbb{P}}(1)$, M is a projection. But this family is $\mathrm{Hom}(M^*, W) \simeq M$, so the projection must in fact be the identity, and we conclude that $\mathcal{E}_M \simeq \mathcal{O}_{\mathbb{P}}(1)$. □

A useful consequence of this is that all primitive spaces for $\det_3$ and $\det_4$ arise from taking certain submatrices of (5.2).

Corollary 5.2.20. *If M is primitive and $c_1(\mathcal{E}_M) = 1$, then M is the image of a projection of* $\mathrm{Hom}(M, \Lambda^2 M)$ *from a subspace of* $\Lambda^2 M$. *M consists of* $(\dim M) \times n$ *matrices of rank* $\dim M - 1$ *for some* $\dim M \le n \le \binom{\dim M}{2}$.

Proof. Because M is primitive, neither $\mathcal{E}_M$ nor $\mathcal{F}_M$ has $\mathcal{O}_{\mathbb{P}}$ as a summand. In particular, $\mathcal{E}_M$ also cannot have $\mathcal{O}_{\mathbb{P}}(1)$ as a summand, or else $\mathcal{E}_M^{**} = \mathcal{F}_M^*(1)$ has the same and $\mathcal{E}_M$ has $\mathcal{O}_{\mathbb{P}}$ as a summand, a contradiction. By Theorem 5.2.17, $\mathcal{E}_M = Q_{\mathbb{P}}$. But $Q_{\mathbb{P}}$ is also the image of the second map $d_2 M \otimes \mathcal{O}_{\mathbb{P}} \to \Lambda^2 M \otimes \mathcal{O}_{\mathbb{P}}(1)$ in the Koszul complex

$$0 \to \mathcal{O}_{\mathbb{P}}(-1) \xrightarrow{d_1} M \otimes \mathcal{O}_{\mathbb{P}} \xrightarrow{d_2} \Lambda^2 M \otimes \mathcal{O}_{\mathbb{P}}(1) \to \cdots.$$

We claim that fiberwise, this map is just multiplication in the Koszul complex. Indeed, explicitly, the corresponding map on global sections sends $y \in M$ to $\sum_i (y_i \wedge y) \otimes x_i \in \Lambda^2 M \otimes M^*$, where $\{y_i\}$ and $\{x_i\}$ are dual bases for M and $M^* = H^0(\mathcal{O}_{\mathbb{P}}(1))$ respectively. In the fiber over $A \in \mathbb{P}$, this map sends $y \in M$ to $\sum_i x_i(A) \cdot (y_i \wedge y) = A \wedge y$, as desired.

Writing $\Lambda^2 M$ as some $W \oplus W'$, we conclude that M is the image of the projection via $\pi_{V,W'}$ of the subspace of $\mathrm{Hom}(M, \Lambda^2 M)$ corresponding to multiplication by members of M in the exterior algebra. If $V \neq M$ or $\dim W \le \dim M$, then M would be imprimitive, so to get the second half of the corollary, it is enough to note that this subspace has rank equal to $\mathrm{rank}(M)$ and specifically to $\dim M - 1$. □

Theorem 5.2.8 follows automatically from the above because when M is a space of 3×3 matrices, the space of maps $M \subset \mathrm{Hom}(M, \Lambda^2 M)$ is parametrized by the generic 3×3 skew-symmetric matrix. For 4×4 matrices, we need to do some extra work because $M \subset \mathrm{Hom}(M, \Lambda^2 M)$ is parametrized by a 4×6 matrix, so there are multiple projections to spaces of 4×4 matrices that we can consider.

5.2.7 Classification for $\det_4$

As a final step before we can apply Corollary 5.2.20 to classify the maximal primitive spaces on $\{\det_n = 0\}$ for $n = 4$, we need to understand the equivalence classes of such (rank $\dim M - 1$) projections of spaces $M \subset \mathrm{Hom}(M, \Lambda^2 M)$. Regard M as a space of maps $M \subset \mathrm{Hom}(M, \Lambda^2 M)$ via Koszul multiplication.

Lemma 5.2.21. *For $W' \subset \Lambda^2 M$, if M' is the projection $\pi_{M,W'}(M)$ and* $\mathrm{rank}(\pi_{M,W'}(M)) = \dim M - 1$, *then $\pi_{M,W'} : M \to M'$ is an isomorphism.*

Proof. Suppose the kernel of $\pi_{M,W'}$ contains a nonzero $A \in M$, then for any $A' \in M$ not in the span of A, the map $\pi_{M,W'}(A')$ given by $\cdot \wedge A' : M \to \Lambda^2 M/W'$ has a kernel of dimension at least two, contradicting the assumption that $\mathrm{rank}(M') = m-1$. □

Lemma 5.2.22. *If $W', W'' \subset \Lambda^2 M$, then $\pi_{M,W'}(M)$ and $\pi_{M,W''}(M)$ are equivalent if and only if W' and W'' are conjugates under the action of $GL(M)$ on $\Lambda^2 M$.*

Proof. In the easier direction, if W' and W'' are conjugate by some $\alpha \in \mathrm{GL}(M)$, then $\Lambda^2\alpha : \Lambda^2 M \to \Lambda^2 M$ induces a map $\beta : (\Lambda^2 M)/W' \to (\Lambda^2 M)/W''$. For any $A \in M$, Koszul multiplication by A is associated with the corresponding diagram:

$$\begin{array}{ccccc} M & \xrightarrow{\wedge A} & \Lambda^2 M & \longrightarrow & (\Lambda^2 M)/W' \\ \Big\downarrow \alpha & & \Big\downarrow \Lambda^2\alpha & & \Big\downarrow \beta \\ M & \xrightarrow{\wedge \alpha(A)} & \Lambda^2 M & \longrightarrow & (\Lambda^2 M)/W'' \end{array}$$

This establishes an equivalence between $\pi_{M,W'}(M)$ and $\pi_{M,W''}(M)$ via (α,β).

In the other direction, suppose $\pi_{M,W'}(M)$ and $\pi_{M,W''}(M)$ are equivalent via some (α,β). It suffices to show that the following diagram commutes

$$\begin{array}{ccc} M & \hookrightarrow & M^* \otimes ((\Lambda^2 M)/W') \\ \Big\downarrow \gamma & & \Big\downarrow \alpha^{-1} \otimes \beta \\ M & \hookrightarrow & M^* \otimes ((\Lambda^2 M)/W'') \end{array}$$

if and only if $\gamma = \lambda\alpha$ for a particular scalar λ, because the diagram obtained by taking adjoints

$$\begin{array}{ccc} M \otimes M & \longrightarrow & (\Lambda^2 M)/W' \\ \Big\downarrow (\lambda\alpha) \otimes \alpha & & \Big\downarrow \beta \\ M \otimes M & \longrightarrow & (\Lambda^2 M)/W'' \end{array}$$

exhibits the fact that M' and M'' are conjugate by $\Lambda^2\alpha$. To see that (5.2.7) commutes, note that for a given $\cdot \wedge A \in \mathrm{Hom}(M, (\Lambda^2 M)/W')$, the kernel is generated by A itself if A is generic, because $\mathrm{rank}(M') = m-1$. Because $\alpha(A)$ is generic if A is generic, the same holds for $\cdot \wedge \alpha(A)$, so the linearity of γ and α implies that γ must be a multiple of α as desired. □

We are now ready to prove Theorem 5.2.9.

Proof of Theorem 5.2.9. By Corollary 5.2.20 and Lemmas 5.2.21 and 5.2.22, it suffices to classify all $\mathrm{GL}(M)$-orbits of 2-planes $W' \subset \Lambda^2 M$, where $\dim M = 4$. Projectivizing, the problem becomes classifying all $\mathrm{PGL}(M)$-orbits of lines in $\mathbb{P}(\Lambda^2 M) = \mathbb{P}^5$.

$\mathrm{PGL}(M)$ is the automorphism group of $G := \mathbb{G}(1,3)$ and cannot send points on G to points off G, or vice versa. For this reason, lines transverse to/tangent to/contained in G can only be mapped to other lines transverse to/tangent to/contained in G, and the action of $\mathrm{PGL}(M)$ is transitive on these three classes, giving exactly three $\mathrm{PGL}(M)$-orbits of lines.

Let $\{x_{ij} \mid i < j\}$ be a basis for $\Lambda^2 M$. To look for representatives of these orbits, it will help to recall that G is cut out the Plucker relation $x_{12}x_{34} - x_{13}x_{24} + x_{14}x_{23} = 0$. Then the line parametrized by $[\alpha x_{12} + \beta x_{34}]$ is transverse to G, and if we take W' to be the corresponding 2-plane in $\Lambda^2 M$ parametrized by $\alpha x_{12} + \beta x_{34}$, then quotienting out by W' corresponds to removing the first and last columns of (5.2), giving the first matrix in the statement of Theorem 5.2.9.

The line parametrized by $[\alpha(x_{12}+x_{34})+\beta x_{13}]$ is tangent to G, so quotienting out the corresponding 2-plane W' corresponds to removing the second column and subtracting the last column from the first in (5.2) to give the second matrix in the statement of Theorem 5.2.9.

Lastly, the line parametrized by $[\alpha x_{12}+\beta x_{13}]$ lies in G, so quotienting out the corresponding 2-plane W' corresponds to removing the first two columns of W' to give the matrix

$$\begin{pmatrix} d & 0 & 0 & 0 \\ 0 & c & d & 0 \\ 0 & -b & 0 & d \\ -a & 0 & -b & -c \end{pmatrix}.$$

But this is imprimitive because the the bottom-right 3×3 matrix is a translate of the space of 3×3 skew-symmetric matrices. On the other hand, the previous two matrices we obtained are primitive by Lemma 5.2.15.

It remains to show that these two matrices are inequivalent. There are many ways to see this; one rather heavy-handed way is to show that one "gives rise" to a boundary component on Det_4, while the other does not, a computation we will carry out at the end of Section 5.4 in Observation 10. □

5.3 Boundary components of Det_3

Among the irreducible components on ∂Det_n, there is the obvious *endomorphism component* $\mathrm{End}(W)\cdot[\det_n]\backslash \mathrm{GL}(W)\cdot[\det_n]$ consisting of degenerate translates of $[\det_n]$. Here, we say that a polynomial p over W is degenerate if there exists some nontrivial linear relation among $\left\{\frac{\partial}{\partial x_i}p\right\}$ for any basis $\{x_i\}$ of W.

For $n=3$, Landsberg et al. showed in [36] that additionally, if we define $P_3 := \lim_{t\to 0}[\det(W_{skew}+t\cdot W_{sym})]$, then $\overline{\mathrm{GL}(W)\cdot P_3}$ is an irreducible component.[1] Only recently was it shown that these are the *only* two irreducible components of ∂Det_3.

Theorem 5.3.1. *There are exactly two irreducible components in the boundary of Det_3, namely* $\mathrm{End}(W)\cdot[\det_3]\backslash(GL(W)\cdot[\det_3])$ *and the $GL(W)$-orbit closure of P_3.*

In this section, we give the proof due to Huttenhain and Lairez in [27], in which they exhibited these two boundary components as the exceptional divisor and proper transform of a particular blowup.

Proving Theorem 5.3.1 was one of the original goals of this work, but during the course of this project, it was proven independently by Huttenhain and Lairez in [27] before all the details of our approach could be worked out. Whereas their approach is by way of resolution of indeterminacies, as sketched at the beginning of Section 5.2, ours is essentially combinatorial. We present their approach in this section and our own in the next.

Recall the setup from the beginning of Section 5.2. Denote the space of 3×3 matrices by W and $\mathrm{End}(W)$ by E. Define the $\mathbb{G}_{\det_3}$-invariant rational map $\phi:\mathbb{P}E \dashrightarrow Det_n$ sending $[a]\in\mathbb{P}E$ to $[\det_3\circ a]$, and denote by B its indeterminacy locus, i.e. the set of all $[a]$ for which $a(W)$ consists solely of singular matrices. The goal is to resolve the indeterminacies of ϕ, specifically on a carefully chosen open subset $U\subset\mathbb{P}(E)$.

For this particular U, we consider the closure of the graph X of $\phi|_U$, and denote by ρ and ψ its projections to U and Det_3. Define Z to be the hypersurface of singular elements of $\mathbb{P}(E)$ and D its preimage under ρ. By our choice of U, it will turn out that ψ surjects onto Det_3 (Lemma 5.3.2), while $\psi(X\backslash D)=\mathrm{GL}(W)\cdot[\det_3]$ by definition, so we conclude that $\partial Det_3\subset\psi(D)$. But ρ will turn out to be the blowup along a *smooth* subvariety $B'\subset B\subset U$ contained inside Z (Lemmas 5.3.3 and 5.3.5). So the preimage D of Z under this blowup map has exactly two components, the exceptional divisor and proper transform, and we're done.

5.3.1 Details of the argument

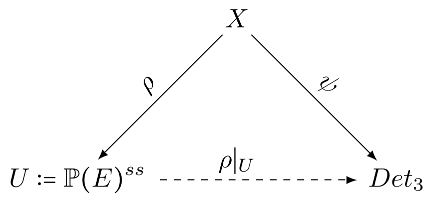
We will take $U=\mathbb{P}(E)^{ss}$, i.e. the set of semistable points under the action of $\mathbb{G}_{det_n}$. This is the set of $[a]\in\mathbb{P}(E)$ for which $0\notin\overline{G_{\det_3}\cdot a}$, or equivalently the set of $[a]\in\mathbb{P}(E)$ for which there exists a non-constant homogeneous $f\in\mathbb{C}[E]^{G_{\det_3}}$ that does not vanish on a.

[1] See Example 5.4.4 in Section 5.4 for a proof. The actual statement of this result in [36] was for all odd n, and in Section 5.5 we discuss this more general result.

Define

$$X = \overline{\{([a],[\det_3 \circ a]) \mid a \in \mathbb{P}(E)^{ss}\}}$$

and projection maps as shown in the following diagram:



Here the projection ρ is also the blowup of $\mathbb{P}(E)^{ss}$ along the ideal sheaf corresponding to the vanishing of $\det_3 \circ a$, which has as support the indeterminacy locus $B \cap \mathbb{P}(E)^{ss}$, and the regular map ψ is a resolution of the indeterminacies of ϕ.

Lemma 5.3.2. *ψ is surjective.*

Proof. Because $\mathrm{im}(\psi)$ contains $\mathrm{GL}(W)\cdot[\det_3]$, it suffices to show that $\mathrm{im}(\psi)$ is closed. Define the projective variety $T := \mathbb{P}(E) \times \mathbb{P}(\mathbb{C}[W]_3)$ and let $\mathbb{G}_{det_3}$ act upon it by $h \cdot ([a],[P]) = ([h \cdot a],[P])$. Obviously, $T^{ss} = \mathbb{P}(E)^{ss} \times \mathbb{P}(\mathbb{C}[W]_3)$, and because ψ is $\mathbb{G}_{det_3}$-invariant, it factors through the GIT quotient and projective variety $T^{ss}//\mathbb{G}_{det_3}$. But we know that the GIT quotient map $\pi : T^{ss} \to T^{ss}//\mathbb{G}_{det_3}$ sends $\mathbb{G}_{det_3}$-invariant closed subsets to closed subsets. So if we write ψ as $\psi' \circ \pi$ for some regular $\psi' : T^{ss}//\mathbb{G}_{det_3} \to \mathbb{P}(\mathbb{C}[W]_3)$, we conclude that $\mathrm{im}(\psi)$ must be closed: as a closed subset of a projective variety, $\pi(X)$ is itself a projective variety, so ψ' sends $\pi(X)$ to a closed subset of $\mathbb{P}(\mathbb{C}[W]_3)$, as desired. □

By Theorem 5.2.8, we have a complete characterization of the components of B: for every $[a] \in B$, there is a $g \in \mathbb{G}^o_{\det_3}$ for which $(g \cdot a)(W)$ is a subspace of one of the following spaces of matrices:

$$U_1^{cmp} = \begin{pmatrix} * & * & 0 \\ * & * & 0 \\ * & * & 0 \end{pmatrix}, (U_1^{cmp})^T = \begin{pmatrix} * & * & * \\ * & * & * \\ 0 & 0 & 0 \end{pmatrix}, U_2^{cmp} = \begin{pmatrix} * & * & * \\ * & 0 & 0 \\ * & 0 & 0 \end{pmatrix}, W_{skew} = \begin{pmatrix} 0 & a & -b \\ -a & 0 & c \\ b & -c & 0 \end{pmatrix}.$$

Denote the corresponding components of B by B_1, B_2, B_3, and B_{skew}, respectively. It turns out that in $\mathbb{P}(E)^{ss}$, the indeterminacy locus of ϕ only depends on B_{skew}.

Lemma 5.3.3. *In B, only component B_{skew} intersects $\mathbb{P}(E)^{ss}$.*

Proof. We verify that B_1, B_2, and B_3 do not meet $\mathbb{P}(E)^{ss}$ by showing that $\overline{\mathbb{G}_{\det_3} \cdot B_i}$ contains zero for $i = 1, 2$. Indeed, note that

$$\lim_{t\to 0} \begin{pmatrix} t & 0 & 0 \\ 0 & t & 0 \\ 0 & 0 & t^{-2} \end{pmatrix} U_0^1 = 0, \quad (U_0^1)^T \lim_{t\to 0} \begin{pmatrix} t & 0 & 0 \\ 0 & t & 0 \\ 0 & 0 & t^{-2} \end{pmatrix} = 0, \quad \lim_{t\to 0} \begin{pmatrix} t^2 & 0 & 0 \\ 0 & t^{-1} & 0 \\ 0 & 0 & t^{-1} \end{pmatrix} U_0^2 = 0.$$

Lastly, we check that B_{skew} meets $\mathbb{P}(E)^{ss}$ by exhibiting a nonconstant homogeneous $\mathbb{G}_{\det_3}$-invariant regular function that does not vanish on a point in B_{skew}.

For general points $M_1, M_2, M_3 \in W$, consider $\tau_0 : E \to \mathbb{C}$ sending $a \in E$ to

$$\mathrm{trace}\left(a(p_1) \cdot \mathrm{adj}(a(p_2)) \cdot a(p_3) \cdot \mathrm{adj}(a(p_1 + p_2 + p_3))\right),$$

where adj denotes the adjugate matrix. Then it is apparent that $\tau_0 \in \mathbb{C}[E]^{\mathbb{G}^o_{\det_3}}$ so that the map $\tau : a \mapsto \tau_0(a) + \tau_0(a^T)$ is $\mathbb{G}_{\det_3}$-invariant. Moreover, for three general linear forms $\ell_1, \ell_2, \ell_3 : W \to \mathbb{C}$ and a point in B_{skew} given by

$$b = \begin{pmatrix} 0 & \ell_1 & -\ell_2 \\ -\ell_1 & 0 & \ell_3 \\ \ell_2 & -\ell_3 & 0 \end{pmatrix},$$

one can check that $\tau(b) \neq 0$, so $B \cap \mathbb{P}(E)^{ss}$ and $B_{skew} \cap \mathbb{P}(E)^{ss}$ indeed intersect. □

Let b be the point in $B_{skew} \cap \mathbb{P}(E)^{ss}$ obtained in the proof of Lemma 5.3.3.

Lemma 5.3.4. $B_{skew} \cap \mathbb{P}(E)^{ss} = [\mathbb{G}_{det_3} \cdot b \cdot GL(W)]$.

Proof. Inclusion from right to left is obvious. For the other direction, suppose that $[a] \in B_{skew} \cap \mathbb{P}(E)^{ss}$. If a is of rank at most 2, then by Corollary 5.2.11, $a(W)$ is a compression space, i.e. $a \in B_1 \cup B_2 \cup B_3$, contradicting Lemma 5.3.3. If a is of rank at most 3, it also cannot lie in $B_1 \cup B_2 \cup B_3$ and, up to a left-action of $\mathbb{G}_{\det_3}$, must have image W_{skew}. So a is a translate of b by a right $\mathrm{GL}(W)$-action. □

The following is the technical crux of the proof.

Lemma 5.3.5. *$B_{skew} \cap \mathbb{P}(E)^{ss}$ is smooth and thus reduced so that ρ is the blowup of $\mathbb{P}(E)^{ss}$ along it.*

Proof. While we know that ρ is the blowup of $\mathbb{P}(E)^{ss}$ along the ideal sheaf $\mathcal{I}$ generated by the conditions imposed by $\det_3 \circ a = 0$ and supported on $B \cap P(E)^{ss} = B_{skew} \cap \mathbb{P}(E)^{ss}$, we need $\mathcal{I}$ to be reduced for ρ to be the blowup of $\mathbb{P}(E)^{ss}$ along $B_{skew} \cap \mathbb{P}(E)^{ss}$. To show this, we will show that $\mathcal{I}$ is smooth.

But this ideal sheaf is fixed under the left and right actions of $\mathbb{G}_{\det_3}$ and $\mathrm{GL}(W)$ respectively, just as $B_{skew} \cap \mathbb{P}(E)^{ss}$ is an orbit under these actions by Lemma 5.3.4. It thus suffices to check at a single point in this orbit, like $[b]$, whether dimension of the tangent plane $T_{[b]}\mathcal{I}$ at that point agrees with dimension of B_{skew}. The latter is equal to $\dim T_{[b]} B_{skew}$, because the dimension of B_{skew} agrees with the tangent space at any point of B_{skew} as B_{skew} is also an orbit.

Both tangent spaces' dimensions can be computed in *Macaulay2*. They are given by

$$T_{[b]}\mathcal{I} = \{c \in T_{[b]}\mathbb{P}(E) \mid \det(b + tc) = O(t^2) \ \forall p \in W\}$$

$$T_{[b]}B_{skew} = \{mb+bc \mid m \in T_1\mathbb{G}_{\det_3}, c \in T_1\mathrm{GL}(W)\} = \{p \in W \mapsto Mb(p)+b(p)N+b(c(p)) \in W \mid M, N \in W, c \in E\},$$

and their dimensions turn out to equal 34. □

Proof of Theorem 5.3.1. Because ρ is a blowup along a smooth subvariety in Z, $D := \rho^{-1}(Z)$ consists of exactly two components, the exceptional divisor and the proper transform. But $\psi(X \backslash D) = \phi(\mathrm{GL}(W)) = \mathrm{GL}(W) \cdot [\det_3]$, so $\partial Det_3 \subset \psi(D)$ which has at most two components, and we're done. □

5.4 An alternative technique

In this section, we give our own alternative proof of Theorem 5.3.1. Firstly, because the boundary of Det_n is $\mathrm{GL}(W)$-stable and $\mathrm{GL}(W)$ is connected, the boundary components are necessarily $\mathrm{GL}(W)$-orbit closures of single polynomials $[q]$. The results from Section 5.1, specifically the existence of an $\mathrm{SL}(W)$-invariant f_0 in $I(\partial Det_n)$, give an easy characterization of what those orbits should look like.

Corollary 5.4.1 (Corollary to Lemma 5.1.6)**.** *The irreducible components of ∂Det_n are of codimension one in Det_n.*

Proof. We make use of the following elementary fact:

Claim 5.4.2 ([48], P. 76, Theorem 7)**.** *If f is a surjective map between irreducible varieties $X \to Y$, then any irreducible component of any fiber $f^{-1}(y)$ for $y \in Y$ has dimension at least* $\dim(X) - \dim(Y)$.

Because f_0 is $\mathrm{SL}(W)$-invariant and nonzero, it does not vanish on $\mathrm{GL}(W) \cdot [\det_n]$, so $f_0 \circ \bar{\sigma} : \mathbb{C} \to \mathbb{C}$ is nonzero and thus surjective. We already know $\bar{\sigma}$ is surjective, so f_0 is also surjective and we're done by the above claim. □

Given a polynomial $[q]$ in ∂Det_n, Corollary 5.4.1 gives us an easy way to check whether $\overline{\mathrm{GL}(W) \cdot [q]}$ is a boundary component.

Lemma 5.4.3. *Let $q \in S^n W^*$ be a point on ∂Det_n. Then $\overline{GL(W) \cdot [q]}$ is an irreducible component of ∂Det_n if and only if* $\dim(\mathrm{Ann}_{\mathfrak{gl}(W)})(q) = 2n^2 - 1$.

Proof. Denote $\mathrm{GL}(W)$ by G. By the orbit-stabilizer theorem, $\dim(\overline{G\cdot[q]}) = \dim(G) - \dim(G_q)$. G/G_q is a homogeneous space, so its dimension is equal to $\dim\mathrm{Lie}(G/G_q) = \mathfrak{g}/\mathrm{Ann}_{\mathfrak{g}}(q)$, where $\mathfrak{g} = \mathrm{Lie}(G)$. Because the boundary components of $\overline{\mathrm{GL}(W)\cdot\det_3}$ are of codimension one and $\dim(\mathrm{Ann}_{\mathfrak{g}}(\det_n)) = 2n^2 - 2$, the result follows. ☐

Example 5.4.4. We can show $\overline{\mathrm{GL}(W)\cdot[P_3]}$ is an irreducible component of ∂Det_3 just by showing that $\dim(\mathrm{Ann}_{\mathfrak{gl}(W)})(q) = 17$. Explicitly, pick a basis $\{x_{ij}\}$ for W and take the corresponding basis for $\mathfrak{gl}(W)$ given by $\{x_{ij}\frac{\partial}{\partial x_{k\ell}}\}$. Act on q by each of these basis elements to produce a vector of 81 polynomials; regard each as a point in $\mathbb{C}^{\binom{9+3-1}{3}}$, and $\dim(\mathrm{Ann}_{\mathfrak{g}}(\det_n))$ is merely the dimension of the kernel of the corresponding 81×165 matrix. A quick check in *Macaulay2* gives the desired result.

So broadly speaking, we will show that no $[q]\in\partial Det_n$ gives rise to a new boundary component. Our motivations for taking this approach to Question 5.0.4 are twofold. Firstly, while the reader shall find that the casework-based arguments that will follow are tremendously involved, the primarily combinatorial techniques that arise in the following give hope that after some streamlining of the argument in future iterations of our proof, this approach can be generalized to Det_4 and beyond. Secondly, this approach was originally attempted by Oeding et al. [45], but their proof incorrectly operated under the assumption that only pure polarizations of $\det_3$ could give rise to non-endomorphism components. A conceptual contribution of our work in this section is to show that in fact sums of polarizations can as well.

To carry out our plan, we first need to understand what the condition $[q]\in\partial Det_3$ imposes on q.

5.4.1 Preliminaries

Because Det_n is constructible, its Zariski and Euclidean closures agree. So every polynomial in Det_n is of the form $\lim_{t\to0} g(t)\cdot[\det_n]$ for $g(t)\in\mathrm{GL}_{n^2}(\mathbb{C}[[t]])$ a formal curve. We first show that any such $g(t)$ can be put in a certain normal form. For $U\subset\mathcal{M}_{n\times n}$, let Id_U denote the linear operator which is the identity on U and kills everything in its complement.

Lemma 5.4.5. *For any $g(t)\in GL_n(\mathbb{C}[[t]])$, there exists a direct sum decomposition $W = U_0\oplus\cdots\oplus U_m$ and integers $d_0,\dots,d_m$ for which $\lim_{t\to0} g(t)\cdot[x] = \lim_{t\to0} h(t)\cdot[x]$, where*

$$h(t) = \sum_{i=0}^{m}\mathrm{Id}_{U_i}\,t^{d_i}.$$

Proof. We will show that there exist $a(t)$, $b(t)\in\mathrm{GL}_n(\mathbb{C}[[t]])$ for which $g(t) = a(t)\cdot h(t)\cdot b(t)$, from which the desired result will follow because $\lim_{t\to0} g(t)\cdot[x] = a(0)\cdot(\lim_{t\to0} h'(t)\cdot[x]) = \lim_{t\to0} h'(t)\cdot[x]$, where $h'(t) = \sum_{i=0}^m \mathrm{Id}_{U_i'}\,t^{d_i}$ for $U_i' = b(0)\cdot U\cdot b^{-1}(0)$.

To find the desired $a(t), b(t)$ for which $g(t)$ as $\sum_{i=0}^m g_i t^{d_i'}$. Take $d_0 = d_0'$ and let r_0 denote the rank of g_0. Pick $a, b\in\mathrm{GL}_n(\mathbb{C})$ for which $a\cdot g_0\cdot b$ is a diagonal matrix whose diagonal consists of r_0 1's and $n-r_0$ 0's. Then there exist $a'(t), b'(t)\in\mathrm{GL}_n(\mathbb{C}[[t]])$ such that $a'(0) = a$, $b_0' = b$, and $a'(t)\cdot g(t)\cdot b'(t)$ is a block diagonal matrix consisting of the $r_0\times r_0$ identity matrix in one block (corresponding to U_0) and a matrix $g'(t)$ in $\mathrm{GL}_{n-r_0}(\mathbb{C}[[t]])$ with entries of order greater than d_0. We are done by the inductive hypothesis on $g'(t)$. ☐

Now note that if we identify each space U_i with a matrix of indeterminates (which we will also call U_i), then

$$g(t)\cdot x = \det\left(\sum_{i=0}^{m} U_i\cdot t^{d_i}\right), \tag{5.5}$$

which we can describe in terms of polarizations of the determinant polynomial.

Definition 5.4.6. For $f\in S^dW^*$ and partition $\pi = (\pi_0,\dots,\pi_m)$ of d, where $\pi_0\geq\cdots\geq\pi_m\geq0$, the *polarization* of f with respect to π, denoted by $\partial_\pi f(x_1,\dots,x_m)$ is the coefficient of $t_1^{\pi_1}\cdots t_m^{\pi_m}$ in the expansion of $f(t_1x_1+\cdots+t_mx_m)$ as a polynomial in $t_1,\dots,t_\ell$.

For notational convenience, if S denotes the set of indices i for which $\pi_i\neq0$, then we will sometimes alternatively write the polarization of f with respect to π as $\partial_\pi f(x_{S_1},\dots,x_{S_k})$.

For our purposes, f will be the determinant polynomial, so we suppress f in the above notation and denote the polarization of the determinant by ∂_π. Furthermore, we will take the arguments of ∂_π to be matrices of indeterminates corresponding to subspaces U_i. The polarization ∂_π then has the additional interpretation as

$$\partial_\pi(U_0, ..., U_m) = \sum_{S_0 \sqcup \cdots \sqcup S_m = [m] : |S_i| = \pi_i \forall i} \det(\mathbf{U}|\mathbf{S}), \tag{5.6}$$

where $\det(\mathbf{U}|\mathbf{S})$ denotes the matrix whose ith column is the ith column of U_j if $i \in S_j$. Because the determinant polynomial is multilinear in the columns, we conclude that for $g(t)$ defined above,

$$g(t) \cdot x = \sum_{\pi = (\pi_0, ..., \pi_m)} \partial_\pi(U_0, t \cdot U_1, ..., t^m \cdot U_m). \tag{5.7}$$

Example 5.4.7. Let U_0 be the space of skew-symmetric matrices $\begin{pmatrix} 0 & a & -b \\ -a & 0 & c \\ b & -c & 0 \end{pmatrix}$ and U_1 the space of symmetric matrices $\begin{pmatrix} d & e & f \\ e & g & h \\ f & h & i \end{pmatrix}$ so that $\mathcal{M}_{3\times 3}(\mathbb{C}) = U_0 \oplus U_1$. Then if $g(t) = \mathrm{Id}_{U_0} + t \cdot \mathrm{Id}_{U_1}$, then

$$g(t) \cdot x = \partial_3(U_0) + t \cdot \partial_{2,1}(U_0, U_1) + O(t^2).$$

$\partial_3(U_1) = \det(U_0) = 0$, while

$$\partial_{2,1}(U_0, U_1) = \begin{vmatrix} 0 & e & f \\ -a & g & h \\ b & h & i \end{vmatrix} + \begin{vmatrix} d & a & f \\ e & 0 & h \\ f & -c & i \end{vmatrix} + \begin{vmatrix} 0 & e & f \\ -a & g & h \\ b & h & i \end{vmatrix}.$$

We want to restrict the range of U_i that we need to consider. If U_0 contains a full-rank matrix, then $\lim_{t \to 0} g(t) \cdot x = g_0 \cdot x \in \mathrm{End}(W) \cdot x$, so we instead assume that $\det_n$ vanishes on U_0 (note that this gives yet another connection between Question 5.2.1 and Question 5.0.4).

Henceforth we will take $n = 3$. To prove Theorem 5.3.1, we proceed by casework on the number of subspaces U_i involved. For two subspaces, we show that Example 5.4.7 gives rise to the only boundary component other than $\mathrm{End}(W) \cdot \det_n \backslash \mathrm{GL}(W) \cdot \det_n$. For three or more subspaces, we first assume that U_0 and U_1 are such that $\partial_{2,1}(U_0, V)$ and $\partial_{1,1,1}(U_0, U_1, V)$ are not identically zero. We show that this is sufficiently restrictive to force any such $g(t) \cdot x$ to lie in one of the two known boundary components. We then show how to lift these assumptions to obtain Theorem 5.3.1 in general.

The following is a fact that we will implicitly and explicitly use in reasoning about orbit closures lying in the endomorphism component.

Observation 7. Suppose U_i appears in polarizations $q_1, q_2, ..., q_k$ in q and the vanishing of all these polarizations imposes m unique linear conditions on U_i. Then if $\dim(U_i) > m$, q is degenerate.

Proof. In this case, we could write q as $\ell_1 p_1 + \cdots \ell_m p_m$ for quadrics $p_1, ..., p_m$ and linear forms $\ell_1, ..., \ell_m$ in the entries of U_i. But then the linear span of $\left\{ \frac{\partial}{\partial_{x_{ij}}} q \right\}$ has dimension at most $m < \dim(U_i)$, so q is necessarily degenerate. □

5.4.2 Two subspaces

Suppose $W = U_0 \oplus U_1$, and suppose U_0 is not a $\mathrm{GL}(W)_{\det_n}$-translate of W_{skew}. By Corollary 5.2.11, we can assume that U_0 is a compression space. Furthermore, obviously q can only contain one polarization, either $\partial_{2,1}(U_0, U_1)$ or $\partial_{1,2}(U_0, U_1)$. If $\partial_{2,1}(U_0, U_1)$ does not vanish, then $\mathrm{rank}(U_0) = 2$, and q is precisely the determinant of $\left((U_0)_1 \quad (U_0)_2 \quad (U_1)_3 \right)$ if the first two columns of U_0 are linearly independent. In this case, q is either degenerate or a $\mathrm{GL}(W)$-translate of $\det_3$. If $\partial_{2,1}(U_0, U_1)$ does vanish so that $\mathrm{rank}(U_0) = 1$, then $q = \partial_{1,2}(U_0, U_1)$ is precisely the determinant of $\left((U_0)_1 \quad (U_0)_2 \quad (U_1)_3 \right)$ if the first column of U_0 is nonzero. Again, q is either degenerate or a $\mathrm{GL}(W)$-translate of $\det_3$.

5.4.3 Three or more subspaces

The coefficient of the lowest order term in the expansion of $g(t)\cdot x$ is some sum of polarizations $q = \sum_{j=1}^{\ell} \partial_{\pi^j}(U_0,...,U_m)$. Henceforth, we will write the arguments to polarizations in increasing order of the indices of the subspaces, e.g. if $\partial_{1,1,1}(U_i,U_j,U_k)$ appears in q, then $i\le j\le k$. Because q is the lowest-order term in the expansion of the determinant of some formal curve, the following trivially holds:

Observation 8. If $\partial_{1,1,1}(U_i,U_j,U_k)$ and $\partial_{1,1,1}(U_{i'},U_{j'},U_{k'})$ appear in q, it cannot be the case that $(i,j,k)\le(i',j',k')$ or $(i,j,k)\ge(i',j',k')$. Moreover, if (a,b,c) is a triple for which $(a,b,c)<(i,j,k)$ for some $\partial_{1,1,1}(U_i,U_j,U_k)$ appearing in q, then $\partial_{1,1,1}(U_a,U_b,U_c)=0$.

Throughout, we will be using the vanishing of such $\partial_{1,1,1}(U_a,U_b,U_c)$ to conclude that U_a, U_b, U_c satisfy certain linear conditions.

Remark 5.4.8. In what follows, we will often say that the vanishing of a polarization, e.g. $\partial_{1,1,1}(U_i,U_j,V)$, imposes certain conditions on V. Here V should not be interpreted as any particular space of matrices; rather, we are saying that the map $\phi: W\to\mathbb{C}$ given by $V\mapsto\partial_{1,1,1}(U_i,U_j,V)$ has a kernel defined by certain linear conditions.

In order for q not to be degenerate, we must have that $\bigcup_{j=1}^{\ell} S^j=\{U_0,...,U_m\}$, where S^j denotes the set of indices i for which $\pi_i^j\ne0$. So in order for q to be non-degenerate, it certainly cannot contain the terms $\partial_{2,1}(U_0,U_i)$ for $i<m$, so we may assume these to vanish.

For now, assume that $\partial_{2,1}(U_0,U_m)$ does not vanish and thus belongs among the polarizations in q. We note that as in Section 5.4.2, we may assume U_0 is a compression space. Otherwise, if $U_0=W_{skew}$, the vanishing of $\partial_{2,1}(U_0,V)$ for $V=U_0,...,U_{m-1}$ imposes too many conditions on V.

Observation 9. $\partial_{2,1}(W_{skew},V)=0$ if and only if $V\subset W_{skew}$.

As before, corollary 5.2.11 thus tells us U_0 must be a compression space. We now proceed by casework depending on whether $U_0\subset U_1^{cmp}$ or $U_0\subset U_2^{cmp}$ (recall that these were defined in Section 5.3).

5.4.4 $U_0\subset U_1^{cmp}$

As in Observation 9, we can check that $\partial_{2,1}(U_1^{cmp},V)=0$ if and only if $V\subset U_1^{cmp}$, which would impose three conditions on V. But then, apart from $\partial_{2,1}(U_0,U_m)$, every other polarization appearing in q would vanish because $U_0,...,U_{m-1}\subset U_1^{cmp}$. It follows that U_0 must be a small enough subspace of U_1^{cmp} that this does not hold. It is straightforward but tedious to verify the following.

Lemma 5.4.9. *If $U_0\subset U_1^{cmp}$ and the vanishing of $\partial_{2,1}(U_0,V)$ imposes exactly two linear conditions on W, then U_0 is a G_{det}-translate of a subspace of $U_1^{cmp}\cap U_2^{cmp}$.*

In this case, the vanishing of $\partial_{2,1}(U_0,V)$ still imposes enough conditions to force q to contain few terms.

Lemma 5.4.10. *If $U_0\subset U_1^{cmp}$ is such that $\partial_{2,1}(U_0,V)=0$ imposes exactly two linear conditions on V, then q consists solely of $\partial_{2,1}(U_0,U_m)$ and possibly $\partial_{1,1,1}(U_0,U_1,U_{m-1})$.*

Proof. By Lemma 5.4.9, we can assume that the vanishing of $\partial_{2,1}(U_0,V)$ forces entries $V_3^1=V_3^2=0$ for $V=U_0,...,U_{m-1}$. If the polarization $\partial_{1,1,1}(U_i,U_j,U_{m-1})$ appears in q, $\partial_{1,1,1}(U_0,U_1,U_k)=0$ for all $0\le k<m-1$. $\partial_{1,2}(U_0,V)$ can be written as

$$\partial_{1,2}(U_0,V)=\begin{vmatrix}(U_0)_1^1 & V_2^1\\ (U_0)_1^2 & V_2^2\end{vmatrix}\cdot V_3^3.$$

This does not factor unless there is a linear dependence between $(U_0)_1^1$ and $(U_0)_1^2$, but that would contradict the fact that $\partial_{2,1}(U_0,V)=0$ imposes two linear conditions on V. We conclude that if $\partial_{2,1}(U_0,V)=0$ imposes exactly two linear conditions on V, then either $U_0,...,U_{m-1}\subset U_1^{cmp}$ or $U_0,...,U_{m-1}\subset U_2^{cmp}$, giving the desired conclusion about terms that appear in q. It is easy to see then that q is either degenerate or a $\mathrm{GL}(W)$-translate of $\det_3$. ☐

The following is again straightforward but tedious to check.

Lemma 5.4.11. *If $U_0 \subset U_1^{cmp}$ and the vanishing of $\partial_{2,1}(U_0, V)$ imposes exactly one linear condition on W, then U_0 is a G_{det}-translate of a subspace of*

$$U^{sq} := \begin{pmatrix} a & b & 0 \\ c & d & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

.

We can assume $U_0 \subset U^{sq}$, in which case all V for which $\partial_{2,1}(U_0, V) = 0$ satisfy $V_3^3 = 0$. For now, assume that $\partial_{1,1,1}(U_0, U_1, V)$ is not identically zero over all such V. We proceed by casework on $\dim(U_0)$, noting that the difficulty of proof increases quite dramatically as $\dim(U_0)$ decreases.

Lemma 5.4.12. *If $U_0 = U_0^3$, then q consists solely of $\partial_{2,1}(U_0, U_m)$ and possibly $\partial_{1,1,1}(U_0, U_1, U_{m-1})$.*

Proof. When $U_0 = U_0^3$, $\partial_{1,2}(U_0^3, V)$ can be written as $-aV_3^2V_2^3 + bV_3^2V_1^3 + cV_3^1V_2^3 - dV_3^1V_1^3$, which vanishes if and only if either $x_{13} = x_{23} = 0$ or $x_{31} = x_{32} = 0$. In either case, $\partial_{1,1,1}(U_0, U_1, V) = 0$ would impose the same conditions on V and thus force an entire row or column of V to vanish, from which we would conclude that q only contains the polarizations $\partial_{2,1}(U_0, U_m)$ and possibly $\partial_{1,1,1}(U_0, U_1, U_{m-1})$. □

Lemma 5.4.13. *If $\dim(U_0) = 1$ and q is non-degenerate, then q can only contain the terms $\partial_{2,1}(U_0, U_m)$, $\partial_{1,1,1}(U_0, U_1, U_{m-1})$, $\partial_{1,1,1}(U_0, U_i, U_{m-2})$, and $\partial_{1,1,1}(U_1, U_1, U_{m-2})$.*

Proof. The result for $\dim(U_0) = 2$ implies the same for $\dim(U_0) = 3$, so we will assume the former holds. After row-reducing, we may assume that

$$U_0 = \begin{pmatrix} a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

so that $\partial_{1,2}(U_0, V) = -aV_3^2V_2^3 - bV_3^1V_1^3$. The vanishing of $\partial_{1,2}(U_0, U_1)$ forces one of $(U_1)_3^2$ or $(U_1)_2^3$ to vanish, and one of $(U_1)_3^1$ or $(U_1)_1^3$ to vanish. If U_1 has a row or column of zeros, the vanishing of $\partial_{1,1,1}(U_0, U_1, U_k)$ for $k < m-1$ forces U_k to have the same, and the claim follows.

By the same reasoning, if instead U_1 satisfies $(U_1)_3^1 = (U_1)_2^3 = 0$, then U_k satisfies these conditions for $k < m-1$, so q cannot contain any other polarization of the form $\partial_{1,1,1}(U_0, U_j, U_k)$. Lastly, suppose to the contrary that $\partial_{1,1,1}(U_i, U_j, U_k)$ appears in q for $j \geq 2$ and $i \geq 1$. Then $\det(U_1) = 0$ means either $(U_1)_1^3$ or $(U_1)_3^2$ vanishes, or U_1 satisfies some third linear condition. If the former, U_1 has a row or column of zeros. If the latter, the vanishing of $\partial_{2,1}(U_1, U_i)$, $\partial_{2,1}(U_1, U_j)$, and $\partial_{2,1}(U_1, U_k)$ forces U_i, U_j, U_k to satisfy this third condition, so $\partial_{1,1,1}(U_i, U_j, U_k) = 0$, a contradiction. □

The remainder of this subsection will be dedicated to resolving the last case of $\dim(U_0) = 1$, in which we may assume after column/row reducing that

$$U_0 = \begin{pmatrix} a & 0 & 0 \\ 0 & \lambda a & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

Assume that the remaining spaces $U_1, ..., U_m$ lie in the complementary space of matrices for which the top-left entry is zero.

Theorem 5.4.14. *If $\dim(U_0) = 1$ and q is non-degenerate, then q gives rise to no new boundary components. In particular, q at most contains the terms $\partial_{2,1}(U_0, U_m)$, $\partial_{1,1,1}(U_0, U_1, U_{m-1})$, $\partial_{1,1,1}(U_0, U_i, U_{m-2})$, $\partial_{1,1,1}(U_1, U_1, U_{m-2})$, and $\partial_{1,1,1}(U_1, U_j, U_k)$ for some i, j, k.*

Proof. Intuitively, the argument is more involved in part because the vanishing of $\partial_{1,1,1}(U_0, U_1, V)$ and $\partial_{2,1}(U_0, V)$ will not guarantee the vanishing of $\partial_{1,2}(U_0, V)$ and $\det(V)$. But as we will show, that our general method used so far must be executed more carefully is actually to be expected, because q can in fact be non-degenerate and give rise to a boundary component.

Note that

$$\partial_{1,2}(U_0, U_1) = -a(\lambda(U_1)_3^1(U_1)_1^3 + (U_1)_3^2(U_1)_2^3) = -a \begin{vmatrix} \lambda(U_1)_3^2 & (U_1)_3^1 \\ (U_1)_1^3 & -(U_1)_2^3 \end{vmatrix},$$

and the maximal linear subspaces of this 2×2 determinant are all of codimension 2. We therefore need to consider the following cases: 1) there are linear dependencies between $(U_1)^2_3$ and $(U_1)^3_1$ and between $(U_1)^3_2$ and $(U_1)^1_3$, 2) there are linear dependencies between $(U_1)^2_3$ and $(U_1)^1_3$ and between $(U_1)^3_2$ and $(U_1)^3_1$. Up to simultaneous row/column operations on U_0 and U_1, we may assume that in 1), either $(U_1)^2_3 = (U_1)^3_1 = 0$, while in 2), $(U_1)^2_3 = (U_1)^1_3 = 0$.

Case 1. There are linear dependencies between $(U_1)^2_3$ and $(U_1)^1_3$ and between $(U_1)^3_2$ and $(U_1)^3_1$.

Up to simultaneous row/column operations on $U_0, ..., U_m$, we may assume $(U_1)^2_3 = (U_1)^3_1 = 0$.

Lemma 5.4.15. *q contains at most one polarization of the form $\partial_{1,1,1}(U_1, U_i, U_j)$ for $i \geq 2$.*

Proof. If q contains such a polarization, then we know $\det(U_1) = 0$, so in addition to the vanishing of $(U_1)^2_3$ and $(U_1)^3_1$, U_1 must satisfy a third linear condition: $\det(U_1)$ has three linear factors corresponding to the vanishing of $(U_1)^3_2$, $(U_1)^1_3$, and $(U_1)^2_1$.

Subcase 1. $(U_1)^3_2 = 0$ (the case of $(U_1)^1_3 = 0$ is identical).

The vanishing of $\partial_{1,1,1}(U_0, U_1, V)$ and $\partial_{2,1}(U_1, V)$ for $V = U_0, ..., U_j$ forces at least one of V^3_1, V^3_2 to vanish. If both must vanish, $\partial_{1,1,1}(U_1, U_i, U_j) = 0$, a contradiction. So the vanishing of $\partial_{2,1}(U_1, V)$ must impose the same single condition on V as that of $\partial_{1,1,1}(U_0, U_1, V)$, namely $V^3_1 = 0$. But we note that

$$\partial_{2,1}(U_1, V) = (U_1)^1_3 \cdot \begin{pmatrix} (U_1)^2_1 & (U_1)^2_2 \\ V^3_1 & V^3_2 \end{pmatrix},$$

so for $\partial_{2,1}(U_1, V) = 0$ to force V^3_2 to vanish as well, we must have that $(U_1)^2_1 = 0$. The vanishing of $\partial_{1,2}(U_0, U_i)$ and $\partial_{1,2}(U_1, U_i)$ further imposes either the single condition of $(U_i)^3_2 = 0$ or the conditions of $(U_i)^2_1 = (U_i)^2_3 = 0$. Consequently, $\partial_{1,1,1}(U_1, U_i, U_j)$ factors as either $(U_1)^1_3(U_i)^2_1(U_j)^3_2$ if $(U_i)^3_2 = 0$, or $(U_1)^1_3(U_i)^3_2(U_j)^2_1$ if $(U_i)^2_1 = (U_i)^2_3 = 0$.

Returning to the original claim, suppose q contains an additional term $\partial_{1,1,1}(U_1, U_{i'}, U_{j'})$ for $i' > i$. Then $\partial_{1,1,1}(U_1, U_i, U_{i'}) = \partial_{1,2}(U_1, U_i) = \partial 1, 2(U_1, U_{i'}) = 0$ implies that $V^3_2 = 0$ or $V^2_1 = V^2_3 = 0$ for $V = U_i, U_{i'}$. But then the vanishing of $\partial_{1,1,1}(U_1, U_i, U_{j'})$ forces at least one of $\partial_{1,1,1}(U_1, U_i, U_j)$ and $\partial_{1,1,1}(U_1, U_{i'}, U_{j'})$ to vanish, as desired.

Subcase 2. $(U_1)^2_1 = 0$.

The vanishing of $\partial_{1,1,1}(U_0, U_1, V)$ and $\partial_{2,1}(U_1, V)$ for $V = U_0, ..., U_j$ forces at least one of V^2_3, V^3_1, V^2_1 to vanish. If all three must, then $\partial_{1,1,1}(U_1, U_i, U_j) = 0$, a contradiction. So the vanishing of $\partial_{1,1,1}(U_0, U_1, V)$ and the vanishing of $\partial_{2,1}(U_1, V)$ must each impose a single condition. Let

$$H = \begin{pmatrix} 0 & H^1_2 & H^1_3 \\ 0 & H^2_2 & 0 \\ 0 & H^3_2 & 0 \end{pmatrix}$$

be the space containing U_1. Then $\partial_{1,1,1}(U_0, H, V) = -a(\lambda H^1_3 V^3_1 + H^3_2 V^2_3)$ while $\partial_{2,1}(H, V) = -H^1_3(H^2_2 V^3_1 - H^3_2 V^2_1)$, so in order for the vanishing of these to impose one condition each on V, there must exist linear dependencies ℓ_1, ℓ_2 between H^1_3 and H^3_2 and between H^2_2 and H^3_2 respectively. $H^3_2 \neq 0$ or we reduce to the previous subcase, so U_1 must satisfy two additional linear conditions. In particular, $\dim(U_1) \leq 2$.

Suppose that the vanishing of $\partial_{1,1,1}(U_0, H, V)$ and $\partial_{2,1}(H, V)$ imposes conditions $V^2_3 = \mu_1 V^3_1$ and $V^2_1 = \mu_2 V^3_1$ (we may assume that $V^3_1 \neq 0$ or else $\partial_{1,1,1}(U_1, U_i, U_j) = 0$). Denoting by $V_{ker} \subset (U_0)^c$ the space containing U_1 which is cut out by these conditions and the vanishing of $(V_{ker})^3_3$, note that $\partial_{1,2}(U_0, V_{ker})$ factorizes into the single variable of U_0 and the forms $(V_{ker})^3_1$ and ℓ_1.

So if V_{ker} additionally satisfies $\ell_1 = 0$, then $\partial_{1,2}(U_1, V_{ker})$ has a factor of $(U_1)^3_1$, and its other factor is of the form

$$\lambda(V_{ker})^1_2(U_1)^3_1 + (V_{ker})^3_2\ell_2. \tag{5.8}$$

Returning to the original claim, suppose q contains another term of the form $\partial_{1,1,1}(U_1, U_{i'}, U_{j'})$ for $i' > i$.

If $(V_{ker})^1_2 \neq 0$ in U_1, then by the above, $U_{i'}$ must satisfy $(U_{i'})^3_1 = (U_{i'})^3_2 = 0$. Replacing U_1 by $U_{i'}$ in the above analysis, we reduce to the previous subcase. Alternatively, if $(V_{ker})^1_2 = 0$, then $\dim(U_1) = 1$ and $U_{i'}$ satisfies $\ell_2 = 0$, in addition to $(U_{i'})^2_3 - \mu_1(U_{i'})^3_1 = (U_{i'})^2_1 - \mu_2(U_{i'})^3_1 = \ell_1 = 0$. The subspace V_{max} cut out by these conditions is such that $\partial_{1,2}(U_1, V_{max}) = 0$. On the other hand, we know by the above that U_i satisfies all but possibly ℓ_2 among these conditions; the subspace $V_{max} \subset V$ cut out by these conditions is such that $\partial_{1,1,1}(U_1, U_{i'}, V)$ has a factor in V of ℓ_2, because $\partial_{1,2}(U_1, U_{i'}) = 0$. U_i thus also satisfies $\ell_2 = 0$.

One can check that $\partial_{1,1,1}(U_1, V_{max}, V_{ker}) = (U_1)^3_2 (U_{max})^3_1 \ell_2$. So because $\partial_{1,1,1}(U_1, U_{i'}, U_{j'}) \neq 0$ while $\partial_{1,1,1}(U_1, U_{i'}, U_j) = 0$, U_j must satisfy $\ell_2 = 0$, in which case $\partial_{1,1,1}(U_1, U_i, U_j) = 0$, a contradiction. □

The last step of the above proof implies the following:

Corollary 5.4.16. *If q contains the polarization $\partial_{1,1,1}(U_1, U_i, U_j)$ for some $i \geq 2$, then $V = U_2, ..., U_{i-1}$ satisfy $V^3_1 = 0$.*

Lemma 5.4.17. *If q is non-degenerate, then $\partial_{1,1,1}(U_1, U_i, U_j)$ cannot appear in q if $i \geq 2$ and $j > 2$.*

Proof. In subcase 1, polarizations of the form $\partial_{2,1}(U_1, U_i)$ cannot appear in q. Otherwise, $\partial_{1,1,1}(U_0, U_1, U_i) = 0$, and because we are assuming $\partial_{1,1,1}(U_0, U_1, V) = 0$ and $\partial_{2,1}(U_1, V) = 0$ impose the same single condition on V, there are at most four terms in q, namely $\partial_{2,1}(U_0, U_m)$, $\partial_{1,1,1}(U_0, U_1, U_{m-1})$, $\partial_{1,1,1}(U_0, U_i, U_j)$ and $\partial_{1,1,1}(U_1, U_{i'}, U_{j'})$.

All polarizations that appear in q and involve U_1 have a linear factor in U_1, namely the condition $(U_1)^1_3 = 0$. If $\partial_{1,1,1}(U_0, U_i, U_j)$ appears in q for $i > 1$, it has a factor of $(U_i)^2_3$ and thus a factor in U_j, even if $i = j$. Likewise, if $\partial_{1,1,1}(U_1, U_{i'}, U_{j'})$ appears in q for $i > 1$, it also decomposes into three linear factors, as we saw in subcase 1.

If $i = j$, then the collection of $\frac{\partial}{\partial x_{ij}} \partial_{1,1,1}(U_0, U_i, U_j)$ has dimension at most 2 because $\partial_{1,1,1}(U_0, U_i, U_j)$ factors. In order for q to not be degenerate, we conclude that $\dim(U_j) + \dim(U_{j'}) + \dim(U_{m-1}) + \dim(U_m) \leq 2 + 1 + 1 + 1 = 5$, with equality holding iff $i = j = j'$ (in which case $\partial_{1,1,1}(U_1, U_{i'}, U_{j'})$ and $\partial_{1,1,1}(U_0, U_i, U_j)$ share a common factor in U_i). On the other hand, $\dim(U_0) + \dim(U_1) + \dim(U_i) + \dim(U_{i'}) \leq 1 + 1 + 1 + 1 = 4$. We conclude that $U_0 \oplus \cdots \oplus U_m \neq W$, a contradiction.

The argument is essentially the same for subcase 2, but we know additionally that $i = i'$ by Corollary 5.4.16. Furthermore, unlike in subcase 1, q may contain $\partial_{2,1}(U_1, U_k)$ for some k. By Observation 8 and our assumptions by definition that $j' > i'$, $j > i$, it follows that $i = i' < j' < j, k < m-1 < m$. Furthermore, in order for the polarizations in q to contain all subspaces $U_0, ..., U_m$, i must be 2.

But all polarizations in q have a linear factor in the highest-indexed space. $\partial_{2,1}(U_0, V) = 0$ imposes exactly one condition on V, so $\partial_{2,1}(U_0, U_m)$ has a linear factor in U_m. $\partial_{1,1,1}(U_0, U_1, V) = 0$ imposes exactly one condition, so $\partial_{1,1,1}(U_0, U_1, U_{m-1})$ has a linear factor in U_{m-1}. By the same argument, $\partial_{2,1}(U_1, U_k)$ has a linear factor in U_k. $\partial_{1,1,1}(U_0, U_i, U_j)$ and $\partial_{1,1,1}(U_1, U_{i'}, U_{j'})$ have the same linear factor of V_{31} in $V = U_i, U_{i'}$. Because $\dim(U_0) = \dim(U_1) = 1$ in subcase 2, $\partial_{1,1,1}(U_0, U_i, U_j)$ and $\partial_{1,1,1}(U_1, U_{i'})$ in fact factorize completely. We conclude that $\dim(U_0), ..., \dim(U_m) = 1$, so $m = 8$. But in this case, at least one subspace will be missing from among the polarizations in q, a contradiction. □

Claim 5.4.18. *If q contains no $\partial_{1,1,1}(U_1, U_i, U_j)$ term, q must be degenerate.*

Proof. In subcase 1, $\partial_{2,1}(U_1, U_k) = 0$, so each of $\partial_{2,1}(U_0, U_m)$, $\partial_{1,1,1}(U_0, U_1, U_{m-1})$, and $\partial_{1,1,1}(U_0, U_i, U_j)$ completely factors, so for $U_0 \oplus \cdots \oplus U_m = W$ to hold, q would have to be degenerate.

In subcase 2, if $\partial_{1,1,1}(U_0, U_1, V) = 0$ imposes two conditions, namely $V^2_3 = V^3_1 = 0$, on $V = U_2, ..., U_{m-2}$, then $\partial_{1,1,1}(U_0, U_i, U_j) = 0$. Moreover, $\partial_{2,1}(U_1, U_k)$ decomposes into factors of $(U_1)^1_3$, $(U_1)^3_2$, and $(U_k)^2_1$, while the collection of $\frac{\partial}{\partial x_{ij}} \partial_{1,1,1}(U_0, U_1, U_{m-1})$ has dimension two, so $\dim(U_1) \leq 3$ and $\dim(U_{m-1}) \leq 2$, while $\dim(U_0) = \dim(U_m) = 1$, so $U_0 \oplus \cdots \oplus U_m \neq W$.

Recall that for $\partial_{1,1,1}(U_0, U_1, V) = 0$ to impose exactly one condition $V^2_3 - \mu_1 V^3_1 = 0$, U_1 must satisfy some additional ℓ_1. But then $\partial_{1,1,1}(U_0, U_1, U_{m-1})$ completely factors, with a factor of $(U_1)^3_2$. $\partial_{1,1,1}(U_0, U_i, U_j)$ also completely factors, while $\partial_{2,1}(U_0, U_m)$ factors because $\dim(U_0) = 1$ and q is non-degenerate. For $V \supset U_k$ satisfying $x_{23} - \mu_1 x_{31}$, $\partial_{2,1}(U_1, V) = 0$ imposes at most two conditions on V, namely $V^3_2 = V^1_3 = 0$. But $\partial_{2,1}(U_1, U_k)$ also has a factor of $(U_1)^3_2$ in U_1, so $\dim(U_1) \leq 2$. In conclusion, $\dim(U_k) + \dim(U_j) + \dim(U_{m-1}) + \dim(U_m) = 4$, so $\dim(U_0) + \cdots + \dim(U_m) \leq 8$, a contradiction. □

Claim 5.4.19. *If q contains $\partial_{1,2}(U_1, U_2)$ and is non-degenerate, then $\overline{GL(W) \cdot [q]} = \overline{GL(W) \cdot [P_3]}$.*

Proof. For subcase 2, as in the proof of Lemma 5.4.17, q only contains terms from $\partial_{2,1}(U_0, U_m)$, $\partial_{1,1,1}(U_0, U_1, U_{m-1})$, $\partial_{1,1,1}(U_0, U_2, U_i)$, $\partial_{2,1}(U_1, U_j)$, and $\partial_{1,1,1}(U_1, U_2, U_2)$. We know the first three of these factorize completely, with a factor of $(U_1)^3_2$ for $\partial_{1,1,1}(U_0, U_1, U_{m-1})$ and $(U_2)^3_1$ for $\partial_{1,1,1}(U_0, U_2, U_i)$. Likewise, for $\partial_{2,1}(U_1, U_j)$, $V_1 \subset V_{max}$, and one can check that $\partial_{2,1}(V_{max}, V)$ has factors $V^2_1 - \mu V^3_1$ and $((V_{max})^3_2)^2$. Lastly, recall from the analysis of subcase 2 in Lemma 5.4.15 that $\partial_{1,2}(U_1, U_2)$ has a factor of $(U_2)^3_1$, and its other factor is of the form $\lambda(U_1)^1_2(U_2)^3_1 + (U_1)^3_2\ell_2$ (see (5.8)).

So $\dim(U_1) \le 2$, $\dim(U_2) \le 2$, and $\dim(U_3) + \cdots + \dim(U_m) \le 4$. so for $U_0, ..., U_m$ to span $\mathbb{C}^9$, these inequalities must be equalities. Consider the decompositions $U_1 = A_1 \oplus B_1$ and $U_2 = A_2 \oplus B_2$, where A_1 and A_2 are spanned by $(U_1)^3_2$ and $(U_2)^3_1$ respectively, and B_1 and B_2 are some complements in U_1 and U_2. What we have checked above is that $\partial_{1,1,1}(U_0, U_1, U_{m-1}) = \partial_{1,1,1}(U_0, U^1_1, U_{m-1})$ and $\partial_{1,1,1}(U_0, U_2, U_i) = \partial_{1,1,1}(U_0, U^1_2, U_i)$.

So consider $q' = \partial_{2,1}(U_0 \oplus A_1 \oplus A_2, B_1 \oplus B_2 \oplus U_3 \oplus \cdots \oplus U_m)$, which we now know is merely a $\mathrm{GL}(W)$-translate of q. This yields one of the two known boundary components; in particular, one can check that $U_0 \oplus U_1 \oplus U_2$ is in fact a G_{det}-translate of W_{skew}, as claimed. The analysis for subcase 1 is similar. □

Finally, suppose that the following Case 2 holds but Case 1 does not.

Case 2. There are linear dependencies ℓ_1, ℓ_2 between $(U_1)^2_3$ and $(U_1)^3_1$, and between $(U_1)^3_2$ and $(U_1)^1_3$ respectively.

If q contains $\partial_{1,1,1}(U_0, U_i, U_j)$ for some $i > 1$, then we know that $\partial_{1,1,1}(U_0, U_1, V) = 0$ for $V = U_i, U_j$, so one of ℓ_1, ℓ_2, or V^3_3 must vanish in V. If all three do, then $\partial_{1,1,1}(U_0, U_i, U_j) = 0$, a contradiction. But if the vanishing of $\partial_{1,1,1}(U_0, U_1, V)$ were to impose only two linear conditions on V, namely $V^3_3 = \mu\ell_1 + \nu\ell_2 = 0$ for some scalars μ, ν, there would have to be a linear dependency between $(U_1)^2_3$ and $(U_1)^3_1$, in which case we reduce to Case 1.

If q contains $\partial_{1,1,1}(U_i, U_j, U_k)$ for some $i \ge 1$, $j > 1$, we will argue in a similar way that such a polarization would have to vanish. It is not hard to see that if there are no linear relations between $(U_1)^1_3$ and $(U_1)^2_3$ and between $(U_1)^3_1$ and $(U_1)^3_2$, then in addition to $ell_1 = \ell_2 = 0$, U_1 satisfies three other linear conditions $m_1 = m_2 = m_3 = 0$. We may assume $\dim(U_1) > 1$, or else we obviously reduce to the previous case. We know that $\partial_{2,1}(U_1, V) = 0$ for $V = U_i, U_j, U_k$, so if V satisfied all $m_1 = m_2 = m_3 = 0$, then we are done. Indeed, the coefficients of $((U_1)^1_3)^2$, $((U_1)^2_3)^2$, and $(U_1)^1_3(U_1)^2_3$ in $\partial_{2,1}(U_1, V)$ for V satisfying $\ell_1 = \ell_2$ cannot be linearly dependent linear forms because exactly one contains each of V^2_1, V^1_2, and V^2_2, respectively. We conclude that $\partial_{1,1,1}(U_i, U_j, U_k)$ would vanish in q, a contradiction.

So q can only consist of some combination of $\partial_{2,1}(U_0, U_m)$, $\partial_{1,1,1}(U_0, U_1, U_{m-1})$, and $\partial_{2,1}(U_1, U_{m-2})$. For q to be non-degenerate, we must have $m \le 4$. By Lemma 7, $\dim(U_{m-2}) \le 3$, $\dim(U_{m-1}) \le 2$, and $\dim(U_m) = 1$. Because U_1 satisfies $(U_1)^3_3 = (U_1)^3_2 = ell_1 = \ell_2 = m_1 = m_2 = m_3 = 0$, $\dim(U_1) \le 2$. In order for $U_0, ..., U_m$ to span $\mathbb{C}^9$, all these inequalities must be equalities. Now consider $q' = \partial_{2,1}(U_0 \oplus U_1, U_{m-2} \oplus U_{m-1} \oplus U_m)$. q and q' are in the same $\mathrm{GL}(W)$ orbit, so again q gives rise to no new boundary components, though as in Claim 5.4.19, it may be the case that $\overline{\mathrm{GL}(W)\cdot[q]} = \overline{\mathrm{GL}(W)\cdot[P_3]}$. □

We now drop our assumption that $\partial_{1,1,1}(U_0, U_1, V)$ is not identically zero over all V for which $\partial_{2,1}(U_0, V) = 0$. This is the case if and only if $U_1 \subset U^{sq}$, in which case we also have that $\partial_{2,1}(U_1, V) = 0$ is also identically zero. So q must be of the form

$$q = \sum_{\mu=1}^{M} \partial_{1,1,1}(U_0, U_{i_\mu}, U_{j_\mu}) + \sum_{\nu=1}^{N} \partial_{1,1,1}(U_1, U_{k_\nu}, U_{\ell_\nu})$$

for $i_\mu, j_\mu, k_\nu, \ell_\nu > 1$.

Note that

$$q' := \sum_{\mu=1}^{M} \partial_{1,1,1}(U_0 \oplus U_1, U_{i_\mu}, U_{j_\mu}) + \sum_{\nu=1}^{N} \partial_{1,1,1}(U_0 \oplus U_1, U_{k_\nu}, U_{\ell_\nu})$$

differs from q by $\sum_{\nu=1}^{N} \partial_{1,1,1}(U_1, U_{i_\mu}, U_{j\mu})$, so if $\dim(U_1) \le \dim(U_0)$, q' is a $\mathrm{GL}(W)$-translate of q and uses one fewer subspace, so we're done.

The only case we still need to consider is thus when $\dim(U_0) = 1$. To show that $\dim(U_1)$ must be 1, suppose to the contrary and suppose q contains some polarization $\partial_{1,1,1}(U_1, U_i, U_j)$ for $i, j > 1$, and let i here be minimal among all such polarizations.

Case 1. $i = j$.

In this case, $\partial_{1,2}(U_1, U_i, U_i)$ is the only polarization in q which contains U_1. Because of the vanishing of $\partial_{1,2}(U_0, U_i)$, in U_i there are linear dependencies ℓ_1, ℓ_2 either between $(U_i)^2_3$ and $(U_i)^3_1$ and between $(U_i)^3_2$ and $(U_i)^1_3$, or between $(U_i)^3_1$ and $(U_i)^3_2$ and between $(U_i)^1_3$ and $(U_i)^2_3$.

In the former case, $\partial_{1,2}(U_1, U_i)$ would completely factorize, because two of its linear factors would correspond to the vanishing of $(U_i)^1_3, (U_i)^2_3$ and $(U_i)^3_1, (U_i)^3_2$, so $\dim(U_1)$ would have to be 1.

In the latter case, similar to before, we may assume after simultaneous row/column operations on $U_0, ..., U_m$ that in U_i, $(U_i)^2_3 = (U_i)^3_1 = 0$. This is enough for $\partial_{1,2}(U_1, U_i)$ to completely factor because two of its linear factors are $(U_i)^1_3$ and $(U_i)^3_2$.

Case 2. $i < j$.

If $\dim(U_1) > 1$, then $\dim(U_0 \oplus U_1) > 2$, so by Lemma 5.4.12 and the vanishing of $\partial_{1,2}(U_0 \oplus U_1, U_i)$, U_i satisfies either $(U_i)^1_3 = (U_i)^2_3 = 0$ or $(U_i)^3_1 = (U_i)^3_2 = 0$. Suppose q contains another term $\partial_{1,1,1}(U_1, U_{i'}, U_{j'})$ for $i' > i$. The vanishing of $\partial_{1,1,1}(U_0 \oplus U_1, U_i, U_{i'})$ and $\partial_{1,1,1}(U_0 \oplus U_1, U_i, U_{j'})$ forces $U_{i'}$ and $U_{j'}$ to satisfy the same pair of conditions, so $\partial_{1,1,1}(U_1, U_{i'}, U_{j'}) = 0$.

Finally, we show that $\partial_{1,1,1}(U_1, U_i, U_j)$ completely factorizes. If $\partial_{1,1,1}(U_0, U_i, U_j) = 0$ imposes $(U_j)^1_3 = (U_j)^2_3$, then $\partial_{1,1,1}(U_1, U_i, U_j) = 0$. If not, $\partial_{1,1,1}(U_1, U_i, U_j)$ has a linear factor in U_j corresponding to the vanishing of both $(U_j)^1_3$ and $(U_j)^2_3$. Furthermore, in order for $\partial_{1,1,1}(U_0, U_i, U_j) = 0$ to impose a single condition, there must exist a linear dependency $(U_i)^3_1 = \mu \cdot (U_i)^3_2$. But then $\partial_{1,1,1}(U_1, U_i, U_j)$ has a factor of $(U_i)^3_2$, so $\partial_{1,1,1}(U_1, U_i, U_j)$ indeed completely factorizes and $\dim(U_1) = 1$ as desired.

Using similar methods to the above, we show in Appendix A.8 that the assumption $\partial_{2,1}(U_0, W) \neq 0$ can be lifted.

5.4.5 $U_0 \subset U_2^{cmp}$

We assume now that U_0 is a subspace of U_2^{cmp} but not a $\mathrm{GL}(W)_{\det_3}$-translate of a subspace of U_1^{cmp}. The analysis simplifies substantially. Firstly, one can check that the following classification holds.

Lemma 5.4.20. *If $U_0 \subset U_2^{cmp}$ U_0 is not a $GL(W)_{det_3}$-translate of a subspace of U_1^{cmp}, and $\partial_{2,1}(U_0, V) = 0$ does not force V to lie in U_2^{cmp}, then U_0 is a $GL(W)_{det_3}$ translate of either*

$$\begin{pmatrix} m & \lambda_1 o & \lambda_2 n \\ n & 0 & 0 \\ o & 0 & 0 \end{pmatrix} \quad or \quad \begin{pmatrix} 0 & \lambda_1 m & \lambda_2 n \\ m & 0 & 0 \\ n & 0 & 0 \end{pmatrix}.$$

Note however that if the former held in Lemma 5.4.20, then if we pick the complement of U_0 to be the space of matrices whose first row is zero, terms not containing U_0 cannot appear in q. In particular, q consists of at most $\partial_{2,1}(U_0, U_m)$ and $\partial_{1,1,1}(U_0, U_1, U_{m-1})$ and therefore does not give rise to any new boundary components.

Lemma 5.4.21. *If* $U_0 = \begin{pmatrix} 0 & \lambda_1 m & \lambda_2 n \\ m & 0 & 0 \\ n & 0 & 0 \end{pmatrix}$, *then q gives rise to no new boundary components.*

Proof. For now, we will assume $\partial_{1,1,1}(U_0, U_1, V)$ does not vanish identically on all V for which $\partial_{2,1}(U_0, U_1, V) = 0$. We claim that q consists of at most $\partial_{2,1}(U_0, U_m)$, $\partial_{1,1,1}(U_0, U_1, U_{m-1})$, and $\partial_{2,1}(U_0, U_1, U_{m-2})$.

Firstly, $\partial_{2,1}(U_0, V) = 0$ imposes three conditions on V, namely $V^2_2 = V^3_3 = 0$ and $\lambda_1 V^2_3 + \lambda_2 V^3_2 = 0$. Now suppose $\partial_{1,1,1}(U_i, U_j, U_k)$ appears in q for $i \geq 1$ and $j > 1$. We claim that $U_1 \subset U_2^{cmp}$ so that $\partial_{1,1,1}(U_0, U_1, V) = 0$ for $V = U_i, U_j, U_k$ forces $V \subset U_2^{cmp}$. In this case, $\partial_{1,1,1}(U_i, U_j, U_k)$ would vanish.

$\partial_{1,2}(U_0, U_1) = 0$ implies that U_1 satisfies either $(U_1)^2_3 = (U_1)^3_2 = 0$, or $(U_1)^1_2 = \lambda_1 \cdot (U_1)^2_1$ and $(U_1)^1_3 = \lambda_2 \cdot (U_1)^3_1$. In the former case, we're done. In the latter, $\det(U_1) = 0$ implies that either $U_1 \subset U_2^{cmp}$, in which case we're again done, or $(U_1)^1_1 = 0$. In this case, the condition $\partial_{2,1}(U_1, V) = 0$ for $V = U_i, U_j, U_k$ implies that either $U_1 \subset U_2^{cmp}$, or $V^1_1 = 0$. Both of these would force $\partial_{1,1,1}(U_i, U_j, U_k)$ to vanish, a contradiction.

Similarly, suppose $\partial_{1,1,1}(U_0, U_j, U_k)$ appeared in q for $j > 1$. $\partial_{1,1,1}(U_0, U_1, V) = 0$ for $V = U_j, U_k$ forces V to satisfy the same one or two conditions that U_1 does as a result of $\partial_{1,2}(U_0, U_1) = 0$, so $\partial_{1,1,1,}(U_0, U_j, U_k)$ must vanish.

Finally, we remove the assumption that $\partial_{1,1,1}(U_0, U_1, V)$ does not vanish identically on all V for which $\partial_{2,1}(U_0, U_1, V) = 0$. In this case

$$U_1 \subset \begin{pmatrix} x & 0 & \lambda_2 y \\ 0 & 0 & 0 \\ y & 0 & 0 \end{pmatrix}.$$

Then either $U_0 \oplus U_1$ is a $\mathrm{GL}(W)_{\det_3}$-translate of the first of the two spaces in Lemma 5.4.20, in which case q at most contains an additional $\partial_{1,1,1}(U_1, U_i, U_j)$ for some i, j, or U_1 is such that $\partial_{2,1}(U_0 \oplus U_1, V) = 0$ forces $V \in U_2^{cmp}$, in which case q still consists of at most $\partial_{2,1}(U_0, U_m)$, $\partial_{1,1,1}(U_0, U_1, U_{m-1})$, and $\partial_{1,1,1}(U_0, U_2, U_{m-2})$.

In any case, one can check that either q is either degenerate or a $\mathrm{GL}(W)$-translate of $\det_3$. □

5.4.6 Next steps

All of the above leads us to pose the following conjecture:

Conjecture 5.4.22. For every n, the irreducible components of Det_n are the endomorphism component and all orbit closures of polynomials of the form $\partial_{n-1,1}(U_0, U_1)$ for complementary subspaces $U_0 \oplus U_1 = W$ such that $\det(U_0) = 0$.

This has yet to be refuted in 4×4 case: of the non-endomorphism components of Det_4, we only know two, both of which arise as such polarizations $\partial_{3,1}(U_0, U_1)$:

Observation 10 ([45]). Denote the two primitive maximal subspaces in Theorem 5.2.9 respectively by U_0 and V_0. Denote by U_0' the imprimitive space of 4×4 matrices obtained from $W_{skew} \subset \mathcal{M}_{3,3}(\mathbb{C})$ in Example 5.2.4. Pick complements U_1, V_1, U_1' for U_0, V_0, U_0' and define $q_1 = \partial_{3,1}(U_0, U_1)$, $q_2 = \partial_{3,1}(V_0, V_1)$, and $q_3 = \partial_{3,1}(U_0', U_1')$. Then $\overline{\mathrm{GL}(W) \cdot [q_1]}$ and $\overline{\mathrm{GL}(W) \cdot [q_3]}$ are irreducible components of Det_4, but $\overline{\mathrm{GL}(W) \cdot [q_3]}$ is not.

Proof. With the computational approach outline in Example 5.4.4, one can check in *Macaulay2* that $\mathrm{Ann}_{\mathfrak{gl}(W)}(q_1)$ and $\mathrm{Ann}_{\mathfrak{gl}(W)}(q_3)$ have dimension 31, while $\mathrm{Ann}_{\mathfrak{gl}(W)}(q_2)$ has dimension 32. So by Lemma 5.4.3 the claim follows. □

5.5 Infinite families of components

It was shown in [36] that the construction $P_n := \partial_{n-1,1}(W_{skew}, W_{sym})$ gives rise to boundary components of Det_n more generally for arbitrary odd dimensions n (note that when n is even, $P_n = 0$).

Theorem 5.5.1. *Let $n \in \mathbb{N}$ be odd and define $P_n = \partial_{n-1,1}(W_{skew}, W_{sym})$. Then $GL(W)\cdot[P_n]$ is an irreducible component of the boundary of Det_n distinct from* $\mathrm{End}(W) \cdot [det_n] \backslash (GL(W) \cdot [det_n])$.

Sketch. That $\mathrm{GL}(W) \cdot [P_n]$ does not lie in the endomorphism component easily follows from checking that $V(P_n)$ is non-degenerate.

For the rest of the claim, first note that $P_n \in \partial Det_n$ because $\det(W_{skew}) = 0$. By Lemma 5.4.3, we wish to show that $\dim(\mathrm{Ann}_{\mathfrak{gl}(W)}(P_n) = \dim(\mathrm{Lie}(G_{P_n})) = 2n^2 - 1$.

Decompose $W = E \otimes E$ into $\Lambda^2 E \otimes S^2 E$.

$$\begin{aligned}
\mathrm{End}(W) &= \mathrm{End}(\Lambda^2 E \otimes S^2 E) \\
&= \mathrm{End}(\Lambda^2 E) \oplus \mathrm{End}(S^2 E) \oplus \mathrm{Hom}(\Lambda^2 E, S^2 E) \oplus \mathrm{Hom}(S^2 E, \Lambda^2 E) \\
&= (\Lambda^2 E \otimes \Lambda^2 E^*) \oplus (S^2 E \otimes S^2 E^*) \oplus (\Lambda^2 E^* \otimes S^2 E) \oplus (S^2 E^* \oplus \Lambda^2 E).
\end{aligned}$$

By several applications of the Littlewood-Richardson rule, we may further decompose these four components into $GL(E)$-submodules as

$$\Lambda^2 E \otimes \Lambda^2 E^* = \mathfrak{gl}(E) \oplus S_{2^2,1^{n-4}}(E)$$

$$S^2 E \otimes S^2 E^* = \mathfrak{gl}(E) \oplus S_{4,2^{n-2}}(E)$$

$$\Lambda^2 E^* \otimes S^2 E = \mathfrak{sl}(E) \oplus S_{3,1^{n-3}}(E)$$

$$S^2E^* \otimes \Lambda^2 E = \mathfrak{sl}(E) \oplus S_{3^2,2^{n-3}}(E).$$

The fact that $\dim(\mathrm{Lie}(G_{P_n})) = 2n^2 - 1$ follows from testing on highest weight vectors.

We will carry out one such computation explicitly. The following result will be useful later in this section when we construct another infinite family of boundary components.

Lemma 5.5.2. $\mathrm{Lie}(G_{P_n})$ *has a component of dimension* $m^2 - 1$ *in* $\mathrm{End}(\Lambda^2 E, S^2 E)$.

Proof. Pick a basis $e_{ij} = e_i \otimes e_j$ for $W = E \otimes E$. This gives a basis $\{T^{ij}_{\hat{k}\hat{l}}\}$ for $\mathrm{End}(E \otimes E)$, where $T^{k\ell}_{\hat{i}\hat{j}}$ sends e_{ij} to $e_{k\ell}$ and vanishes on all other $e_{i'j'}$. One can check that

$$T^{1n}_{\hat{n}\hat{n}} - T^{n1}_{\hat{n}\hat{n}} + \sum_{i=2}^{n-1} \left(T^{1j}_{\hat{j}\hat{n}} - T^{j1}_{\hat{j}\hat{n}} + T^{1j}_{\hat{n}\hat{j}} - T^{j1}_{\hat{n}\hat{j}} \right) \tag{5.9}$$

is a highest weight vector of $\mathfrak{sl}(E) = S_{2,1^{n-2}}E \subset \mathrm{End}(\Lambda^2 E, S^2 E)$ in the decomposition of $\mathrm{End}(E \otimes E)$ as a $\mathrm{GL}(E)$-module.

The Lie algebra action of $T^{k\ell}_{\hat{i}\hat{j}}$ on $S^n W^*$ is given by $e_{k\ell} \cdot \partial_{e_{ij}}$, so in particular, the highest weight vector in (5.9) kills P_n as desired. □

□

To find infinite families for even n, we can try looking for infinite families of spaces of singular matrices that are not compression spaces. As noted in Section 5.2, one way to produce such a space is via Example 5.2.4: for any even n, if W_{skew} denotes the space of $(n-1) \times (n-1)$ skew-symmetric matrices, then define

$$W_{pad} := \begin{pmatrix} W_{skew} & * \\ \mathbf{0} & * \end{pmatrix}, \quad W^c_{pad} := \begin{pmatrix} W_{sym} & \mathbf{0} \\ * & 0 \end{pmatrix}.$$

W_{pad} is clearly a space of singular $n \times n$ matrices, and $W_{pad} \oplus W^c_{pad} = W$. Define $Q_n := \partial_{n-1,1}(W_{pad}, W^c_{pad})$. For instance, for $n = 4$,

$$W_{pad} = \begin{pmatrix} 0 & \alpha_{12} & \alpha_{13} & x_1 \\ -\alpha_{12} & 0 & \alpha_{23} & x_2 \\ -\alpha_{13} & -\alpha_{23} & 0 & x_3 \\ 0 & 0 & 0 & y \end{pmatrix}, \qquad W^c_{pad} = \begin{pmatrix} \beta_{11} & \beta_{12} & \beta_{13} & 0 \\ \beta_{12} & \beta_{22} & \beta_{23} & 0 \\ \beta_{13} & \beta_{23} & \beta_{33} & 0 \\ z_1 & z_2 & z_3 & 0 \end{pmatrix},$$

and

$$Q_n = \begin{vmatrix} 0 & \alpha_{12} & \beta_{13} & x_1 \\ -\alpha_{12} & 0 & \beta_{23} & x_2 \\ -\alpha_{13} & -\alpha_{23} & \beta_{33} & x_3 \\ 0 & 0 & z_3 & y \end{vmatrix} + \begin{vmatrix} 0 & \beta_{12} & \alpha_{13} & x_1 \\ -\alpha_{12} & \beta_{22} & \alpha_{23} & x_2 \\ -\alpha_{13} & \beta_{23} & 0 & x_3 \\ 0 & z_2 & 0 & y \end{vmatrix} + \begin{vmatrix} \beta_{11} & \alpha_{12} & \alpha_{13} & x_1 \\ \beta_{12} & 0 & \alpha_{23} & x_2 \\ \beta_{13} & -\alpha_{23} & 0 & x_3 \\ z_1 & 0 & 0 & y \end{vmatrix}.$$

As we saw in Observation 10, $\mathrm{GL}(W) \cdot [Q_4]$ is a nontrivial boundary component of Det_4. We now show that this holds for all even n.

Theorem 5.5.3. *Let $n \in \mathbb{N}$ be even. Then $GL(W) \cdot [Q_n]$ is an irreducible component of the boundary of Det_n distinct from* $\mathrm{End}(W) \cdot det_n \backslash (GL(W) \cdot det_n)$.

Proof. As in the proof of Theorem 5.5.1, checking that $\overline{\mathrm{GL}(W) \cdot Q_n}$ is distinct from $\mathrm{End}(W) \cdot \det_n \backslash (\mathrm{GL}(W) \cdot \det_n)$ amounts to checking that $V(Q_n)$ is not a cone.

For the rest of the claim, as before, we want to show $\dim(\mathrm{Lie}(G_{Q_n})) = 2n^2 - 1$. For some linear coordinate ℓ, let $E = E' \oplus \ell$. Then $E \otimes E$ decomposes into the blocks

$$\begin{pmatrix} \Lambda^2 E' \oplus S^2 E' & A \\ B & \ell \otimes \ell \end{pmatrix}, \tag{5.10}$$

where $A = E' \otimes \ell$ and $B = \ell \otimes E'$. Pick bases $\{e^{\Lambda}_{ij}\}$, $\{e^S_{ij}\}$, $\{a_i\}$, $\{b_i\}$, and x for $\Lambda^2 E'$, $S^2 E'$, A, B, and ℓ respectively; these together form a basis for $E \otimes E$.

	Λ^2E'	S^2E'	A	B	ℓ
$\partial_{e^{\Lambda}_{ij}}$	1	1	0	0	1
	1	0	1	1	0
$\partial_{e^{S}_{ij}}$	2	0	0	0	1
∂_{a_i}	2	0	0	1	0
∂_{b_i}	2	0	1	0	0
∂_x	2	1	0	0	0

Table 5.1: Gradings for partial derivatives of Q_n

For convenience, we say that $p \in S^dW^*$ is $(d_1, d_2, d_3, d_4, d_5)$-graded if p is of degree $d_1, ..., d_5$ in Λ^2E', S^2E', A, B, and ℓ respectively. Table 5.1 summarizes the gradings for the various partial derivatives of Q_n

Because $\mathrm{Lie}(G_{Q_n}) = \mathrm{Ann}_{\mathfrak{gl}(W)}\, Q_n$, we want to show that there are $2n^2 - 1$ linear dependencies among all $T^{k\ell}_{\hat{i}\hat{j}}.Q_n$, where the $T^{k\ell}_{\hat{i}\hat{j}}$ are defined with respect to the basis we have chosen for $E \otimes E$.

By the gradings in Table 5.1, there can only exist linear dependencies among

1) $\{e^{\Lambda}_{ij} \cdot \partial_{e^{\Lambda}_{k\ell}} Q_n\}$, $\{e^{S}_{ij} \cdot \partial_{e^{S}_{k\ell}} Q_n\}$, $\{a_j \cdot \partial_{a_i} Q_n\}$, $\{b_j \cdot \partial_{b_i} Q_n\}$, and $x \cdot \partial_x Q_n$

2) $\{b_k \cdot \partial_{e^{S}_{ij}} Q_n\}$ and $\{x \cdot \partial_{a_i} Q_n\}$

3) $\{a_k \cdot \partial_{e^{S}_{ij}} Q_n\}$ and $\{x \cdot \partial_{b_i} Q_n\}$

4) $\{e^{\Lambda}_{ij} \cdot \partial_{e^{S}_{k\ell}} Q_n\}$

5) $\{e^{\Lambda}_{ij} \cdot \partial_{a_k} Q_n\}$

6) $\{e^{\Lambda}_{ij} \cdot \partial_{b_k} Q_n\}$

7) $\{e^{S}_{ij} \cdot \partial_{a_k} Q_n\}$ and $\{b_j \cdot \partial_x Q_n\}$

8) $\{e^{S}_{ij} \cdot \partial_{b_k} Q_n\}$ and $\{a_j \cdot \partial_x Q_n\}$

In 1), there is one linear dependency among members of $\{a_j \cdot \partial_{a_i} Q_n\}$ and $\{b_j \cdot \partial_{b_i} Q_n\}$, and one among $x \cdot \partial_x Q_n$ and members of $\{e^{S}_{j} \cdot \partial_{e^{S}_{i}} Q_n\}$. Additionally, all $(n-1)^2$ members of $\{e^{\Lambda}_{j} \cdot \partial_{e^{\Lambda}_{i}} Q_n\}$ lie in the span of $\{e^{S}_{j} \cdot \partial_{e^{S}_{i}} Q_n\}$, $\{a_j \cdot \partial_{a_i} Q_n\}$, $\{b_j \cdot \partial_{b_i} Q_n\}$, and $x \cdot \partial_x Q_n$. We conclude that there are exactly $1 + 1 + (n-1)^2 = (n-1)^2 + 2$ dependencies in case 1).

In 2), all $n-1$ members of $\{x \cdot \partial_{a_i} Q_n\}$ are already in the span of $\{b_k \cdot \partial_{e^{S}_{ij}} Q_n\}$, so there are $n-1$ dependencies in case 2). Analogously, there are $n-1$ dependencies in case 3).

In 4), Lemma 5.5.2 tells us that there exactly $(n-1)^2 - 1$ dependencies.

In 5), every collection of $(n-1) - 1$ members of $\{e^{\Lambda}_{ij} \cdot \partial_{a_k} Q_n\}$ has a unique linear dependency, for a total of $n-1$ dependencies. The same holds in case 6). On the other hand, one can check that in cases 7) and 8), no dependencies exist.

We conclude that $\dim(\mathrm{Lie}(G_{Q_n})) = 2(n-1)^2 + 1 + 4(n-1) = 2n^2 - 1$ as desired. □

Bibliography

[1] S. Aaronson, A. Wigderson. Algebrization: A New Barrier in Complexity Theory. ACM Transactions on Theory of Computing 1(1), 2009.

[2] M.D. Atkinson, Primitive spaces of matrices of bouned rank, II. *J. Austral. Math. Soc.* 34 (1983), 306-315.

[3] T. Baker, J. Gill, and R. Solovay. Relativizations of the P=?NP question. SIAM J. Comput., 4:431442, 1975

[4] J.-Y. Cai and V. Choudhary. Some results on matchgates and holographic algorithms, *Automata, languages, and programming. Part 1, Lecture Notes in Comput. Sci.*, vol 4051, Springer, Berlin, 2006, pp. 703-714.

[5] J.-Y. Cai and V. Choudhary. Valiant's holant theorem and matchgate tensors, *Theory and applications of models of computation, Lecture Notes in Comput. Sci.*, vol. 3959, Springer, Berlin, 2006, pp. 248-261.

[6] J.-Y. Cai and V. Choudhary. Valiant's holant theorem and matchgate tensors, *Theoret. Comput. Sci.* 384 (2007), no. 1, 22-32.

[7] J.-Y. Cai, V. Choudhary, and P. Lu. On the theory of matchgate computations, *Theory Comput. Syst.*, 45(1):108132, 2009. Preliminary version in CCC'07.

[8] J.-Y. Cai and Z. Fu. (2014). A collapse theorem for holographic algorithms with matchgates on domain size at most 4. *Information and Computation*, 239, 149-169.

[9] J.-Y. Cai and A. Gorenstein. Matchgates revisited. *CoRR*, abs/1303.6729, 2013.

[10] J.-Y. Cai and P. Lu. Holographic algorithms: from art to science, *STOC '07– Proceedings of the 39th Annual ACM Symposium on Theory of Computing*, AMC, New York, 2007, pp. 401-410.

[11] J.-Y. Cai and P. Lu: Holographic algorithms: the power of dimensionality resolved. In: Arge, L., Cachin, C., Jurdzinski, T. et al. (eds.) *Proceedings of ICALP*. Lecture Notes in Computer Science, vol. 4596, pp. 631642. Springer, Berlin (2007)

[12] J.-Y. Cai and P. Lu. On symmetric signatures in holographic algorithms, *STACS 2007, Lecture Notes in Comput. Sci.*, vol. 4393, Springer, Berlin, 2007, pp. 429-440.

[13] J.-Y. Cai and P. Lu. Basis collapse in holographic algorithms, *Computational Complexity* 17 (2008), no. 2, 254-281.

[14] J.-Y. Cai, P. Lu, and M. Xia. Holographic reduction, interpolation and hardness. *Computational Complexity*, 21(4):573604, 2012.

[15] S. Chen. Basis Collapse for Holographic Algorithms Over All Domain Sizes. To appear in *STOC* '16. `http://arxiv.org/abs/1511.00778`

[16] C.C. Chevalley. Algebraic Theory of Spinors. Columbia University Press, New York, 1954.

[17] A. Church, A note on the Entscheidungsproblem. Journal of Symbolic Logic, 1 (1936), pp 4041.

[18] S. Cook. The complexity of theorem-proving procedures, in *Conference Record of Third Annual ACM Symposium on Theory of Computing*, ACM, New York, 1971, 151158.

[19] D. Eisenbud, J. Harris. Vector spaces of matrices of low rank, Adv. in Math, 70 (1988), no. 2, 135- 155.

[20] G. Frobenius, Uber die Darstellung der endlichen Gruppen durch lineare Substitutionen, Sitzungsber Deutsch. Akad. Wiss. Berline (1897), 994-1015.

[21] Z. Fu, F. Yang: Holographic algorithms on bases of rank 2. `http://arxiv/abs/1303.7361`. In submission.

[22] I.M. Gelfand, M.M. Kapranov, and A.V. Zelevinsky, Discriminants, resultants, and multidimensional determinants, Mathematics: Theory & Applications, Birkhauser, Boston, 1994.

[23] B. Grenet, An Upper Bound for the Permanent versus Determinant Problem, *Theory of Computing* (2014), Accepted.

[24] A. Grothendieck (1957). Sur La Classification Des Fibres Holomorphes Sur La Sphere de Riemann. American Journal of Mathematics, 79(1), 121138.

[25] A. Gupta, P. Kamath, N. Kayal, and R. Saptharishi. Arithmetic circuits: A chasm at depth three, Electronic Colloquium on Computational Complexity (ECCC) 20 (2013), 26.

[26] R. Howe, $(\mathrm{GL}_n, \mathrm{GL}_m)$-duality and symmetric plethysm. *Proc. Indian Acad. Sci. (Math. Sci.)* 97 (1987), 85-109.

[27] J. Huttenhain, Pierre Lairez. The boundary of the orbit of the 3×3 determinant polynomial. `http://arxiv.org/abs/1512.02437`

[28] K. Iwama and H. Morizumi. An explicit lower bound of 5n o(n) for Boolean circuits. In *Mathematical foundations of computer science 2002*, volume 2420 of Lecture Notes in Comput. Sci., pages 353364. Springer, Berlin, 2002.

[29] P.W. Kasteleyn: The statistics of dimers on a lattice. Physica 27, 12091225 (1961)

[30] G. Kempf, Instability in invariant theory. *Ann. of Math.* 108 (1978), 299-316.

[31] S. Kumar. Geometry of orbits of permanents and determinants, Comment. Math. Helv. 88 (2013), no. 3, 759788

[32] J.M. Landsberg. Geometry and complexity. Lecture notes.

[33] J. M. Landsberg. Geometric complexity theory: an introduction for geometers. In: Ann. Univ. Ferrara. Sez. VII Sci. Mat. (2015) 61.1, pp. 65117.

[34] J.M. Landsberg, J. Morton, and S. Norine. Holographic algorithms without matchgates. preprint arXiv:0904.0471, 2009.

[35] J.M. Landsberg and Giorgio Ottaviani, Equations for secant varieties of Veronese and other varieties, Ann. Mat. Pura Appl. (4) 192 (2013), no. 4, 569606.

[36] J.M. Landsberg, L. Manivel, and N. Ressayre. Hypersurfaces with degenerate duals and the geometric complexity theory program, Comment. Math. Helv. 88 (2013), no. 2, 469484.

[37] J.M. Landsberg and N. Ressayre. Permanent v. determinant: an exponential lower bound assuming symmetry and a potential path towards Valiants conjecture. In *Proceedings of the 2016 ACM Conference on Innovations in Theoretical Computer Science*, pp. 29-35.

[38] J.M. Landsberg, J. Morton, and S. Norine. Holographic algorithms without matchgates. *Tensors and Multilinear Algebra* 438 (2013), no. 2, pp. 782-795.

[39] A. Li and M. Xia: A theory for Valiant's matchcircuits (extended abstract). *In Proc. 25th Symp. Theoretical Aspects of Comp. Sci. (STACS'08)*, pp. 491502. Schloss Dagstuhl, 2008.

[40] G. Malod and N. Portier, Characterizing Valiant's algebraic complexity classes, Journal of Complexity. 24 (2008), 16-38.

[41] M. Marcus and F. C. May, The permanent function, Canad. J. Math. 14 (1962), 177189.

[42] Y. Matsushima: Espaces homogenes de Stein des groupes de Lie complexes. Nagoya Math. J. 18 153164 (1961)

[43] T. Muir. A Treatise on the Theory of Determinants.Macmillan and Co (1882).

[44] K.D. Mulmuley and M. Sohoni, Geometric complexity theory. I. An approach to the P vs. NP and related problems, SIAM J. Comput. 31 (2001), no. 2, 496526 (electronic).

[45] L. Oeding, personal communication, September 10, 2015

[46] A. Razborov and S. Rudich, Natural proofs. J. Comput. Syst. Sci. 55 (1997), no. 11, 2435.

[47] A.L. Onishchik and E.B. Vinberg, *Lie groups and algebraic groups*, Springer Series in Soviet Mathematics, Springer-Verlag, Berlin, 1990, Translated from Russian and with a preface by D.A. Leites.

[48] I. Shafarevich. *Basic algebraic geometry* 1. Springer-Verlag, Berlin 1994.

[49] J.R. Stembridge, Nonintersecting paths, Pfaffians, and plane partitions, Adv. Math. 83 (1990), no. 1, 96131.

[50] H.N.V. Temperley and M.E. Fisher. Dimer problem in statistical mechanicsan exact result. Philos. Mag. 6, 10611063 (1961)

[51] S. Toda. Classes of arithmetic circuits capturing the complexity of computing the determinant. *IEICE Transactions on Information and Systems*, E75-D:116124, 1992.

[52] A. Turing, On computable numbers, with an application to the Entscheidungsproblem. Proceedings of the London Mathematical Society, Series 2, 42 (1936-7), pp 230265.

[53] S.P. Vadhan. The complexity of counting in sparse, regular and planar graphs, *SIAM J. Comput.* 31(2): 398427, 2001.

[54] L. G. Valiant (1979a). Completeness classes in algebra. In *Proc. 11th ACM Symposium on Theory of Computing*, 249261.

[55] L. G. Valiant (1979b). The complexity of computing the permanent. *Theor. Comp. Sci.* 8, 189-201.

[56] L. G. Valiant, Reducibility by algebraic projections, Logic and Algorithmic: an International Symposium held in honor of Ernst Specker, vol. 30, Monogr. No. 30 de l'Enseign. Math., 1982, 365-380.

[57] L.G. Valiant. Expressiveness of matchgates, *Theoret. Comput Sci.* 289 (2002), no. 1, 457-471.

[58] L.G. Valiant. Quantum circuits that can be simulated classically in polynomial time. SIAM J. Comput., 31(4):12291254, 2002. Preliminary version in STOC'01.

[59] L.G. Valiant. Accidental algorithms. In *FOCS*, pages 509517. IEEE Computer Society, 2006.

[60] L.G. Valiant. Holographic algorithms (extended abstract). In *Proc. 45th IEEE Symposium on Foundations of Computer Science*, pp. 306315 (2004).

[61] L.G. Valiant. Holographic circuits, *Automata, languages and programming, Lecture Notes in Comput. Sci.*, vol. 3580, Springer, Berlin, 2005, pp. 1-15.

[62] L.G. Valiant. Some observations on holographic algorithms, Proc. 9th Latin American Theoretical Informatics Symposium, LATIN 2010. LNCS, Vol 6034 Springer-Verlag (2010), 577-590.

[63] Joachim von zur Gathen, Permanent and determinant, Linear Algebra Appl. 96 (1987), 87100.

[64] V. S. Varadarajan. Supersymmetry for Mathematicians: An Introduction (Courant Lecture Notes). American Mathematical Society (2004), Chapter 5.

[65] M. Xia, P. Zhang, and W. Zhao, The complexity of counting on 3-regular planar graphs, *Theoret. Comput. Sci.*, 384 (2007), pp. 111125.

Appendices

Appendix A

A.1 Complexity Classes

Theoretical computer scientists use the word *language* in an alphabet Σ to refer to subsets of $\Sigma^* = \cup_{n=1}^{\infty} \Sigma^n$. For example, the set $\{0^k 1^k \mid k \in \mathbb{N}\}$ is a language over the Boolean alphabet.

The basic model of computation used in complexity theory is the Turing machine, but for the

Definition A.1.1. A *deterministic Turing machine* M on an alphabet Σ consists of a read-only input tape and k work tapes each containing a one-dimensional array of cells on which characters of Σ can be printed, $k+1$ heads over each of these tapes, a collection Q of possible internal states including an "accept" and "reject" state, and a transition map $Q \times \Sigma^{k+1} \to Q \times \Sigma^k \times \{L, R\}^{k+1}$ which, given an internal state and all character read by the heads, specifies the characters that the work tape heads should print on their respective cells, the new internal state of the machine, and the directions that each of the heads should move next on their respective tapes.

A Turing machine M is said to decide a language L over Σ if for every $x \in \{0,1\}^*$, M with initially empty worktapes and an input tape consisting only of the characters of x eventually reaches the "accept" state if $x \in L$ and eventually reaches the "reject" state otherwise.

Definition A.1.2. A language L is in P if there exists a deterministic Turing machine M that decides L in time $\mathsf{poly}(n)$, where n is the length of the input.

Definition A.1.3. A language L over Σ is in NP if there exists some polynomial p and a language L' over $\Sigma \cup \{\$\}$ in P such that for every $x \in L$, there exists a corresponding *certificate* $s_x \in \{0,1\}^{p(|x|)}$ for which $x\$s_x \in L'$.

Two other complexity classes that we encounter in this work are $\#\mathsf{P}$ and $\oplus\mathsf{P}$, which can be thought of respectively as the "counting" and "counting modulo 2" versions of NP: instead of determining whether a certificate for a given input x exists verifying that $x \in L$, one is asked to count the number or parity of such certificates.

Definition A.1.4. A function $f : \Sigma^* \to \mathbb{N}$ is in $\#\mathsf{P}$ if there exists some polynomial p and a language $L \in \mathsf{P}$ for which $f(x)$ equals the number of $s_x \in \{0,1\}^{p(|x|)}$ for which $x\$s_x \in L$.

Definition A.1.5. A function $f : \Sigma^* \to \{0,1\}$ is in $\oplus\mathsf{P}$ if there exists some polynomial p and a language $L \in \mathsf{P}$ for which $f(x)$ equals the parity of the number of $s_x \in \{0,1\}^{p(|x|)}$ for which $x\$s_x \in L$.

A.2 Weakly Skew Circuits

The following was first defined in [51].

Definition A.2.1. A *weakly skew circuit* is an arithmetic circuit C such that for any computation gate labeled with $\times$, removal of that node in C separates the connected component of one of its incoming neighbors from the rest of C. We call the other incoming neighbor a *reusable gate*.

In other words, each multiplication gate must have one argument which is computed solely for that gate. The motivation for this is that in the proof of Lemma 4.1.6, there is not much reuse of results of intermediate computations.

Definition A.2.2. A sequence of polynomials (p_n) is said to lie in VP_{ws} if there exists a corresponding sequence of weakly skew circuits (C_n) computing (p_n) for which $\text{size}(C_n), \deg(C_n) \leq \mathsf{poly}(n)$.

We now prove that $(\det_n)$ is VP_{ws}-complete.

Proof of Theorem 4.1.8. To show $(\det_n) \in \mathsf{VP}_{ws}$, one just has to check that in the proof of Lemma 4.1.6, matrix multiplication can be simulated even by weakly skew circuits.

To show $(\det_n)$ is VP_{ws}-hard, we first prove that the following construction is possible. In a weighted directed acyclic graph G, define the *weight* of a path to be the product of the weights on its edges. Define the *weight* of a pair of vertices (s,t) to be the sum of the weights of all paths connecting s and t.

Lemma A.2.3. *Let C be a weakly skew circuit of size n with multiple outputs. Then there exists a weighted directed acyclic graph G with at most $n+1$ vertices, exactly one of which has indegree zero, such that for every reusable gate p in C, there is a vertex $t_p \in G$ for which the weight of (s,t_p) is the same polynomial that is computed by p.*

Proof. Naturally, the argument here is inductive and proceeds by casework on α. For brevity, we will only illustrate the argument for when α is a multiplication gate. Suppose removal of p separates C into disjoint circuits C_1 and C_2 which compute polynomials p_1 and p_2 respectively, and suppose C_2 is the component whose gates are not reusable. Inductively, we obtain graphs G_1 and G_2 and pairs of vertices (s,t_1) and (t,s_2) whose weights agree with p_1 and p_2 respectively. One can check that $G = (G_1 \sqcup G_2)/(t_1 \sim s_2)$ satisfies the desired properties. □

Now take any $f \in \mathsf{VP}_{ws}$ computed by some weakly skew C of size n and look at the corresponding graph G guaranteed by Lemma A.2.3. We will show that m arises as an $(n+1)\times(n+1)$ determinant, specifically of an adjacency matrix. Modify G by gluing together s and t, negating the weight of every edge, and attaching a loop to every vertex other than $s = t$; denote the adjacency matrix of this new graph by A. By construction, $\det_n(A) = -f$, so by adding an extra block of size 1 and single entry -1 to A, we get the desired $(n+1)\times(n+1)$ matrix. □

A.3 Algebraic Peter-Weyl Theorem

In this section we prove Theorem 4.4.3. For any finite-dimensional representation $\rho : G \to \mathrm{GL}(V)$, define the matrix coefficient map $i_V : V^* \otimes V \to \mathbb{C}[G]$ by $i_V(\phi \otimes v)(g) = \phi(gv)$. The image of i_V is the *space of matrix coefficients of V*. We show that each isotypic component of type V_λ in $\mathbb{C}[G]$ is the space of matrix coefficients of V_λ, from which the reductivity of G implies the decomposition in Theorem 4.4.3.

Lemma A.3.1. *If V is irreducible and finite-dimensional, then i_V is $G\times G$-equivariant and injective. Moreover, $i_V(V^* \otimes V)$ equals the isotypic component of type V under the right-action of G and the isotypic component of type V^* under the left-action of G.*

Proof. We check $G\times G$-equivariance directly: for $(\phi, v) \in V^* \otimes V$, and $(g_1, g_2) \in G \times G$, we have that

$$i_V(g_1\phi \otimes g_2 v)(g) = (g_1\phi)(g g_2 v) = \phi(g^{-1} g g_2 v) = (g_1 \cdot (i_V(\phi \otimes v)) \cdot g_2)(g)$$

for all $g \in G$. For injectivity, this just follows from the fact that $V^* \otimes V$ is an irreducible $G\times G$-module so that i_V cannot have a kernel.

For the last part, we will prove it for the right-action of G; the other half is analogous. We want to show that if $\iota : V \hookrightarrow \mathbb{C}[G]$ is any G-equivariant embedding under the right-action, then its image is contained inside the space of matrix coefficients. Indeed, for any $g \in G$, $\iota(v)(g) = \iota(gv)(1) = i_V(\alpha \otimes v)(g)$, where $\alpha \in V^*$ is defined by $v \mapsto \iota(v)\cdot 1$, so we're done. □

A.4 Schur-Weyl Duality and Representations of GL(V)

Let V be any n-dimensional vector space. An understanding of the finite-dimensional irreducible GL(V)-modules is crucial to the above search for obstructions. Consider the actions of GL(V) and $\mathbb{S}_d$ on $V^{\otimes d}$:

$$T(v_1 \otimes \cdots \otimes v_d) = Tv_1 \otimes \cdots \otimes Tv_n$$

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

for $T \in \mathrm{GL}(V)$ and $\sigma \in \mathbb{S}_d$. It is apparent that these two actions commute. In fact, not only do the actions of these two groups commute with each other, but in fact we shall show they are each other's centralizers using the double centralizer theorem. From this follows the famous Schur-Weyl duality, which gives a direct-sum decomposition of $V^{\otimes d}$ encapsulating the relationship between finite-dimensional irreducible GL(V)-modules and $\mathbb{C}[\mathbb{S}_d]$-modules.

Theorem A.4.1. *For vector space V,*

$$V^{\otimes d} = \bigoplus_{\pi} S_\pi V \otimes [\pi]$$

as a $GL(V) \times S_d$-module, where the sum is taken over all partitions π of n, and $S_\pi V$ denotes the Weyl module $\mathrm{Hom}_{\mathbb{S}_d}([\pi], V^{\otimes d})$. *Note further that $S_\pi(V) = 0$ if $\ell(\pi) \geq \dim(V)$*

The rest of this section will be devoted to proving this.

Definition A.4.2. If $S \subset \mathrm{End}(V)$, then the *centralizer* $S' \subset \mathrm{End}(V)$ is defined by

$$S' := \{X \in \mathrm{End}(V) \mid Xs = sX \ \forall \ s \in S\}.$$

Theorem A.4.3 (Double Centralizer Theorem)**.** *If $\mathcal{A} \subset \mathrm{End}(V)$ is a completely reducible associative algebra, then $\mathcal{A} = \mathcal{A}''$.*

Proof. Inclusion from left to right is obvious: every element of $\mathcal{A}'$ commutes with every element of $\mathcal{A}$, so every element of $\mathcal{A}$ commutes with every element of $\mathcal{A}'$, i.e. $\mathcal{A} \subset (\mathcal{A}')'$.

In the other direction, take any $T \in \mathcal{A}''$, and select a basis $v_1 ..., v_n$ of V. We claim there exists $a \in \mathcal{A}$ for which $T(v_j) = av_j$ for all j. Let $w = v_1 \oplus \cdots \oplus v_n \in V^{\oplus n}$ and consider an $\mathcal{A}$-equivariant projection $\pi : V^{\oplus n} \to \mathcal{A}w$. By definition of $\mathcal{A}$-equivariance, $\pi \in \mathcal{A}'$, so $T(\pi(w)) = \pi(T(w))$. But $T(\pi(w)) = T(w)$ while $\pi(T(w)) \in \mathcal{A}w$, so

$$T(v_1) \oplus \cdots \oplus T(v_n) = T(w) = aw = av_1 \oplus \cdots \oplus av_n$$

for some $a \in \mathcal{A}$ as desired. ☐

We now use this to show that GL(V) and $\mathbb{S}_d$ are each other's centralizers.

Lemma A.4.4. $\mathrm{End}_{GL(V)}(V^{\otimes d}) = \mathbb{C}[\mathbb{S}_d]$.

Proof. By the double centralizer theorem, it is enough to show that the algebra $\mathrm{End}_{\mathbb{C}[\mathbb{S}_d]}(V^{\otimes d})$ is generated by GL(V). But under the natural identification of $\mathrm{End}(V^{\otimes d})$ with $(V \otimes V^*)^{\otimes d}$, $\mathrm{End}_{\mathbb{C}[\mathbb{S}_d]}(V^{\otimes d})$ is identified with $S^d(V \otimes V^*)$. Inside $(V \otimes V^*)^{\otimes d}$, the action of g belongs in $S^d(V \otimes V^*)$ for each g, and GL(V) collectively generates $S^d(V \otimes V^*)$, so we're done. ☐

It remains to describe the isotypic components of G and $\mathrm{End}_G(W)$ for $W = V^{\otimes d}$ and $G = \mathrm{GL}(V)$.

Lemma A.4.5. *For reductive group G and G-module W, the isotypic components of G and* $\mathrm{End}_G(W)$ *in W are the same. Moreover, if U is one such isotypic component arising from irreducible representations A of G and B of* $\mathrm{End}_G(W)$, *then $U = A \otimes B$ as a $G \times \mathrm{End}_G(W)$-module, and*

$$A = \mathrm{Hom}_{\mathrm{End}_G(W)}(B, U), \qquad B = \mathrm{Hom}_G(A, U).$$

Proof. It is enough to show that for any irreducible G-module A, we can take the corresponding $\mathrm{End}_G(W)$-module to be $\mathrm{Hom}_G(A,W)$, and that correspondingly, for any irreducible $\mathrm{End}_G(W)$-module B, we can take the corresponding G-module to be $\mathrm{Hom}_{\mathrm{End}(G)}(B,W)$. We only show the former; the latter is entirely analogous.

Specifically, we need to check 1) that $\mathrm{Hom}_G(A,W)$ is indeed an irreducible $\mathrm{End}_G(W)$-module, and 2) that $A \otimes \mathrm{Hom}_G(A,W)$ is the isotypic component of A in W.

For 1), it is enough to show that the action of $\mathrm{End}_G(W)$ on $\mathrm{Hom}_G(A,W)$ is transitive, i.e. for any $s,t \in \mathrm{Hom}_G(A,W)$, there is some $a \in \mathrm{End}_G(W)$ for which $at = s$. We can extend the map defined on tA by multiplication by st^{-1} to an $a' \in \mathrm{End}_G(W)$ so that $a't$ defines an isomorphism $A \to sA$. Post-composing this with left-multiplication by s^{-1} gives an isomorphism and thus, by Schur's lemma, multiplication by a nonzero scalar λ. Define $a = c^{-1}a'$, proving transitivity.

For 2), suppose $U = A \otimes B$ were the isotypic component of A. B certainly lies inside $\mathrm{Hom}_G(A,W)$: to any $b \in B$, associate the G-equivariant map $a \mapsto a \otimes b$. On the other hand, by definition of isotypic component, all G-equivariant embeddings of A in W land in U, so $\mathrm{Hom}_G(A,W) \subseteq B$, and we're done. □

Theorem A.4.1 thus follows from Lemma A.4.5.

A.5 Complex Algebraic Groups

Definition A.5.1. An affine variety G over $\mathbb{C}$ for which there exist morphisms of varieties $\mu : G \times G \to G$, $e : \mathrm{Spec}(k) \to G$, and $\iota : G \to G$ which give G the structure of a group, then G is a *affine complex algebraic group.*

Let G be an affine complex algebraic group.

Definition A.5.2. G is *reductive* if every G-module decomposes as a direct sum of irreducible G-modules.

Example A.5.3. Denote the coordinate ring of the variety $\mathbb{C}^{n^2+1}$ by $\mathbb{C}[x_{11}, x_{12}, ..., x_{nn}, y]$. $\mathrm{GL}(\mathbb{C}^n)$ is an algebraic group cut out by the equation $\det((x_{ij}))\cdot y = 1$, which enforces the condition that the determinants of the matrices (x_{ij}) in $\mathrm{GL}(\mathbb{C}^n)$ are nonzero. As we show in Theorem A.4.1 Appendix A.4, $\mathrm{GL}(\mathbb{C}^n)$ is reductive.

Definition A.5.4. A module is *simple/irreducible* if it contains no proper-submodules. An algebra $\mathcal{A}$ (resp. algebraic group G) is *completely reducible* (resp. *reductive*) if every $\mathcal{A}$-module (resp. G-module) has a direct-sum decomposition into irreducible modules.

A useful fact about reductive groups that we use in Section 5.1 to prove Theorem 5.1.2 is the following:

Theorem A.5.5 (Matsushima, [42])**.** *If G is a reductive group and H a subgroup, then G/H is affine if and only if H is reductive.*

A.6 Planarizing Matchgates

In the proof of Lemma 3.4.3, we made several initial constructions of transducers to achieve certain row and column operations but noted that those constructions, specifically those shown in Figures 3.2b, 3.2c, 3.2d, needed to be modified because they were not planar. Following the technique of Cai and Gorenstein [9], we planarize those matchgates by replacing every edge crossing with the so-called *crossover gadget* X shown in Figure A.1.

Because the standard signature $\underline{X}$ of the crossover gadget is given by $\underline{X}^{0000} = 1$, $\underline{X}^{0101} = 1$, $\underline{X}^{1010} = 1$, $\underline{X}^{1111} = -1$, and $\underline{X}^{\sigma} = 0$ for all other $\sigma \in \{0,1\}^4$ so that the standard signature remains invariant under any cyclic permutation of the external nodes, the orientation of the copy of X placed over an edge crossing does not matter.

We first make precise our operation of planarizing matchgates, following the terminology of [9]. If an edge $\{u,v\}$ of weight w crosses t other edges, replace each of the t crossings by a crossover gadget and replace the edge by $t+1$ edges connecting adjacent crossover gadgets. Of these $t+1$ edges, assign t of them to have weight 1 and the remaining one to have weight w. Call the union of the $t+1$ edges the *u-v-passage.*

Given a non-planar matchgate Γ, denote the matchgate obtained from planarizing Γ by Γ'.

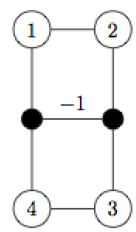


Figure A.1: Crossover gadget. Unlabeled edges are of weight 1; labeled vertices are external nodes. Figure from [9].

Observation 11. Let M be a perfect matching of Γ' whose contribution c to $\text{PerfMatch}(\Gamma')$ is nonzero, and let $M' \subset M$ denote the edges not belonging to crossover gadgets. Then M' is the union of u-v-passages corresponds to a perfect matching of Γ whose contribution to $\text{PerfMatch}(\Gamma)$ is $\pm c$.

Proof. If an edge incident to one of the external nodes, say node 1, of a crossover gadget is present in M, then the edge incident to node 3 of the crossover gadget must be present in M as well, as $\underline{X}^\sigma = 0$ if $\sigma_1 \neq \sigma_3$. We conclude that M is a union of u-v-passages. The corresponding perfect matching of Γ has contribution $\pm c$ because each of the nonzero entries of $\underline{X}$ is ± 1. □

We need to verify that the entries of $\underline{\Gamma}$ and $\underline{\Gamma'}$ are equal except for a select number of entries which differ by a factor of -1.

Lemma A.6.1. *Let Γ be the K-input, K-output transducer shown in Figure 3.2b with signature $L_3^{j,k}$. There exists a planar matchgate whose standard signature agrees with $\underline{\Gamma}$ on the main diagonal entries and entry $(1^K \oplus e_j \oplus e_k, 1^K)$, and agrees with $\underline{\Gamma}$ everywhere else up to sign.*

Proof. Take the desired matchgate to be Γ'. Note that every subgraph of Γ has at most one perfect matching. In other words, each entry of $\underline{\Gamma}$ arises from at most a single perfect matching. Therefore, by Observation 11, $\underline{\Gamma}$ and $\underline{\Gamma'}$ agree everywhere up to sign. Now consider any main diagonal entry $\underline{\Gamma'}^\sigma_\sigma = \text{PerfMatch}(\Gamma' \backslash Z)$. If M is a perfect matching of $\Gamma' \backslash Z$ making a nonzero contribution to $\text{PerfMatch}(\Gamma' \backslash Z)$, it corresponds to a perfect matching of $\Gamma \backslash Z$ making a nonzero contribution to $\text{PerfMatch}(\Gamma \backslash Z)$. But the only such perfect matching does not contain the edge between left node j to and left node k. Thus, the contribution of this matching and that of M are both equal to 1.

If $\underline{\Gamma}$ and $\underline{\Gamma'}$ disagree on entry $(1^K \oplus e_j \oplus e_k, 1^K)$, modify Γ by multiplying the weight of the edge connecting left nodes j and k by -1, and take the desired matchgate to be the corresponding Γ'. □

Lemma A.6.2. *Let Γ be the K-input, K-output transducer shown in Figure 3.2c with signature L_4. There exists a planar matchgate whose standard signature agrees with $\underline{\Gamma}$ everywhere up to sign.*

Proof. Take the desired matchgate to be Γ'. As in the proof of Lemma A.6.1, every subgraph of Γ has at most one perfect matching, so we already know $\underline{\Gamma}$ and $\underline{\Gamma'}$ agree everywhere up to sign. □

Lemma A.6.3. *Let Γ be the K-input, K-output transducer shown in Figure 3.2d with signature L_5^j. There exists a planar matchgate whose standard signature agrees with $\underline{\Gamma}$ on the main diagonal entries and entry $(1^K \oplus e_j, 1^K \oplus e_{q_{i+1}})$, and agrees with $\underline{\Gamma}$ everywhere else up to sign.*

Proof. Take the desired matchgate to be Γ'. As in the proof of Lemma A.6.1, every subgraph of Γ has at most one perfect matching and $\underline{\Gamma}$ and $\underline{\Gamma'}$ agree on the main diagonal entries. If they disagree on entry $(1^K \oplus e_j, 1^K \oplus e_{q_{i+1}})$, modify Γ by multiplying the weight of the edge connecting left node j to right node $i+1$ by -1, and take the desired matchgate to be the corresponding Γ'. □

A.7 Basic Facts about Spinors

A.7.1 Clifford algebras and the spin representation(s)

The route we take to define spinors is from the point of view of Clifford algebras. For convenience, we will work over $\mathbb{R}$, though the results that follow are valid over all fields of characteristic not equal to 2.

Let V be a vector space of dimension n equipped with a quadratic form Q, and let B denote the polarization of Q, i.e. $B(x,y) = Q(x+y) - Q(x) - Q(y)$. Denote by $T(V)$ the tensor algebra of V.

Definition A.7.1. The *Clifford algebra* $C(V)$ associated to V is the k-algebra $T(V)/I$, where I is the ideal in T generated by elements of the form $x \otimes x - Q(x)$.

We can decompose the grading $T(V) = \bigoplus_{r=0}^{\infty} V^{\otimes r}$ as a direct sum of the even and odd gradings, denoted $T_+(V)$ and $T_-(V)$ respectively. Correspondingly, denote by $C_+(V)$ the algebra $T_+(V)/(T_+(V) \cap I)$, and $C_-(V)$ similarly.

We begin with some elementary observations. First, pick an orthogonal basis $\{e_1, ..., e_n\}$ for $C(V)$ such that $B(e_i, e_j) = 0$ for $i \neq j$ and $B(e_i, e_i) = Q(e_i)$. Equivalently, we have that $e_i^2 = Q(e_i)$ and $e_i e_j + e_j e_i = 0$. The map sending $e_{i_1} \cdots e_{i_k}$ to $e_{i_1} \wedge \cdots \wedge e_{i_k}$ for all $\{i_1, ..., i_k\} \subset [n]$ gives the following:

Lemma A.7.2. *$C(V)$ and ΛV are naturally isomorphic as vector spaces.*

The following is a fairly standard fact.

Lemma A.7.3. *Define $C^{p,q}$ to be the Clifford algebra of a vector space W equipped with the quadratic form $Q(v) = \sum_{i=1}^{p} v_i^2 - \sum_{j=p+1}^{q} v_j^2$. Every real Clifford algebra $C(V)$ is isomorphic to some $C^{p,q}$ for $p+q=n$, where $\max(p,q)$ is called the* isotropy index*.*

We are now ready to define the spin representation.

Lemma A.7.4. *If n is even, then $C(V)$ is a central simple algebra. If n is odd, then $C_+(V)$ is a central simple algebra.*

Proof. First assume that n is even and say that $C(V) = C^{p,q}$ for $p < q$, picking generators $e_1, ..., e_n$ as above but assuming that we've diagonalized Q in such a way that $\epsilon_{n-1} \neq \epsilon_n$. Define elements $f_1 = e_1 \cdots e_{n-1}$ and $f_2 = e_1 \cdots e_{n-2} e_n$; note that these satisfy $f_1 f_2 + f_2 f_1 = 0$ and $f_1^2 = -f_2^2$ and thus generate $C^{1,1} \simeq M(2)$. So $C^{p,q} = C^{p-1,q-1} \otimes M(2) = C^{0,q-p} \otimes M(2^p)$, and it remains to show that $C^{0,q-p}$ is central and simple. But this follows by Bott periodicity, by which $C^{0,2m+8} = C^{0,2n} \otimes M(16)$.

Next assume that n is odd. Define elements $g_i = e_1 e_i$ for $1 < i < n$; note that these satisfy $g_i g_j + g_j g_i = 0$ and $g_i^2 = -\epsilon_1 \epsilon_i$. So if $C(V) = C^{p,q}$, then $C_+(M)$ is isomorphic to $C^{p,q-1}$ or $C^{p-1,q}$, which we know from the case of n even to be central simple. □

Now let G denote the orthogonal group of transformations preserving B.

Definition A.7.5. The **Clifford group** Γ of G is the set of $s \in C(M)$ for which conjugation of any $x \in M$ gives something inside M. Let χ be the linear representation of Γ sending s to this conjugation map; call this the **vector representation of** Γ.

In fact, χ sends Γ not just into $\mathrm{End}(C(M))$ but into G: $Q(sxs^{-1}) = (sxs^{-1})^2 = sx^2s^{-1} = Q(x)$. One can show that if the rank of Q is even (resp. odd), $\chi(\Gamma)$ is all of G (resp. the even-graded elements of G with determinant one), though we'll omit this proof.

For n even, because $C(V)$ is simple, there is a single simple representation up to isomorphism, and we call this the *spin representation*. For n odd, because $C_+(V)$ is simple, there is likewise a unique spin representation of $C_+(V)$.

Lemma A.7.6. *If $C(V)$ is not simple, there are exactly two ways to extend the spin representation ρ_+ of $C_+(V)$ to one of $C(V)$.*

Proof. It is enough to prove the following claim.

Claim A.7.7. *If V is of odd dimension $n = 2m+1$, then the center of $C(V)$ is $k[\epsilon]$ for some odd element ϵ for which $\epsilon^2 = (-1)^{m+q}$.*

Proof. For $I \subset [n]$, define $e_I = \prod_{i\in I} e_i$. If $i \in I$, then $e_I e_i = (-1)^{|I|+1} e_i e_I$; otherwise $e_I e_i = (-1)^{|I|} e_i e_I$. Given an element $x \neq 1$ in the center of $C(V)$, write it as $\sum_I a_I e_I$. Then the fact that $xe_j = e_j x$ for all j implies that $a_I = 0$ for all I except $I = [n]$, so take $\epsilon = e_1 \cdots e_n$. It is straightforward to verify that $\epsilon^2 = (-1)^{m+q}$. □

Because $C(V)$ is not simple so that ϵ^2 must be a square, and because we are working over $\mathbb{R}$, $\epsilon^2 = 1$. To finish the proof of the lemma, we can now write any $x \in C(V)$ uniquely as $x_1 + x_2\epsilon$ for $x_1, x_2 \in C_+(V)$. The maps $x \mapsto x_1 + x_2$ and $x \mapsto x_1 - x_2$ are homomorphisms $\phi_1, \phi_2 : C(V) \to C_+(V)$. Now simply post-compose these with ρ^+ to obtain two distinct representations ρ_1, ρ_2 of $C(M)$. Suppose we had another representation of $C(M)$ extending ρ_+, call it ρ' and denote $\rho'(\epsilon)$ by σ. We know that σ^2 is the identity map on the space S of spinors, so S decomposes into eigenspaces of eigenvalue 1 or -1 with respect to σ. These eigenspaces are preserved under the action of C_+, so one of them must be zero for ρ_+ to be simple. We conclude that $\sigma = \pm I$ and thus that ρ' agrees with either ρ_1 or ρ_2. □

A.7.2 Pure spinors

The particular kinds of spinors which form the bridge between geometry and holographic algorithms are the *pure spinors*, which will correspond to maximal linear subvarieties of the quadric $Q = 0$. We will assume henceforth that Q is of maximal isotropic index and that, for simplicity, $n = 2m$ is even. Pick a splitting of V into maximal totally isotropic subspaces $N \oplus P$ and denote the spin representation by ρ.

Lemma A.7.8. *The intersection between a minimal left ideal $\mathfrak{a}$ and a minimal right ideal $\mathfrak{b}$ of $C(V)$ is a one-dimensional vector space.*

Proof. A theorem due to Brauer tells us that there exist idempotents e, e' for which $\mathfrak{a}$ and $\mathfrak{b}$ are respectively generated in $C(V)$ by e and e'. Because e, e' are idempotents, $\rho(e), \rho(e')$ are projections.

We claim that $H := \ker\rho(e)$ is a hyperplane. It is clear that $\mathfrak{a}$ kills H, and conversely any element $x \in C(V)$ which kills H is equal to $x \cdot e$, as $\rho(e)$ acts as the identity on $\mathrm{im}\rho(e)$. So $\mathfrak{a}$ is precisely the annihilator of H, and H is thus a hyperplane by minimality of $\mathfrak{a}$.

Likewise, we claim that $D := \mathrm{im}\rho(e')$ is a line. It is clear that $\mathfrak{b}$ sends D to itself, and conversely any element $x \in C(V)$ which sends S to D is equal to $e' \cdot x$, as e' acts as the identity on D. So $\mathfrak{b}$ is precisely the set of elements sending S to D, and D is thus a line by minimality of $\mathfrak{a}$.

To summarize, $\mathfrak{a} \cap \mathfrak{b}$ is the set of $x \in C(V)$ killing H and projecting S into D. Picking a generator y for D, we see that for $v \notin H$, $\rho(x) \cdot v = \lambda y$, so x is uniquely determined by λ, and $\mathfrak{a} \cap \mathfrak{b}$ is indeed a one-dimensional vector space. □

Let Z be any other maximal totally isotropic subspace of V, and say Z generates the subalgebra C^Z of $C(V)$. Define f and f_Z to be products of the basis elements of P and Z, respectively; by Lemma A.7.2, f and f_Z are uniquely defined up to a multiplicative factor.

Lemma A.7.9. *$C(V) \cdot f$ and $f_Z \cdot C(V)$ are minimal left and right ideals respectively.*

So by Lemma A.7.8, these two ideals intersect on some line $S_Z \cdot f$, and we call $S_Z \subset C(V)$ the line of *representative spinors of Z*.

Definition A.7.10. A *pure spinor* is any spinor representative of some maximal totally isotropic subspace Z.

It turns out these spinors are "representative" in the following sense.

Proposition A.7.11 ([16])**.** *If u_Z is a representative spinor of Z, then Z is precisely the set of elements in V which kill u_Z.*

Proof. Recall that for M of even dimension, $\chi(\Gamma) = G$, meaning there must exist some $s \in \Gamma$ by which P can be conjugated to obtain Z. As such, we'll assume that $Z = P$, in which case the corresponding line of pure spinors contains $u_Z = 1$. But $\rho(x)$ for any x is multiplication by the component of x inside N, and this map is zero iff this component is zero, i.e iff $x \in P = Z$. □

There is therefore a one-to-one correspondence between maximal totally isotropic subspaces of V and lines of pure spinors.

A.8 Lifting Assumptions on U_0 and U_1

We proceed by casework on the polarization $\partial_{1,1,1}(U_\alpha, U_\beta, U_m)$ that U_m appears in as well as the smallest k for which $\partial_{1,1,1}(U_0, U_\gamma, V)$ is not zero for all V on which $\partial_{1,1,1}(U_\alpha, U_\beta, V)$ vanishes.

First, we note the following general fact:

Lemma A.8.1. *If $\partial_{2,1}(U_0, U_i) = 0$ for all $0 \leq i \leq m$, then* $\operatorname{rank}(U_0)$.

Proof. Suppose to the contrary that U_0 contains matrices of rank at least 2. The vanishing of $\partial_{2,1}(U_0, U_i)$ for each i imposes the same nontrivial linear condition on each U_i, so $U_0 \oplus \cdots \oplus U_m \neq W$. □

After row/column operations, we may assume that $U_0 \subset \begin{pmatrix} a & 0 & 0 \\ b & 0 & 0 \\ c & 0 & 0 \end{pmatrix}$.

A.8.1 $\alpha = 0$

Lemma A.8.2. *If $\dim(U_0) = 3$ and $\partial_{2,1}(U_0, U_m) = 0$, then either $V(q)$ is a cone or q lies in the orbit of det_3.*

Proof. If $\dim(U_0) = 3$, pick the complement of U_0 in $\mathbb{C}^9$ to be the space of matrices for which the leftmost column is zero; $U_1, \ldots, U_m$ all lie in this space. We conclude that no term of the form $\partial_{1,1,1}(U_i, U_j, U_k)$ for $i \geq 1$ appears in q, and q consists at most of $\partial_{1,1,1}(U_0, U_1, U_m)$ and $\partial_{1,1,1}(U_0, U_2, U_{m-1})$.

□

Lemma A.8.3. *If $\dim(U_0) = 2$ and $\partial_{2,1}(U_0, U_m) = 0$, then q gives rise to no new boundary components. In particular, q consists at most of $\partial_{1,1,1}(U_0, U_1, U_m)$ and either $\partial_{1,1,1}(U_0, U_2, U_{m-1})$ or $\partial_{2,1}(U_1, U_{m-1})$ for some i, j.*

Proof. The vanishing of $\partial_{1,2}(U_0, U_1)$ forces either 1) the middle and rightmost columns of U_1 to be linearly dependent, or 2) $(U_1)^3_2$ and $(U_1)^3_3$ to vanish.

Suppose 1) holds. After applying row/column operations, we may assume that the middle column of U_1 is zero. We will show that no polarization of the form $\partial_{1,1,1}(U_1, U_i, U_j)$ can appear in q. Suppose such a term appeared for which $1 < i < j$. Then $\partial_{1,1,1}(U_0, U_1, V) = 0$ forces the rightmost column of V to be zero unless the entries of the middle column of U_1 are all linearly dependent, i.e. $((U_1)^1_2, (U_1)^2_2, (U_1)^3_2) = x \cdot (1, \lambda, \mu)$ for some indeterminate x. In this case, the entries of the rightmost columns of U_i and U_j are likewise related. The vanishing of $\partial_{1,2}(U_1, U_i)$ then implies a) $(U_1)^3_1 = 0$, b) the rightmost column of U_i vanishes, or c) $\mu \cdot (U_i)^1_2 = (U_i)^2_2$.

If a) $(U_1)^3_1 = 0$, then $\partial_{1,1,1}(U_1, U_i, U_j) = 0$ and we're done. If b) the rightmost column of U_i vanishes or c) $\mu \cdot (U_i)^1_2 = (U_i)^2_2$, then $\partial_{1,1,1}(U_0, U_i, U_j) = 0$ implies either the rightmost column of U_j vanishes as well, or the middle column of U_i satisfies the same relations as does that of U_1. In either case, $\partial_{1,1,1}(U_1, U_i, U_j)$ again vanishes.

The same analysis can be carried out if $\partial_{1,2}(U_1, U_i)$ appears in q for $i > 1$. Assuming the entries of the middle column of U_1 are all linearly dependent as before, we see that the vanishing of $\partial_{1,2}(U_0, U_i)$ implies that of $\partial_{1,2}(U_1, U_i)$.

Likewise, if any polarization $\partial_{2,1}(U_1, U_i)$ appears in q, the vanishing of $\partial_{1,1,1}(U_0, U_1, U_i)$ either forces the rightmost column of U_i to vanish, or at least forces $(U_i)^1_3$ and $(U_i)^2_3$ to satisfy the same relation that $(U_1)^1_2$ and $(U_1)^2_2$ do, which is enough for $\partial_{2,1}(U_1, U_i)$ to vanish.

We conclude that if 1) holds, q at most consists of $\partial_{1,1,1}(U_0, U_1, U_m)$ and $\partial_{1,1,1}(U_0, U_i, U_j)$ for some i, j.

Now suppose 2) holds but 1) does not. If q contains some term of the form $\partial_{1,1,1}(U_1, U_i, U_j)$ for $1 < i \leq j$, then $\det(U_1) = 0$ implies $(U_1)^3_1 = 0$. $\partial_{2,1}(U_1, U_i) = \partial_{2,1}(U_1, U_j) = 0$ implies $(U_i)^3_1 = (U_j)^3_1 = 0$. Finally, note that $\partial_{1,1,1}(U_0, U_1, V) = 0$ forces $V^3_2 = V^3_3 = 0$ so that q does not contain more than one term containing U_0. We conclude that q at most consists of $\partial_{1,1,1}(U_0, U_1, U_m)$ and $\partial_{2,1}(U_1, U_j)$. □

The remainder of this section is devoted to proving the hardest case, namely when $\dim(U_0) = 1$.

Theorem A.8.4. *If* $\dim(U_0) = 1$ *and* $\partial_{2,1}(U_0, U_m) = 0$*, then* q *gives rise to no new boundary components.*

After row/column operations, we may assume that $U_0 = \begin{pmatrix} a & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$. To prove Theorem A.8.4 will proceed by casework on the β for which $\partial_{2,1}(U_0, U_\beta, U_m)$ appears in q.

Case 1. $\beta = 1$.

Proof. Suppose that $\partial_{1,1,1}(U_1, U_a, U_b)$ appears for $a, b \geq 2$. We will show that $V(q)$ is a cone or q is a $\mathrm{GL}(W)$-translate of $\det_3$. $\partial_{1,2}(U_0, U_1) = 0$ implies that in U_1, $\begin{pmatrix} x_{22} & x_{32} \\ x_{32} & x_{33} \end{pmatrix}$ is singular, so after simultaneous row/column operations on $U_0, ..., U_m$, we may assume $x_{32} = x_{33} = 0$ in U_1. $\det(U_1) = 0$ then implies either 1) the last column of U_1 is zero, or 2) $M := \begin{pmatrix} x_{21} & x_{22} \\ x_{31} & x_{32} \end{pmatrix}$ is singular. Assume first that there are no no linear dependencies between x_{22} and x_{33} in U_1. Then if 1) holds, the vanishing of $\partial_{1,1,1}(U_0 \oplus U_1, U_1, V)$ forces the rightmost column of V to be zero, so $\partial_{1,1,1}(U_1, U_a, U_b) = 0$, a contradiction. If 2) holds so that there is some linear dependence between the columns of M in U_1, the vanishing of $\partial_{1,1}(U_0, U_1, V)$ forces $x_{23} = x_{33} = 0$ in V, and the vanishing of $\partial_{2,1}(U_1, V)$ then forces such a V to have the same linear dependence as U_1 between the columns of M. We conclude again that $\partial_{1,1,1}(U_1, U_a, U_b) = 0$.

Finally, assume $(U_1)_2^2 = \lambda \cdot (U_1)_2^3$ for some $\lambda \in \mathbb{C}$. Then in order for $\det(U_1)$ to vanish, either 1) holds, $(U_1)_3^2 = (U_1)_3^3 = 0$, or $(U_1)_1^2 = \lambda \cdot (U_1)_1^3$. In any case, $\partial_{1,1,1}(U_0, U_1, V) = 0$ implies that $V_3^2 = \lambda \cdot V_3^3$ for $V = U_a, U_b$, and $\partial_{1,2}(U_0, U_a)$ then implies that either $(U_a)_3^2 = (U_a)_3^3 = 0$ or $(U_a)_2^2 = \lambda (U_a)_2^3$.

We claim that either every term in q omits the variable W_2^1 or W_1^2 and W_1^3 only ever appears together in the linear factor $W_2^1 - \lambda \cdot W_1^2$. This is trivially true for the polarizations in q containing U_0.

Indeed, it is straightforward to check that this is the case for all polarizations in q when 1) holds because of the vanishing of $\partial_{2,1}(U_1, V)$ for $V = U_a, U_b$. On the other hand, if 1) doesn't hold and $(U_1)_3^2 = (U_1)_3^3 = 0$, then $\partial_{1,1,1}(U_0, U_1, V)$ is identically zero, a contradiction. If neither of these scenarios holds and $(U_1)_1^2 = \lambda \cdot (U_1)_1^3$, then the vanishing of $\partial_{2,1}(U_1, V)$ for $V = U_a, U_b$ implies every term in q omits the variable W_2^1.

In conclusion, q would at most contain $\partial_{1,1,1}(U_0, U_1, U_m)$, $\partial_{1,1,1}(U_0, U_\alpha, U_\beta)$ and $\partial_{2,1}(U_1, U_\gamma)$. One can check that these terms completely factor; for $V(q)$ not to be a cone, $U_0 \oplus \cdots \oplus U_m \leq 1 + 3 + 3$, but then $U_0 \oplus \cdots \oplus U_m \neq W$ a contradiction. ☐

Case 2. $\beta = 2$.

Proof. $\partial_{1,1,1}(U_0, U_1, V) \equiv 0$ implies that $U_1 \in U_2^{cmp}$. We can assume that $U_1, ..., U_m$ lie in the complement of U_0 for which $(U_i)_1^1 = 0$, and that $U_0 = \begin{pmatrix} a & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$.

Pick the complement of U_0 in $\mathbb{C}^9$ to be the space of matrices (x_{ij}) for which $x_{11} = 0$; $U_0, ..., U_{m-1}$ all lie in this space. Then the vanishing of $\partial_{1,2}(U_0, U_2)$ implies that the minor $\left| \begin{pmatrix} (U_2)_2^2 & (U_2)_3^2 \\ (U_2)_2^3 & (U_2)_3^3 \end{pmatrix} \right|$ is zero. After simultaneous row/column operations on $U_0, ..., U_m$, we may assume further that $(U_2)_2^3 = (U_2)_3^3 = 0$.

Lemma A.8.5. *If* $\dim(U_1) = 1$ *and* $\partial_{2,1}(U_0, W)$ *is identically zero, then* q *omits a variable of* U_1*.*

Proof. We may assume U_1 is of the form $\begin{pmatrix} 0 & W_2^1 & 0 \\ \lambda W_2^1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$, where $x = W_2^1$. We claim that q must omit the variable W_1^2. This is certainly the case for a polarization of the form $\partial_{2,1}(U_1, U_i)$, or any polarization in q containing U_0. If $\partial_{1,1,1}(U_i, U_j, U_k)$ appears in q for $1 \leq i < j \leq k$, the vanishing of $\partial_{2,1}(U_1, V)$ and $\partial_{1,1,1}(U_0, U_2, V)$ for $V = U_i, U_j, U_k$ implies that $\partial_{1,1,1}(U_1, U_j, U_k)$ also omits the variable W_1^2, completing the proof of our claim. ☐

Lemma A.8.6. *If* $\dim(U_1) \geq 2$*, then either* q *omits a variable or consists of at most three terms:* $\partial_{1,1,1}(U_0, U_2, U_m)$*,* $\partial_{1,1,1}(U_1, U_i, U_j)$*, and* $\partial_{1,1,1}(U_2, U_k, U_\ell)$ *for some* i, j, k, ℓ*.*

Proof. We will assume

$$U_1 = \begin{pmatrix} 0 & W_2^1 & 0 \\ W_1^2 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \tag{A.1}$$

. Indeed, if U_1 had additional nonzero entries, the vanishing of $\partial_{2,1}(U_0 \oplus U_1, V)$ would impose two linear conditions, and as we saw in Lemma 5.4.9, there would be at most two polarizations in q of the form $\partial_{2,1}(U_1, U_j, U_k)$, and no polarizations of the form $\partial_{2,1}(U_i, U_j, U_k)$ for $i \geq 2$. One could then readily check that either $V(q)$ is a cone or q is a GL(W)-translate of $\det_3$.

Returning to (A.1), the argument now is similar to that of Lemma A.8.5. Firstly, the vanishing of $\partial_{1,2}(U_1, U_2)$ implies that either $(U_2)_3^2$ or $(U_2)_1^3$ must be zero.

If $(U_2)_1^3 = 0$, then $\partial_{1,1,1}(U_0, U_2, V) = 0$ for $V = U_0, ..., U_{m-1}$ forces either both $(U_i)_2^3$ and $(U_i)_3^3$ to vanish for all i, or some linear combination of them to. If the former, any other polarization containing U_0 would vanish, any term of the form $\partial_{1,1,1}(U_1, U_i, U_j)$ would omit W_2^1. If however $\partial_{1,1,1}(U_0, U_2, V) = 0$ imposed some condition $\ell = 0$ on $V = U_0, ..., U_{m-1}$, this would mean there is a linear dependence between $(U_2)_2^2$ and $(U_2)_3^2$. After simultaneous row/column operations on $U_0, ..., U_{m-1}$, we can assume $(U_2)_2^2 = 0$ so that $\ell = (U_i)_2^3$. One can then check that q omits W_2^2, and we're done.

On the other hand, if $(U_2)_3^2 = 0$, we claim that q omits either W_1^2 or W_2^1. Indeed, $\partial_{1,1,1}(U_0, U_2, V) = 0$ and $\partial_{2,1}(U_1, V)$ impose the same single condition on V, so no term of the form $\partial_{2,1}(U_1, U_i)$ appears in q. If some $\partial_{1,2}(U_1, U_j)$ appears in q, $\partial_{1,2}(U_0, U_j) = 0$ forces either $(U_j)_2^3$ or $(U_j)_3^2$ to vanish, either of which would force $\partial_{1,2}(U_1, U_j)$ to omit one of W_1^2 or W_2^1. If some $\partial_{1,1,1}(U_1, U_2, U_k)$ appeared in q, then $\det(U_2) = 0$ forces $(U_2)_1^3$, $(U_2)_2^2$, or $(U_2)_3^1$ to vanish. The first or last would imply $\partial_{1,1,1}(U_1, U_2, U_k)$ omits one of W_1^2 or W_2^1. $(U_2)_2^2 = 0$ would imply $\partial_{1,1,1}(U_0, U_2, W)$ is identically zero, a contradiction.

Finally, if some $\partial_{1,1,1}(U_1, U_j, U_k)$ appeared in q for $j > 2$, then $\partial_{1,2}(U_0 \oplus U_1, U_j) = 0$ forces one of the following pairs to vanish: $(U_j)_3^1$ and $(U_j)_3^2$, $(U_j)_3^2$ and $(U_j)_2^3$, or $(U_j)_1^3$ and $(U_j)_2^3$. The first or last would force $\partial_{1,1,1}(U_1, U_j, U_k)$ to omit one of W_2^1 or W_1^2.

It remains to handle the exceptional case where $(U_j)_3^2 = (U_j)_2^3 = 0$. Here we claim that q still gives rise to no new boundary components.. Firstly note that $\partial_{2,1}(U_i, U_j) = 0$ for all $i \leq 2$. Suppose some $\partial_{1,1,1}(U_0, U_k, U_\ell)$ and $\partial_{1,1,1}(U_1, U_i, U_j)$ appeared in q for $k > 2$. If $k \geq j$, then the vanishing of $\partial_{1,2}(U_0, U_k)$ forces $\partial_{1,1,1}(U_1, U_i, U_j)$ to omit one of W_2^1 or W_1^2. If $k < j$, then the vanishing of $\partial_{1,1,1}(U_1, U_i, U_k)$ forces the same. We conclude that q contains at most three terms: $\partial_{1,1,1}(U_0, U_2, U_m)$, $\partial_{1,1,1}(U_1, U_i, U_j)$, and $\partial_{1,1,1}(U_2, U_k, U_\ell)$ for some i, j, k, ℓ, and one readily checks that $V(q)$ is a cone or q is a GL(W)-translate of $\det_3$ in this case. ☐

☐

Case 3. $\beta > 2$.

Proof. This case follows easily from the previous one. If $\dim(U_1) = 1$ or $\dim(U_2) = 1$, then q will omit a variable from U_1 or U_2, by Lemma A.8.5. Therefore, because $U_1, U_2 \subset U_2^{cmp}$, $\dim(U_1) = \dim(U_2) = 2$, i.e. $U_0 \oplus U_1 \oplus U_2 = U_2^{cmp}$. But then V for which $\partial_{2,1}(U_0 \oplus U_1 \oplus U_2, V) = 0$ must lie in U_2^{cmp}, so q contains no polarizations $\partial_{1,1,1}(U_i, U_j, U_k)$ for $i \geq 2$, except possibly some $\partial_{2,1}(U_2, U_k)$. As we saw in the proof of Lemma A.8.6, if q contains some $\partial_{1,1,1}(U_1, U_j, U_k)$ and does not omit any variables, then the only polarization in q that contains U_0 is $\partial_{1,1,1}(U_0, U_\beta, U_m)$. So as in Lemma A.8.6, the only terms that q contains if it does not omit any variables are $\partial_{1,1,1}(U_0, U_\beta, U_m)$, $\partial_{1,1,1}(U_1, U_i, U_j)$, and $\partial_{1,1,1}(U_2, U_k, U_\ell)$ for some i, j, k, ℓ, so we're done. ☐

A.8.2 $\alpha = 1$

Case 1. $\beta = 1$.

Proof. Suppose that $\gamma = 2$. Then we are effectively in the setting of Case 2. In particular, if $\dim(U_1) = 1$, the proof of Lemma A.8.5 exactly carries over to show that q omits one of W_2^1 or W_1^2. If $\dim(U_2) = 2$, then if $(U_2)_1^3 = 0$, the proof of Lemma A.8.6 also exactly carries over to show that q omits W_2^2. If $(U_2)_3^2 = 0$ however, we have to slightly modify the argument. In this case, recall that $\partial_{1,1,1}(U_0, U_2, V) = 0$ and $\partial_{2,1}(U_1, V) = 0$ impose the same single linear condition on V. Whereas in the proof of Lemma A.8.6 we could then conclude that q contains no term of the form $\partial_{2,1}(U_1, U_i)$, here we know that q contains no term of the form $\partial_{1,1,1}(U_0, U_2, U_i)$. Instead of showing that q either omits W_1^2 or W_2^1 from U_1 or consists of the terms $\partial_{1,1,1}(U_0, U_\beta, U_m)$, $\partial_{1,1,1}(U_1, U_i, U_j)$, and $\partial_{1,1,1}(U_2, U_k, U_\ell)$ for some i, j, k, ℓ, one can show that q omits W_2^2 or consists of the terms $\partial_{1,1,1}(U_1, U_1, U_m)$, $\partial_{1,1,1}(U_0, U_i, U_j)$, and $\partial_{1,1,1}(U_2, U_k, U_\ell)$.

Now suppose that $\gamma > 2$. As in the proof for Case 3, we can assume that $U_0 \oplus U_1 \oplus U_2 = U_2^{cmp}$, and the analysis for $\gamma = 2$ above likewise carries over to the case of $\gamma > 2$. □

Case 2. $\beta \geq 2$.

Proof. By Lemma A.8.3 applied to $U_0 \oplus U_1$, we conclude that the only terms in q that do not contain U_0 are $\partial_{1,1,1}(U_1, U_2, U_m)$ and either $\partial_{1,1,1}(U_1, U_3, U_{m-1})$ or $\partial_{2,1}(U_2, U_{m-1})$. In particular, β cannot exceed 2.

We first note that only one polarization in q can contain U_0. Suppose to the contrary there were two, $\partial_{1,1,1}(U_0, U_i, U_j)$ and $\partial_{1,1,1}(U_0, U_k, U_\ell)$, where $i < k$. Then taking U_0 in Lemma A.8.3 to be $U_0 \oplus U_1$, we see that $\partial_{1,1,1}(U_0 \oplus U_1, U_2, V) = 0$ for $V = U_k, U_\ell$ forces U_k, U_ℓ to have rightmost columns with entries all linearly dependent. $\partial_{1,1,1}(U_0, U_i, V) = 0$ for $V = U_k, U_\ell$ then either implies that their rightmost columns vanish, or that $V_2^2 = V_2^3 = 0$. Either forces $\partial_{1,1,1}(U_0, U_k, U_\ell)$ to vanish.

We conclude that q consists at most of $\partial_{1,1,1}(U_1, U_2, U_m)$, $\partial_{1,1,1}(U_0, U_i, U_j)$ for some i, j, and either $\partial_{1,1,1}(U_1, U_3, U_{m-1})$ or $\partial_{2,1}(U_2, U_{m-1})$. So either $V(q)$ is a cone or q is a $\mathrm{GL}(W)$-translate of $\det_3$. □

A.8.3 $\alpha \geq 2$

Note that α cannot exceed 2, or else $U_0 \oplus \cdots \oplus U_\alpha$ is a space of rank 1 matrices, yet $\dim(U_0 \oplus \cdots \oplus U_\alpha) > 3$, a contradiction. So take $\alpha = 2$.

Case 1. $\beta = 2$.

Proof. Just as Case 1 followed almost immediately from Case 3, so too does this case follow from Case 2. q consists of at most $\partial_{2,1}(U_2, U_m)$, $\partial_{1,1,1}(U_1, U_i, U_j)$, and $\partial_{1,1,1}(U_0, U_k, U_\ell)$ for some i, j, k, ℓ, and we can check that q gives rise to no new boundary components. □

Case 2. $\beta \geq 3$.

Proof. If U_m appears in $\partial_{1,1,1}(U_2, U_i, U_m)$ for some $i \geq 3$, then $\partial_{2,1}(U_0 \oplus U_1 \oplus U_2, W)$ is identically zero, so we may assume

$$U_0 \oplus U_1 \oplus U_2 = \begin{pmatrix} a & 0 & 0 \\ b & 0 & 0 \\ c & 0 & 0 \end{pmatrix}. \tag{A.2}$$

But then by Lemma A.8.2, q contains at most two terms with U_2, namely $\partial_{1,1,1}(U_2, U_3, U_m)$ and $\partial_{1,1,1}(U_2, U_4, U_{m-1})$. In addition, q has at most one term containing U_0 and at most one containing U_1. This can be proved with the exact same technique as in Case 2. It follows that q contains at most four terms: $\partial_{1,1,1}(U_2, U_3, U_m)$, $\partial_{1,1,1}(U_2, U_4, U_{m-1})$, $\partial_{1,1,1}(U_0, U_i, U_j)$, and $\partial_{1,1,1}(U_1, U_k, U_\ell)$, and it is easy to check that either $V(q)$ is a cone or q is a $\mathrm{GL}(W)$-translate of $\det_3$. □

Notations

$\mathbb{C}[X]$	Coordinate ring of X
$\mathbb{C}[X]_d$	Degree-d component of $\mathbb{C}[X]$
$\underline{\Gamma}$	Standard signature of matchgate Γ
$[p]$	Point in $\mathbb{P}V$ corresponding to the line through $p \in V$
$S^d W^*$	Degree-d homogeneous polynomials over W
$\mathbb{C}^*$	Units of $\mathbb{C}$
V(p)	Affine variety in W cut out by p
G	complex affine algebraic group, usually reductive
$\mathcal{M}_{m,n}(k)$	Space of matrices with entries in field k
I_n	$n \times n$ identity matrix
Γ^σ_ζ	Entry of matrix Γ in row σ and column ζ
$\mathrm{Lie}(G)$	Lie algebra associated to Lie group G
$\mathfrak{g}.x$	Induced action of Lie algebra
X_{sing}	Singular locus of X
$G \cdot X$	Orbit of X under action of G
$\overline{G \cdot X}$	Zariski closure of orbit of X
$X//G$	GIT quotient
$\mathbb{S}_n$, $\mathcal{A}_n$	Symmetric and alternating groups on n elements
$\mathrm{Wt}(G, V)$	Weights of G-module V
$\mathrm{span}(Z)$	Span of columns of a matrix M indexed by Z
$S_\pi(V)$	Irreducible $\mathrm{GL}(V)$-module associated to partition π
V^{ss}	Semi-stable points of V under action of some G
$G(k, n)$	Grassmannian of k-planes in $\mathbb{A}^n$.
$\partial_\pi f$	π-th polarization of polynomial f
$\mathbb{S}_m$	Variety of pure spinors