Holant* Dichotomy on Domain Size 3: A Geometric Perspective

Jin-Yi Cai ✉
Department of Computer Sciences, University of Wisconsin-Madison, WI, USA

Jin Soo Ihm ✉
Department of Computer Sciences, University of Wisconsin-Madison, WI, USA

Abstract

Holant problems are a general framework to study the computational complexity of counting problems. It is a more expressive framework than counting constraint satisfaction problems (CSP) which are in turn more expressive than counting graph homomorphisms (GH). In this paper, we prove the first complexity dichotomy of $\text{Holant}_3^*(\mathcal{F})$ where $\mathcal{F}$ is an arbitrary set of symmetric, real valued constraint functions on domain size 3. We give an explicit tractability criterion and prove that, if $\mathcal{F}$ satisfies this criterion then $\text{Holant}_3^*(\mathcal{F})$ is polynomial time computable, and otherwise it is #P-hard, with no intermediate cases. We show that the geometry of the tensor decomposition of the constraint functions plays a central role in the formulation as well as the structural internal logic of the dichotomy.

2012 ACM Subject Classification Theory of computation → Complexity theory and logic

Keywords and phrases Holant problem, Complexity dichotomy, Higher domain

Digital Object Identifier 10.4230/LIPIcs.ICALP.2025.148

Category Track B: Automata, Logic, Semantics, and Theory of Programming

Related Version *Full Version*: https://arxiv.org/abs/2504.14074 [12]

Acknowledgements We sincerely thank the reviewers for their thoughtful and thorough feedback.

1 Introduction

Holant problems were introduced in [14] as a broad framework to study the computational complexity of counting problems. Counting CSP is a special case of Holant problems [18, 4, 3, 19, 13, 21, 8, 5]. In turn, counting CSP includes counting graph homomorphisms (GH), introduced by Lovász [26, 24], which is a special case with a single binary constraint function. Typical Holant problems include counting all matchings, counting perfect matchings #PM (including all weighted versions), counting cycle covers, counting edge colorings, and many other natural problems. It is strictly more expressive than GH; for example, it is known that #PM cannot be expressed in the framework of GH [22, 9].

The complexity classification program of counting problems is to classify as broad a class of problems as possible according to their inherent computational complexity within these frameworks. Let $\mathcal{F}$ be a set of (real or complex valued) constraint functions defined on some domain set D. It defines a Holant problem $\text{Holant}(\mathcal{F})$ as follows. An input consists of a graph $G = (V, E)$, where each $v \in V$ has an associated $\mathbf{F} \in \mathcal{F}$, with incident edges to v labeled as input variables of $\mathbf{F}$. The output is the sum of products of evaluations of the constraint functions over all assignments over D for the variables. The goal of the complexity classification of Holant problems is to classify the complexity of $\text{Holant}(\mathcal{F})$. A complexity dichotomy theorem for counting problems classifies every problem in a broad class of problems $\mathcal{F}$ to be either polynomial time solvable or #P-hard.

© Jin-Yi Cai and Jin Soo Ihm;
licensed under Creative Commons License CC-BY 4.0

There has been tremendous progress in the classification of counting GH and counting CSP [18, 19, 13, 3, 21, 8, 5, 20, 1, 23, 7]. Much progress was also made in the classification of Holant problems, particularly on the Boolean domain ($|D| = 2$), i.e., when variables take 0-1 values (but constraint functions take arbitrary values, such as partition functions from statistical physics). This includes the dichotomy for all complex-valued symmetric constraint functions [10] and for all real-valued not necessarily symmetric constraint functions [27]. On the other hand, obtaining higher domain Holant dichotomy has been far more challenging. There is a huge increase in difficulty in proving dichotomy theorems for domain size > 2, as already seen in decision CSP of domain size 3, a major achievement by Bulatov [2]. Toward proving these dichotomies one often first considers restricted classes of Holant problems assuming some particular set of constraint functions are present. Two sets stand out: (1) the set of equality functions $\mathcal{EQ}$ of all arities (this is the class of all counting CSP problems) and (2) the set of all unary functions $\mathcal{U}$, i.e., functions of arity one. Indeed, $\#\mathrm{CSP}(\mathcal{F}) = \mathrm{Holant}(\mathcal{F} \cup \mathcal{EQ})$; i.e., counting CSP are the special case of Holant problems with $\mathcal{EQ}$ assumed to be present. In this paper we study (2): $\mathrm{Holant}_3^*(\mathcal{F}) := \mathrm{Holant}_3(\mathcal{F} \cup \mathcal{U})$, for an arbitrary set $\mathcal{F}$ of symmetric real-valued constraint functions on domain size 3.

Previously there were only two significant Holant dichotomies on higher domains. One is for a single ternary constraint function that has a strong symmetry property called domain permutation invariance [11]. That work also solves a decades-old open problem of the complexity of counting edge colorings. The other is a dichotomy for $\mathrm{Holant}_3^*(f)$ where f is a single symmetric complex-valued ternary constraint function on domain size 3 [15]. Extending this dichotomy to an arbitrary constraint function, or more ambitiously, to a set of constraint functions has been a goal for more than 10 years without much progress.

In this paper, we extend the result in [15] to an arbitrary set of real-valued symmetric constraint functions. In [25] an interesting observation was made that an exceptional form of complex-valued tractable constraint functions does not occur when the function is real-valued. By restricting ourselves to a set $\mathcal{F}$ of real-valued constraint functions, we can bypass a lot of difficulty associated with this exceptional form. Another major source of intricacy is related to the interaction of binary constraint functions with other constraint functions in $\mathcal{F}$. We introduce a new geometric perspective that provides a unifying principle in the formulation as well as a structural internal logic of what leads to tractability and what leads to #P-hardness. After discovering some new tractable classes of functions aided by the geometric perspective, we are able to prove a $\mathrm{Holant}_3^*(\mathcal{F})$ dichotomy. This dichotomy is dictated by the geometry of the tensor decomposition of constraint functions.

Suppose $\mathbf{G}$ is a binary constraint function and $\mathbf{F}$ is a tenary constraint function, with $\mathbf{F} = \mathbf{u}^{\otimes 3} + \mathbf{v}^{\otimes 3}$ its tensor decomposition. One of the simplest constructions possible with $\mathbf{G}$ and $\mathbf{F}$ is to connect $\mathbf{G}$ at the three edges of $\mathbf{F}$; the resulting constraint function is $\mathbf{G}^{\otimes 3}\mathbf{F}$ which has tensor decomposition $(\mathbf{Gu})^{\otimes 3} + (\mathbf{Gv})^{\otimes 3}$. We see that this gadget construction plays nicely with the tensor decomposition. Generalizing this idea, suppose $\mathcal{B}$ is a set of binary constraint functions and $\mathcal{T}$ is a set of ternary constraint functions. Let $\langle \mathcal{B} \rangle$ be the monoid generated by $\mathcal{B}$. We may consider the orbit $\mathcal{O}$ of $\mathcal{T}$ under the monoid action of $\langle \mathcal{B} \rangle$, such that $\mathbf{G} \in \langle \mathcal{B} \rangle$ acts on $\mathbf{F} \in \mathcal{T}$ by $\mathbf{G} : \mathbf{F} \mapsto \mathbf{G}^{\otimes 3}\mathbf{F}$. Although the constraint functions in $\mathcal{O}$ are the results of a very simple gadget construction, we show that $\mathcal{O}$ contains sufficient information about the interaction of binary constraint functions and other constraint functions, and the simplicity allows us to analyze it by considering the geometry of the vectors of the tensor decomposition of the constraint functions in $\mathcal{O}$.

Compared to the Boolean domain dichotomy theorem, stated in explicit recurrences on the values of the signatures (see Theorem 2.12 in [6]) the dichotomy theorem (Theorem 3.1) we wish to prove has a more non-explicit form, which is also more conceptual. This is

$$\begin{array}{ccccccc} & & & f_{\mathsf{BBB}} & & & \\ & & f_{\mathsf{BBG}} & & f_{\mathsf{BBR}} & & \\ & f_{\mathsf{BGG}} & & f_{\mathsf{BGR}} & & f_{\mathsf{BRR}} & \\ f_{\mathsf{GGG}} & & f_{\mathsf{GGR}} & & f_{\mathsf{GRR}} & & f_{\mathsf{RRR}} \end{array}$$

Figure 1 Notation for expressing a symmetric ternary domain 3 constraint functions. This notation can be extended for higher arity signatures by using a larger triangle.

$$\begin{array}{ccccccc} & & & 2 & & & \\ & & 2 & & -1 & & \\ & 2 & & -1 & & 5 & \\ 2 & & -1 & & 5 & & -7 \end{array}$$

(a) $\mathbf{F}_1$

$$\begin{array}{ccccccc} & & & -7 & & & \\ & & 5 & & -1 & & \\ & 5 & & 5 & & 2 & \\ -7 & & -1 & & 2 & & 2 \end{array}$$

(b) $\mathbf{G}_1$

$$\begin{array}{ccccccc} & & & -2 & & & \\ & & 1 & & 1 & & \\ & 1 & & -2 & & 1 & \\ -2 & & 1 & & 1 & & -2 \end{array}$$

(c) $\mathbf{H}_1$

$$\begin{array}{ccccc} & & 3+2\sqrt{2} & & \\ & -3+2\sqrt{2} & & -\sqrt{2} & \\ 3+2\sqrt{2} & & -\sqrt{2} & & -4\sqrt{2} \end{array}$$

(d) $\mathbf{B}_1$

Figure 2 Ternary constraint functions $\mathbf{F}_1$, $\mathbf{G}_1$, $\mathbf{H}_1$, and a binary constraint function $\mathbf{B}_1$.

informed by the geometric perspective, but it also causes some difficulty in its proof, when we try to extend to a set of constraint functions of arbitrary arities. We introduce a new technique to overcome this difficulty. First (and this is quite a surprise), it turns out that a dichotomy of two constraint functions of arity 3 is easier to state and prove than the dichotomy of one binary and one ternary constraint functions. Also they can be proven independently of each other. This is a departure from all previous proofs of dichotomy theorems in this area. Second, using the unary constraint functions available in Holant*, any symmetric constraint function $\mathbf{F}$ of arity 4 defines a linear transformation from $\mathbb{R}^3$ to the space of symmetric constraint functions of arity 3, which corresponds to the ternary constraint functions constructible by connecting a unary function to $\mathbf{F}$. In particular, the image of this map, $\mathscr{F}$, is a linear subspace. In particular, the image $\mathscr{F}$ of this map is a linear subspace. Considering the space $\mathscr{F}$ instead of specific sub-functions allows us to bypass the difficulty from the non-explicit form of the dichotomy statement, which is in terms of tensor decompositions up to an orthogonal transformation. We show that a dichotomy of two ternary constraint functions and the fact that $\mathscr{F}$ is closed under linear combinations imply that $\mathscr{F}$ must be of a very special form for Holant$^*_3(\mathscr{F})$ to be tractable, which in turn implies that $\mathbf{F}$ must possess a certain regularity.

While the tractability criterion in Theorem 3.1 is stated in a conceptual and succinct way, the tractable cases are actually quite rich and varied. We present here specific examples of new tractable cases. Denote the domain by $D = \{\mathsf{B}, \mathsf{G}, \mathsf{R}\}$. We use the notation in Figure 1 to denote a symmetric ternary constraint function on domain D. Consider the four constraint functions $\mathbf{F}_1, \mathbf{G}_1, \mathbf{H}_1, \mathbf{B}_1$ in Figure 2. It is not obvious that Holant$^*_3(\mathbf{F}_1, \mathbf{G}_1, \mathbf{H}_1, \mathbf{B}_1)$ is polynomial-time computable.

We apply the orthogonal transform $T = \frac{1}{\sqrt{6}} \begin{bmatrix} \sqrt{2} & \sqrt{2} & \sqrt{2} \\ 1 & 1 & -2 \\ \sqrt{3} & -\sqrt{3} & 0 \end{bmatrix}$ which transforms $\mathbf{F}_1, \mathbf{G}_1$ and $\mathbf{H}_1$ to be supported on $\{\mathsf{B}, \mathsf{G}\}^*$, $\{\mathsf{B}, \mathsf{R}\}^*$, $\{\mathsf{G}, \mathsf{R}\}^*$ respectively, Their tensor decompositions have a revealing structure. Ignoring the scalar constants, we have[1]

$$\begin{aligned} \mathbf{F}'_1 &= T^{\otimes 3}\mathbf{F}_1 = 3\sqrt{3}(1,0,0)^{\otimes 3} + 6\sqrt{6}(0,1,0)^{\otimes 3} = 3\sqrt{3}\,\mathbf{e}_1^{\otimes 3} + 6\sqrt{6}\,\mathbf{e}_2^{\otimes 3} \\ \mathbf{G}'_1 &= T^{\otimes 3}\mathbf{G}_1 = (1,0,\mathrm{i})^{\otimes 3} + (1,0,-\mathrm{i})^{\otimes 3} = (\mathbf{e}_1 + \mathrm{i}\mathbf{e}_3)^{\otimes 3} + (\mathbf{e}_1 - \mathrm{i}\mathbf{e}_3)^{\otimes 3} \\ \mathbf{H}'_1 &= T^{\otimes 3}\mathbf{H}_1 = (0,1,\mathrm{i})^{\otimes 3} + (0,1,-\mathrm{i})^{\otimes 3} = (\mathbf{e}_2 + \mathrm{i}\mathbf{e}_3)^{\otimes 3} + (\mathbf{e}_2 - \mathrm{i}\mathbf{e}_3)^{\otimes 3} \end{aligned}$$

[1] We use i to denote the imaginary unit. Complex numbers do appear, even though the signatures are all real valued. This is similar to eigenvalues.

$$\begin{array}{ccccccc} & & & -3 & & & \\ & & 1 & & -5 & & \\ & -3 & & -5 & & 2 & \\ 1 & & -5 & & 2 & & 10 \end{array}$$

(a) $\mathbf{F}_2$

$$\begin{array}{ccccccc} & & & 5 & & & \\ & & 11 & & 4 & & \\ & 5 & & 4 & & 2 & \\ 11 & & 4 & & 2 & & 1 \end{array}$$

(b) $\mathbf{G}_2$

$$\begin{array}{ccccc} & & 4+2\sqrt{2} & & \\ & -2+2\sqrt{2} & & -4+\sqrt{2} & \\ 4+2\sqrt{2} & & -4+\sqrt{2} & & -2-4\sqrt{2} \end{array}$$

(c) $\mathbf{H}_2$

$$\begin{array}{ccccc} & & 2-2\sqrt{2} & & \\ & 0 & & 2+\sqrt{2} & \\ -2+2\sqrt{2} & & -2-\sqrt{2} & & 0 \end{array}$$

(d) $\mathbf{B}_2$

Figure 3 Ternary constraint functions $\mathbf{F}_2$, $\mathbf{G}_2$ and bianry constraint functions $\mathbf{H}_2, \mathbf{B}_2$.

The vectors in tensor decompositions show that geometrically, $\mathbf{F}'_1$, $\mathbf{G}'_1$, and $\mathbf{H}'_1$ are associated with three coordinate planes. The function $\mathbf{B}'_1 = T^{\otimes 2}\mathbf{B}_1$ written in matrix form is $\begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$, where the (i, j) entry is the function value $\mathbf{B}'_1(i, j)$, for i, j in the new domain set. Applying Theorem 3.1 we can conclude that $\{\mathbf{F}_1, \mathbf{G}_1, \mathbf{H}_1, \mathbf{B}_1\}$ is in tractable class $\mathscr{E}$.

For the second example we apply the orthogonal transform $T = \frac{1}{\sqrt{6}} \begin{bmatrix} \sqrt{2} & \sqrt{2} & -\sqrt{2} \\ 1 & 1 & 2 \\ -\sqrt{3} & \sqrt{3} & 0 \end{bmatrix}$ to the constraint functions in Figure 3.

$$T^{\otimes 3}\mathbf{F}_2 = 3\sqrt{3}((1, \mathrm{i}, 0)^{\otimes 3} + (1, -\mathrm{i}, 0)^{\otimes 3}) + 4\sqrt{2}\mathbf{e}_3^{\otimes 3}, \quad T^{\otimes 3}\mathbf{G}_2 = (\sqrt{3}, \sqrt{6}, 0)^{\otimes 3} + 6\sqrt{2}\mathbf{e}_3^{\otimes 3}$$

and $T^{\otimes 3}\mathbf{H}_2$ and $T^{\otimes 3}\mathbf{B}_2$ can be written in matrix form $\begin{bmatrix} 1 & 1 & 0 \\ 1 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ and $\begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & -1 \\ 1 & -1 & 0 \end{bmatrix}$ respectively, up to scalar constants. Applying Theorem 3.1 we can conclude that $\{\mathbf{F}_2, \mathbf{G}_2, \mathbf{H}_2, \mathbf{B}_2\}$ is in tractable class $\mathscr{D}$.

Our new algorithm also solves some natural problems. Consider the following problem. For $n \in \mathbb{N}$, $i \neq j \in \{\mathsf{B}, \mathsf{G}, \mathsf{R}\}$, and any $a, b \in \mathbb{R}$, let $\mathsf{PARITY}_{a,b}^{n,i,j} : \{\mathsf{B}, \mathsf{G}, \mathsf{R}\}^n \to \mathbb{R}$ be the function

$$\mathsf{PARITY}_{a,b}^{n,i,j}(\mathbf{x}) = \begin{cases} a & \text{if } \mathbf{x} \in \{i, j\}^n \text{ and } \mathbf{x} \text{ contains even number of } i \\ b & \text{if } \mathbf{x} \in \{i, j\}^n \text{ and } \mathbf{x} \text{ contains odd number of } i \\ 0 & \text{otherwise} \end{cases}$$

Let $(\neq)_{pq;r} : \{\mathsf{B}, \mathsf{G}, \mathsf{R}\}^2 \to \{0, 1\}$ for distinct $p, q, r \in \{\mathsf{B}, \mathsf{G}, \mathsf{R}\}$ be the function

$$(\neq)_{pq;r}(x, y) = \begin{cases} 1 & \text{if } x, y \in \{p, q\} \text{ and } x \neq y \\ 1 & \text{if } x = y = r \\ 0 & \text{otherwise} \end{cases}$$

Let $\mathcal{F} = \{\mathsf{PARITY}_{a,b}^{n,i,j} : i \neq j \in \{\mathsf{B}, \mathsf{G}, \mathsf{R}\}, a, b \in \mathbb{R}\} \cup \{(\neq)_{pq;r} : \text{distinct } p, q, r \in \{\mathsf{B}, \mathsf{G}, \mathsf{R}\}\}$. There is a related constraint satisfaction decision problem, where

$$\mathcal{F}^b = \{\mathsf{PARITY}_{a,b}^{n,i,j} : i \neq j \in \{\mathsf{B}, \mathsf{G}, \mathsf{R}\}, a, b \in \{0, 1\}\} \cup \{(\neq)_{pq;r} : \text{distinct } p, q, r \in \{\mathsf{B}, \mathsf{G}, \mathsf{R}\}\},$$

and we ask if an $\mathcal{F}^b$ signature grid has a nonzero assignment. It is not even immediately obvious whether this decision problem is solvable in polynomial time. Theorem 3.1 tells us that $\mathcal{F}$ is in class $\mathscr{E}$ and thus $\text{Holant}_3^*(\mathcal{F})$ is computable in polynomial time, which implies that the decision problem is also solvable in polynomial time.

2 Preliminaries

2.1 Definitions

Let D be a finite domain set, and $\mathcal{F}$ be a set of constraint functions, also called signatures. Each $\mathbf{F} \in \mathcal{F}$ is a mapping from $D^k \to \mathbb{C}$ for some arity k. If the image of $\mathbf{F}$ is contained in $\mathbb{R}$, we say $\mathbf{F}$ is real-valued.

A *signature grid* $\Omega = (G, \mathcal{F}, \pi)$ consists of a graph $G = (V, E)$ where each vertex is labeled by a function $\mathbf{F}_v \in \mathcal{F}$ and π is the labeling. The arity of $\mathbf{F}_v$ must match the degree of v. The Holant problem on instance Ω is to evaluate

$$\text{Holant}_\Omega = \sum_\sigma \prod_{v \in V} \mathbf{F}_v(\sigma|_{E(v)}), \tag{1}$$

where the sum is over all edge assignments $\sigma : E \to D$ and $E(v)$ is the edges adjacent to v, and $\mathbf{F}_v(\sigma|_{E(v)})$ is the evaluation of $\mathbf{F}_v$ on the ordered input tuple $\sigma|_{E(v)}$.

A signature $\mathbf{F}_v$ is listed by its values lexicographically as a table, or it can be expressed as a tensor in $(\mathbb{C}^{|D|})^{\otimes \deg(v)}$. We can identify a unary function $\mathbf{F}(x) : D \to \mathbb{C}$ with a vector $\mathbf{u} \in \mathbb{C}^{|D|}$. Given two vectors $\mathbf{u}$ and $\mathbf{v}$ of dimension $|D|$, the tensor product $\mathbf{u} \otimes \mathbf{v}$ is a vector in $\mathbb{C}^{|D|^2}$, with entries $u_i v_j$ for $1 \le i, j \le |D|$. For matrices $A = (a_{ij})$ and $B = (b_{kl})$ the tensor product (or Kronecker product) $A \otimes B$ is defined similarly; it has entries $a_{ij} b_{kl}$ indexed by $((i,k),(j,l))$ lexicographically. We write $\mathbf{u}^{\otimes k}$ for $\mathbf{u} \otimes \cdots \otimes \mathbf{u}$ with k copies of $\mathbf{u}$. $A^{\otimes k}$ is similarly defined.

A signature $\mathbf{F}$ of arity k is *degenerate* if $\mathbf{F} = \mathbf{u}_1 \otimes \cdots \otimes \mathbf{u}_k$ for some vectors $\mathbf{u}_i$. Such a signature is very weak; there is no interaction between the variables. If every signature in $\mathcal{F}$ is degenerate, then Holant_Ω for any $\Omega = (G, \mathcal{F}, \pi)$ is computable in polynomial time in a trivial way: Simply split every vertex v into $\deg(v)$ vertices each assigned a unary $\mathbf{F}_i$ and connected to the incident edge. Then Holant_Ω becomes a product over each component of a single edge. Thus degenerate signatures are weak and should be properly understood as made up by unary signatures. To concentrate on the essential features that differentiates tractability from intractability, Holant* was introduced in [13, 14]. These are the problems where all unary signatures are assumed to be present, i.e. $\text{Holant}^*(\mathcal{F}) = \text{Holant}(\mathcal{F} \cup \mathcal{U})$ where $\mathcal{U}$ is the set of all unary signatures. We note that for real valued $\mathcal{F}$ the complexity of Holant$^*(\mathcal{F})$ is unchanged whether we use real valued or complex valued $\mathcal{U}$ [25](Lemma 9), and hence in this paper we use real valued $\mathcal{U}$. In the proof of #P-hardness, we freely use complex valued unary functions and apply the known Holant* dichotomy theorems that may use complex valued unary functions.

2.2 Holographic Transformation

To describe the idea of holographic transformations, it is convenient to consider bipartite graphs. For a general graph, we can always transform it into a bipartite graph while preserving the Holant value, as follows: for each edge in the graph, we replace it by a path of length 2, and assign to the new vertex the binary Equality function $(=_2)$.

We use the notation $\text{Holant}(\mathcal{R}|\mathcal{G})$ to denote the Holant problem on bipartite graphs $H = (U, V, E)$, where each signature for a vertex in U or V is from $\mathcal{R}$ or $\mathcal{G}$, respectively. An input instance for the bipartite Holant problem is a bipartite signature grid and is denoted as $\Omega = (H; \mathcal{R}|\mathcal{G}; \pi)$. Signatures in $\mathcal{R}$ are considered as row vectors (or covariant tensors); signatures in $\mathcal{G}$ are considered as column vectors (or contravariant tensors).

For a $|D| \times |D|$ matrix T and a signature set $\mathcal{F}$, define

$$T\mathcal{F} = \{\mathbf{G} : \exists \mathbf{F} \in \mathcal{F} \text{ of arity n, such that } \mathbf{G} = T^{\otimes n}\mathbf{F}\},$$

and similarly for $\mathcal{F}T$. Whenever we write $T^{\otimes n}\mathbf{F}$ or $T\mathcal{F}$, we view the signatures as column vectors; similarly $\mathbf{F}T^{\otimes n}$ or $\mathcal{F}T$ as row vectors. A holographic transformation by T is the following operation: given a signature grid $\Omega = (H; \mathcal{R}|\mathcal{G}; \pi)$, for the same graph H, we get a new grid $\Omega' = (H; \mathcal{R}T|T^{-1}\mathcal{G}; \pi)$ by replacing each signature in $\mathcal{R}$ or $\mathcal{G}$ with the corresponding signature in $\mathcal{R}T$ or $T^{-1}\mathcal{G}$.

▶ **Theorem 2.1** (Valiant's Holant Theorem [28]). *If there is a holographic transformation mapping signature grid* Ω *to* Ω'*, then* $\text{Holant}_\Omega = \text{Holant}_{\Omega'}$.

Therefore, an invertible holographic transformation does not change the complexity of the Holant problem in the bipartite setting. Furthermore, if T is orthogonal, then $(=_2)T^{\otimes 2} = T^{\intercal} I T = I$, so it preserves binary equality. This means that an orthogonal holographic transformation can be used freely in the standard setting.

▶ **Corollary 2.2.** *Suppose* T *is an orthogonal matrix,* $T^{\intercal}T = I$*, and let* $\Omega = (G, \mathcal{F}, \pi)$ *be a signature grid. Under a holographic transformation by* T*, we get a new signature grid* $\Omega' = (G, T\mathcal{F}, \pi)$ *and* $\text{Holant}_\Omega = \text{Holant}_{\Omega'}$.

2.3 Notation

For two nonzero vectors $\mathbf{x}, \mathbf{y} \in \mathbb{C}^n$, we write $\mathbf{x} \sim \mathbf{y}$ to denote projective equality, i.e. $\mathbf{x} = \lambda\mathbf{y}$ for some nonzero $\lambda \in \mathbb{C}$. Throughout this paper, the symbol $\langle \mathbf{u}, \mathbf{v} \rangle$ for $\mathbf{u}, \mathbf{v} \in \mathbb{C}^n$ denotes the dot product, i.e. $\langle \mathbf{u}, \mathbf{v} \rangle = \sum u_i v_i$. We say $\mathbf{u}, \mathbf{v} \in \mathbb{C}^n$ are orthogonal if $\langle \mathbf{u}, \mathbf{v} \rangle = 0$

A signature $\mathbf{F}$ of arity k is *symmetric* if $\mathbf{F}(x_1, \ldots, x_k) = \mathbf{F}(x_{\sigma(1)}, \ldots, x_{\sigma(k)})$ for all $\sigma \in S_k$, the symmetric group. In this paper, if not further specified, a signature $\mathbf{F}$ is assumed to be real-valued, symmetric, and on domain 3. We consider a signature $\mathbf{F}$ and its nonzero multiple $c\mathbf{F}$ as the same signature, since replacing $\mathbf{F}$ by $c\mathbf{F}$ only introduces a easily computable global factor in the Holant value.

A symmetric signature $\mathbf{F}$ on k Boolean variables $\{0, 1\}$ can be expressed as $[f_0, f_1, \ldots, f_k]$ where f_i is the value of $\mathbf{F}$ on inputs of Hamming weight i. In this paper, we focus on signatures on domain size 3, and we use the symbols $\{\mathsf{B}, \mathsf{G}, \mathsf{R}\}$ to denote the domain elements. A binary signature $\mathbf{F}$ (not necessarily symmetric) can be expressed as a $|D| \times |D|$ matrix $M_\mathbf{F}$, where the entry $(i, j) \in D \times D$ is the value of $\mathbf{F}(i, j)$. For the ease of notation, we use the term matrix and binary signature interchangeably, and use $\mathbf{F}$ to refer to both a signature and its matrix $M_\mathbf{F}$. To fix an ordering, binary signature on domain 3 is expressed as $\mathbf{F} = \begin{bmatrix} f_{\mathsf{BB}} & f_{\mathsf{BG}} & f_{\mathsf{BR}} \\ f_{\mathsf{GB}} & f_{\mathsf{GG}} & f_{\mathsf{GR}} \\ f_{\mathsf{RB}} & f_{\mathsf{RG}} & f_{\mathsf{RR}} \end{bmatrix}$. If $\mathbf{F}$ is a symmetric signature, then $\mathbf{F}$ is a symmetric matrix.

Let $\mathbf{G}$ be a binary signature and $\mathbf{F}$ be a symmetric signature of arity $k \geq 2$. We use $\mathbf{G}^{\otimes k}\mathbf{F}$ to denote the gadget constructed by attaching a $\mathbf{G}$ at the edges of $\mathbf{F}$. If $\mathbf{F}$ is written in a tensor form, i.e. $\mathbf{F} = \mathbf{v}_1^{\otimes k} + \cdots + \mathbf{v}_s^{\otimes k}$ for $\mathbf{v}_i \in \mathbb{C}^{|D|}$ we can easily check that $\mathbf{G}^{\otimes k}\mathbf{F} = (\mathbf{G}\mathbf{v}_1)^{\otimes k} + \cdots + (\mathbf{G}\mathbf{v}_s)^{\otimes k}$. Another gadget construction is connecting a unary signature. Let $\mathbf{u} \in \mathbb{C}^{|D|}$ and $\mathbf{F}$ be a symmetric signature on domain D of arity k. Then, $\langle \mathbf{F}, \mathbf{u} \rangle$ is the arity $k-1$ gadget obtained by connecting $\mathbf{u}$ to any edge of $\mathbf{F}$. Since $\mathbf{F}$ is symmetric, the choice of the edge does not matter.

We use Holant_2 to denote the Holant problem on Boolean domain $\{0, 1\}$, and Holant_3 to denote the Holant problem on domain $\{\mathsf{B}, \mathsf{G}, \mathsf{R}\}$. We say two sets of signatures $\mathcal{F}$ and $\mathcal{G}$ are *compatible* if $\text{Holant}(\mathcal{F} \cup \mathcal{G})$ is tractable.

For a domain 3 signature, we use the symbol $\mathbf{F}^{* \to \{i,j\}}$ to denote the Boolean domain signature obtained by restricting the inputs of $\mathbf{F}$ to be from $\{i, j\} \subset \{\mathsf{B}, \mathsf{G}, \mathsf{R}\}$ and identifying i and j with 0 and 1 respectively. For a set of signatures $\mathcal{F}$, $\mathcal{F}^{* \to \{i,j\}} := \{\mathbf{F}^{* \to \{i,j\}} : \mathbf{F} \in \mathcal{F}\}$.

Let $\mathbf{F}$ be a domain 3 signature. We define $\operatorname{supp} \mathbf{F}$ to be the set of inputs for which $\mathbf{F}$ is nonzero. We say $\mathbf{F}$ is an EBD signature (a signature defined essentially on a Boolean domain) if $\operatorname{supp} \mathbf{F} \subseteq \{i, j\}^*$ for some $\{i, j\} \subset \{\mathsf{B}, \mathsf{G}, \mathsf{R}\}$. We say $\mathbf{F}$ is *domain separated* to $\{\mathsf{B}, \mathsf{G}\}$ and $\{\mathsf{R}\}$, written $\mathbf{F}$ is $\mathsf{BG}|\mathsf{R}$, if $\operatorname{supp} \mathbf{F} \subseteq \{\mathsf{B}, \mathsf{G}\}^* \cup \{\mathsf{R}\}^*$. In other words, $\mathbf{F}$ is zero on inputs that take values from both $\{\mathsf{B}, \mathsf{G}\}$ and $\{\mathsf{R}\}$. It is possible that $\operatorname{supp} \mathbf{F} \subseteq \{\mathsf{B}\}^* \cup \{\mathsf{G}\}^* \cup \{\mathsf{R}\}^*$, in which case $\mathbf{F} = a\mathbf{e_1}^{\otimes n} + b\mathbf{e_2}^{\otimes n} + c\mathbf{e_3}^{\otimes n}$ for some $a, b, c \in \mathbb{R}$. We call such $\mathbf{F}$ a GenEQ

signature. We similarly define domain separation to $\{i, j\}$ and $\{k\}$ and write $ij|k$ for any distinct $i, j, k \in \{\mathsf{B}, \mathsf{G}, \mathsf{R}\}$. We refer to a matrix as $\mathsf{BG}|\mathsf{R}$ if it can be viewed as a $\mathsf{BG}|\mathsf{R}$ binary signature. For example, $M = \begin{bmatrix} * & * & 0 \\ * & * & 0 \\ 0 & 0 & * \end{bmatrix}$ is a $\mathsf{BG}|\mathsf{R}$ matrix.

Denote by $\mathcal{E}$ the set of all functions $\mathbf{F}$ such that if $\mathbf{F}$ has arity n, then $\operatorname{supp} \mathbf{F} \subseteq \{a, b, c\}$ for $a = (a_1, \ldots, a_n), b = (b_1, \ldots, b_n), c = (c_1, \ldots, c_n) \in \{\mathsf{B}, \mathsf{G}, \mathsf{R}\}^n$ such that for all $1 \leq i \leq n$, a_i, b_i, c_i are all distinct. We think of $\mathcal{E}$ as a generalized form of GenEQ function to not necessarily symmetric functions.

We use $\mathcal{D}$ to denote the set of 3×3 matrices such that the first two columns are linearly dependent and also the first two rows are linearly dependent. We can easily check that $\mathcal{D}$ is closed under multiplication. If M is a $\mathsf{BG}|\mathsf{R}$ matrix, we can see that $M\mathcal{D}, \mathcal{D}M \subseteq \mathcal{D}$.

We use O_h to denote the symmetry group of an octahedron. As a subgroup of the real 3×3 orthogonal group $O(3)$, O_h consists of the matrices $\begin{bmatrix} \epsilon_1 & 0 & 0 \\ 0 & \epsilon_2 & 0 \\ 0 & 0 & \epsilon_3 \end{bmatrix}$, $\begin{bmatrix} \epsilon_1 & 0 & 0 \\ 0 & 0 & \epsilon_2 \\ 0 & \epsilon_3 & 0 \end{bmatrix}$, $\begin{bmatrix} 0 & \epsilon_1 & 0 \\ \epsilon_2 & 0 & 0 \\ 0 & 0 & \epsilon_3 \end{bmatrix}$, $\begin{bmatrix} 0 & \epsilon_1 & 0 \\ 0 & 0 & \epsilon_2 \\ \epsilon_3 & 0 & 0 \end{bmatrix}$, $\begin{bmatrix} 0 & 0 & \epsilon_1 \\ \epsilon_2 & 0 & 0 \\ 0 & \epsilon_3 & 0 \end{bmatrix}$, $\begin{bmatrix} 0 & 0 & \epsilon_1 \\ 0 & \epsilon_2 & 0 \\ \epsilon_3 & 0 & 0 \end{bmatrix}$ for $\epsilon_1, \epsilon_2, \epsilon_3 \in \{1, -1\}$.

We call a signature/matrix of the form $\begin{bmatrix} 0 & 0 & a \\ 0 & 0 & b \\ c & d & 0 \end{bmatrix}$ for some $a, b, c, d \in \mathbb{R}$ a $\mathsf{Swap}_{\mathsf{BG;R}}$ signature/matrix. The intuition is that a signature of this form swaps the domains $\{\mathsf{B}, \mathsf{G}\}$ and $\{\mathsf{R}\}$. Similarly, we call $\begin{bmatrix} 0 & a & 0 \\ b & 0 & c \\ 0 & d & 0 \end{bmatrix}$ as $\mathsf{Swap}_{\mathsf{BR;G}}$ and $\begin{bmatrix} 0 & a & b \\ c & 0 & 0 \\ d & 0 & 0 \end{bmatrix}$ as $\mathsf{Swap}_{\mathsf{GR;B}}$.

Let $\mathsf{S}^k(\mathbb{C}^n)$ be the set of complex-valued arity-k signature over a domain of size n. The *rank* of $A \in \mathsf{S}^k(\mathbb{C}^n)$ is defined as $\operatorname{rank}(A) := \min\{s : A = \sum_{i=1}^{s} \mathbf{y}_i^{\otimes k}\}$. Several properties of symmetric rank are shown in [17]: rank is well defined and $\operatorname{rank}(\sum_{i=1}^{s} \mathbf{y}_i) = s$ for linearly independent $\mathbf{y}_1, \ldots, \mathbf{y}_s^{\otimes k} \in \mathbb{C}^n$. One can also easily show that the tensor decomposition into linearly independent vectors is unique up to scaling and reordering.

2.4 Known Dichotomy Theorems

An explicit $\operatorname{Holant}_2^*$ dichotomy for a set of symmetric Boolean domain signatures is known [6] (see Section A for the statement). The following dichotomy for a single singature of domain 3 is our starting point.

▶ **Theorem 2.3** (Theorem 2 in [25]). *Let $\mathbf{F}$ be a real-valued symmetric ternary function over domain $\{\mathsf{B}, \mathsf{G}, \mathsf{R}\}$. Then, $\operatorname{Holant}_3^*(\mathbf{F})$ is computable in polynomial time if there exists a real orthogonal matrix T such that one of the following conditions holds. Otherwise, it is #P-hard.*

$\mathfrak{A}$. $T^{\otimes 3}\mathbf{F} = a\mathbf{e_1}^{\otimes 3} + b\mathbf{e_2}^{\otimes 3} + c\mathbf{e_3}^{\otimes 3}$ *for some* $a, b, c \in \mathbb{R}$.

$\mathfrak{B}$. $cT^{\otimes 3}\mathbf{F} = \epsilon(\boldsymbol{\beta_0}^{\otimes 3} + \overline{\boldsymbol{\beta_0}}^{\otimes 3}) + \lambda\mathbf{e_3}^{\otimes 3}$ *where* $\boldsymbol{\beta_0} = (1, \mathfrak{i}, 0)^\intercal$, $\epsilon \in \{0, 1\}$ *and for some* $c, \lambda \in \mathbb{R}$ *and* $c \neq 0$.

3 Statement of the Dichotomy Theorem

Let $\mathcal{F}$ be a set of nondegenerate, real-valued, symmetric signatures over domain $\{\mathsf{B}, \mathsf{G}, \mathsf{R}\}$.

▶ **Theorem 3.1.** $\operatorname{Holant}_3^*(\mathcal{F})$ *is computable in polynomial time if there exists a real orthogonal T, such that one of the following conditions holds. In all other cases,* $\operatorname{Holant}_3^*(\mathcal{F})$ *is #P-hard.*

$\mathscr{A}$. *Every signature in $\mathcal{F}$ has arity ≤ 2.*

$\mathscr{B}$. $T\mathcal{F} \subseteq \mathcal{E}$.

$\mathscr{C}$. *For all $\mathbf{F} \in T\mathcal{F}$ of arity ≥ 3, $\operatorname{supp} \mathbf{F} \subseteq \{\mathsf{B}, \mathsf{G}\}^*$, and*
For all binary $\mathbf{G} \in T\mathcal{F}$, either $\mathbf{G} \in \mathcal{D}$ or $\mathbf{G}$ is $\mathsf{BG}|\mathsf{R}$, and
$\operatorname{Holant}_2^*((T\mathcal{F})^{*\to\{\mathsf{B},\mathsf{G}\}})$ *is tractable.*

$\mathscr{D}$. *For all* $\mathbf{F} \in T\mathcal{F}$ *of arity* ≥ 3, $\mathbf{F}$ *is* $\mathsf{BG|R}$, *and*
For all binary $\mathbf{G} \in T\mathcal{F}$, *either* $\mathbf{G}$ *is* $\mathsf{BG|R}$ *or* $\mathbf{G}$ *is* $\mathsf{Swap}_{\mathsf{BG;R}}$, *and*
$\mathrm{Holant}_2^*((T\mathcal{F})^{*\to\{\mathsf{B},\mathsf{G}\}})$ *is tractable.*

$\mathscr{E}$. *Let* $\mathcal{F}_{ij} = \{\mathbf{F} \in T\mathcal{F} : \operatorname{supp} \mathbf{F} \subseteq \{i,j\}^*\}$. *Let* $\mathcal{R} = T\mathcal{F} - (\mathcal{F}_{\mathsf{BG}} \cup \mathcal{F}_{\mathsf{BR}} \cup \mathcal{F}_{\mathsf{GR}})$.
$\mathcal{R} \subseteq \mathbb{R}O_h$, *and* $\langle \mathcal{R}' \rangle \subseteq O_h$, *where* $\langle \mathcal{R}' \rangle$ *is the monoid generated by* $\mathcal{R}' = \mathbb{R}\mathcal{R} \cap O_h$, *and*
For all $i, j \in \{\mathsf{B},\mathsf{G},\mathsf{R}\}$, $\mathrm{Holant}_2^*(\langle \mathcal{R}' \rangle^{*\to\{i,j\}} \cup \mathcal{F}_{ij})$ *is tractable, and*
For all $i, j \in \{\mathsf{B},\mathsf{G},\mathsf{R}\}$, $\mathrm{Holant}_2^*((\bigcup_{\mathbf{G}\in\langle\mathcal{R}'\rangle} \mathbf{G}(T\mathcal{F}))^{*\to\{i,j\}})$ *is tractable.*

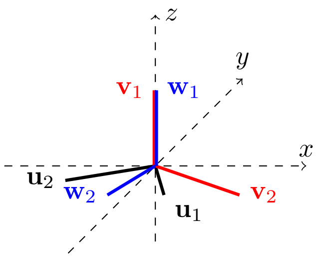
In classes $\mathscr{C}$, $\mathscr{D}$, and $\mathscr{E}$, we refer to the Holant_2^* tractability. By the Boolean domain dichotomy, it is necessary that those Holant_2^* problems are tractable.

In case (b) of class $\mathscr{E}$, it seems like we may need to use the asymmetric Holant_2^* dichotomy (which is known [16]), because while the signatures in $\mathcal{R}$ are symmetric, $\langle \mathcal{R} \rangle$ may contain asymmetric signatures. We claim that is not necessary. Let $\mathbf{G} \in O_h$. If $\mathbf{G}^{*\to\{i,j\}}$ is nondegenerate and asymmetric, then $\mathbf{G}^{*\to\{i,j\}} = \pm\left[\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right]$. We can deduce that $\mathbf{G}^{*\to\{i,j\}}$ is *universally compatible*, i.e., when appended to any Holant_2^* tractable class results in a tractable set, without referring to any asymmetric Holant_2^* dichotomy. For type $\mathrm{I}(a,b)$ (see Definition A.1 and Theorem A.2), we see that $\left[\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right] = \frac{1}{4a^2+b^2}\left[\begin{smallmatrix} b & -2a \\ -2a & -b \end{smallmatrix}\right]\left[\begin{smallmatrix} 2a & b \\ b & -2a \end{smallmatrix}\right]$. The first matrix in the symmetric signature notation is $[b, -2a, -b]$, which satisfies the recurrence $a \cdot b + b(-2a) = a \cdot (-b)$. The second matrix in the symmetric signature notation is $[2a, b, -2a]$, which is the other specified form. For type II, we see that $\left[\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right] = \left[\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right]\left[\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right]$, and $[1, 0, -1]$ and $[0, 1, 0]$ are type II.

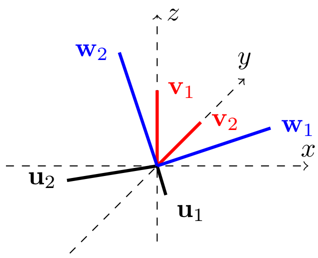
This is not a coincidence. We may view the Holant_2^* dichotomy in a geometric way. Consider the tractable type $\mathrm{I}(a,b)$. There are two norm 1 orthogonal vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^2$ such that any signature of arity ≥ 3 of type $\mathrm{I}(a,b)$ is of the form $c\mathbf{u}^{\otimes n} + d\mathbf{v}^{\otimes n}$. We may represent type $\mathrm{I}(a,b)$ signatures of arity ≥ 3 by two orthogonal lines in the $\mathbb{R}^2$ plane. Consider the gadget construction of connecting a binary signature $\mathbf{G}$ to all of the edges of a signature. In the tensor decomposition form, we have $\mathbf{G}^{\otimes n}(c\mathbf{u}^{\otimes n} + d\mathbf{v}^{\otimes n}) = c(\mathbf{G}\mathbf{u})^{\otimes n} + d(\mathbf{G}\mathbf{v})^{\otimes n}$. For the binary signatures, we can check that the tractable signatures correspond to the linear transformations that fix the union of the two orthogonal lines as a set. This is easily verified for the case of type $\mathrm{I}(0,1)$, when $\mathbf{u} = \mathbf{e}_1$ and $\mathbf{v} = \mathbf{e}_2$, since $[*, 0, *]$ signature corresponds to scaling $\mathbf{e}_1$ and $\mathbf{e}_2$, while $[0, *, 0]$ signature corresponds to reflection along $x = y$ line, and similarly for other type $I(a,b)$. The asymmetric signature above is a $\pi/2$ rotation, which fixes any pair of orthogonal lines, so it is tractable with all $\mathrm{I}(a,b)$ types.

The tractable type II may be viewed as a circle in the $\mathbb{R}^2$ plane. The justification is that any type II signature can be written as $(\mathbf{u} + i\mathbf{v})^{\otimes n} + (\mathbf{u} - i\mathbf{v})^{\otimes n}$ for two orthogonal $\mathbf{u}, \mathbf{v} \in \mathbb{R}^2$ of the same norm, and all such signatures can be constructed from one type II signature. The set of linear transformations that fix the circle is the orthogonal group $O(2)$. We see that up to a scalar factor, the signature $[x, y, -x]$ corresponds to a reflection matrix and the signature $[x, 0, x]$ to the identity. Since any rotation can be written as a product of two reflections, they are also compatible with type II signatures.

A similar analogy can be made for domain 3 signatures as well, since the tensor decomposition forms in Theorem 2.3 are also about orthogonality of the vectors. Similar to the Boolean domain tractable signatures, a tractable domain 3 signature also can be represented as a set of vectors in the three dimensional space. For instance, let $\mathbf{u}_1^{\otimes 3} + \mathbf{u}_2^{\otimes 3}$, $\mathbf{v}_1^{\otimes 3} + \mathbf{v}_2^{\otimes 3}$ and $\mathbf{w}_1^{\otimes 3} + \mathbf{w}_2^{\otimes 3}$, be three signatures. The idea is depicted in Figure 4, and this intuition is formalized in Lemma 5.3. Class $\mathscr{C}$ says that the signatures of arity 3 or higher must live in some plane, and we can assume it is the xy-plane after an orthogonal transformation. The compatible binary signatures are those fixing the xy-plane ($\mathsf{BG|R}$) or a degenerate transformation on the xy-plane. The class $\mathscr{D}$ says that the signatures of arity 3 or higher are



(a) $\mathbf{u}_1, \mathbf{u}_2, \mathbf{v}_2, \mathbf{w}_2$ are on the xy plane and $\mathbf{v}_1, \mathbf{w}_1$ are on the z axis.



(b) $\mathbf{u}_1, \mathbf{u}_2$ are on xy plane, $\mathbf{v}_1, \mathbf{v}_2$ are on the yz plane and $\mathbf{w}_1, \mathbf{w}_2$ are on the xz plane.

Figure 4 Geometric idea behind the dichotomy of $\text{Holant}_3^*(\mathcal{F})$.

formed by the xy-plane and the z-axis. The compatible binary signatures, after the same orthogonal transformation, are those fixing the xy-plane and z-axis line ($\mathsf{BG|R}$) or mapping between the xy-plane and the z-axis ($\mathsf{Swap}_{\mathsf{BG;R}}$). The class $\mathscr{E}$ says that the signatures of arity 3 or higher must live in one of xy-plane, yz-plane, or xz-plane. The compatible binary signatures are those permuting the three planes without stretching. Such transformations are precisely the group O_h.

4 Tractability

In this section, we prove tractability by giving a polynomial time algorithm for each of the classes in Theorem 3.1. By Corollary 2.2, $\text{Holant}_3^*(\mathcal{F})$ is tractable if and only if $\text{Holant}_3^*(T\mathcal{F})$ is tractable. Also, we may always assume that the given signature grid is connected, as the Holant value of any signature grid is the product over connected components.

Classes $\mathscr{A}$ and $\mathscr{B}$ are tractable by standard arguments. If every signature in $\mathcal{F}$ has arity ≤ 2 (class $\mathscr{A}$), then, the graph of the signature grid Ω is a disjoint union of paths and cycles. By matrix multiplication, we can compute the Holant value for a path. The Holant value for a cycle is obtained by taking the trace of a path.

Suppose $T\mathcal{F} \subseteq \mathcal{E}$ for an orthogonal T (class $\mathscr{B}$). Let Ω be a $T\mathcal{F}$ signature grid and e any edge. For any $\mathsf{B}, \mathsf{G}, \mathsf{R}$ assignment to e, the assignment must propagate uniquely to all edges, which implies that there are at most three assignments to the whole grid that can result in a nonzero Holant value.

4.1 Class $\mathscr{C}$

Suppose $\mathcal{F}$ satisfies the conditions of $\mathscr{C}$ by an orthogonal T. Let Ω be a connected signature grid over $T\mathcal{F}$. We may assume Ω is not in Class $\mathscr{A}$. If a unary signature is connected to another unary signature then this is the entire Ω and we are done. If a unary signature is connected to a binary signature then they become another unary constraint which we can compute its signature. So, by induction, we can assume any unary remaining is connected to some signature $\mathbf{F} \in T\mathcal{F}$ of arity ≥ 3, which has $\operatorname{supp} \mathbf{F} \subseteq \{\mathsf{B}, \mathsf{G}\}^*$, and thus we can replace the unary restricted to $\{\mathsf{B}, \mathsf{G}\}^*$. If there is a chain of binary signatures $\mathbf{G}_1, \mathbf{G}_2, \ldots, \mathbf{G}_k$, we replace it by a single binary signature $\mathbf{G}$ by taking the matrix product. If all $\mathbf{G}_i$ are $\mathsf{BG|R}$, then $\mathbf{G}$ is also $\mathsf{BG|R}$. In addition, the matrix of $\mathbf{G}^{*\to\{\mathsf{B},\mathsf{G}\}}$ is equal to the product of the matrices of $\mathbf{G}_i^{*\to\{\mathsf{B},\mathsf{G}\}}$. If there is some i such that $\mathbf{G}_i \in \mathcal{D}$, then $\mathbf{G} \in \mathcal{D}$ as well, since $\mathcal{D}$ is closed under multiplication and also closed under left or right multiplication by a $\mathsf{BG|R}$ matrix. Note that if $\mathbf{G} \in \mathcal{D}$, then $\mathbf{G}^{*\to\{\mathsf{B},\mathsf{G}\}}$ is degenerate. Now since any binary

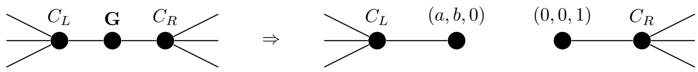


Figure 5 Factorization of $\mathsf{Swap}_{\mathsf{BG;R}}$ signature $\mathbf{G} = \left[\begin{smallmatrix} 0 & 0 & a \\ 0 & 0 & b \\ c & d & 0 \end{smallmatrix}\right]$ in type (1) assignments.

$\mathbf{G}$ remaining can only be connected to signatures $\mathbf{F}$ of arity ≥ 3 with $\operatorname{supp} \mathbf{F} \subseteq \{\mathsf{B}, \mathsf{G}\}^*$, we can replace $\mathbf{G}$ by $\mathbf{G}^{*\to\{\mathsf{B},\mathsf{G}\}}$. Thus we obtain an equivalent signature grid Ω' as a Holant_2^* instance on domain $\{\mathsf{B}, \mathsf{G}\}$. Then condition 3. implies that the Holant value of Ω' can be computed in polynomial time.

4.2 Class $\mathscr{D}$

Suppose $\mathcal{F}$ satisfies the conditions of $\mathscr{D}$ by an orthogonal T, and let Ω be a connected $T\mathcal{F}$ signature grid. If Ω does not use any binary signature of $\mathsf{Swap}_{\mathsf{BG;R}}$, then all signatures are $\mathsf{BG|R}$. Then, if any edge gets assigned B or G, all other edges must also be assigned B or G for the assignment to result in a nonzero value. Similarly, any assignment of R to an edge must propagate as R to all other edges for the assignment to result in a nonzero value. Therefore, the Holant value is sum of all assignments taking values in $\{\mathsf{B}, \mathsf{G}\}$ and an assignment that only assigns R. The first sum can be computed in polynomial since $\mathrm{Holant}_2^*((T\mathcal{F})^{*\to\{\mathsf{B},\mathsf{G}\}})$ is tractable, and the second value is computable in polynomial time.

Now, suppose Ω contains $\mathsf{Swap}_{\mathsf{BG;R}}$ signatures. First, note that a product of two $\mathsf{Swap}_{\mathsf{BG;R}}$ signatures is $\mathsf{BG|R}$ signature that is degenerate on $\{\mathsf{B}, \mathsf{G}\}$: $\left[\begin{smallmatrix} 0 & 0 & a \\ 0 & 0 & b \\ c & d & 0 \end{smallmatrix}\right]\left[\begin{smallmatrix} 0 & 0 & e \\ 0 & 0 & f \\ g & h & 0 \end{smallmatrix}\right] = \left[\begin{smallmatrix} ag & ah & 0 \\ bg & bh & 0 \\ 0 & 0 & ce+df \end{smallmatrix}\right]$. Hence, we may reduce any chain $\mathbf{G}_1, \mathbf{G}_2, \cdots, \mathbf{G}_k$ of $\mathsf{Swap}_{\mathsf{BG;R}}$ signatures of length $k \geq 2$ by matrix multiplication to a single $\mathbf{G}$ if k is even and a length 2 chain $\mathbf{G}, \mathbf{G}_k$ if k is odd, where $\mathbf{G}$ is a $\mathsf{BG|R}$ signature that is degenerate on $\{\mathsf{B}, \mathsf{G}\}$. In particular, $\mathbf{G}^{*\to\{\mathsf{B},\mathsf{G}\}}$ is compatible with $\mathcal{F}^{*\to\{\mathsf{B},\mathsf{G}\}}$. Therefore, we may assume that we have a signature grid Ω' such that any $\mathsf{Swap}_{\mathsf{BG;R}}$ signature is connected to two $\mathsf{BG|R}$ signatures. Now, suppose we gather each connected component of $\mathsf{BG|R}$ signatures as a cluster, so that the connections between clusters are by $\mathsf{Swap}_{\mathsf{BG;R}}$ signatures. If we imagine a graph with vertices being the clusters and edges being the $\mathsf{Swap}_{\mathsf{BG;R}}$ signatures, then this graph must be bipartite. Otherwise, suppose there is an odd cycle of clusters $C_1, C_2, \cdots, C_k, C_1$. In any nonzero assignment, each cluster can take only values from $\{\mathsf{B}, \mathsf{G}\}$ or R. If C_1 gets $\{\mathsf{B}, \mathsf{G}\}$, then C_2 must have R because of the $\mathsf{Swap}_{\mathsf{BG;R}}$ connection. Therefore, if there is an odd cycle, C_1 will have an incoming R, resulting in a zero evaluation. Similar argument shows that an assignment of R to C_1 evaluates to zero.

Let the bipartition be $L \sqcup R$. There are only two types of nonzero assignments: (1) all clusters in L get $\{\mathsf{B}, \mathsf{G}\}$ and all clusters in R get R; (2) all clusters in L get R and all clusters in R get $\{\mathsf{B}, \mathsf{G}\}$. For both types of assignments, the $\mathsf{Swap}_{\mathsf{BG;R}}$ signatures connecting a cluster $C_L \in L$ and $C_R \in R$ factors into two degenerate signatures. In particular, suppose $\mathbf{G} = \left[\begin{smallmatrix} 0 & 0 & a \\ 0 & 0 & b \\ c & d & 0 \end{smallmatrix}\right]$ is a $\mathsf{Swap}_{\mathsf{BG;R}}$ signature connecting C_L and C_R. Let Ω' be obtained from Ω by removing $\mathbf{G}$ and connecting the unary $(a, b, 0)$ to C_L and $(0, 0, 1)$ to C_R, as described in Figure 5. Then, since C_L only takes values in $\{\mathsf{B}, \mathsf{G}\}$ and C_R only takes values in R, the sum of type (1) assignments is the same as the Holant value of Ω'. Therefore, to evaluate the sum of type (1) assignments, we may factor all the $\mathsf{Swap}_{\mathsf{BG;R}}$ signatures in this way, evaluate in $\{\mathsf{B}, \mathsf{G}\}$ domain Holant_2^* on L and assign R to all of R, and multiply the two resulting values. Both can be done in polynomial time since $\mathcal{F}^{*\to\{\mathsf{B},\mathsf{G}\}}$ is assumed to be tractable. Similarly, we may compute the contribution of type (2) assignments, and the Holant value is just the sum of those two values.

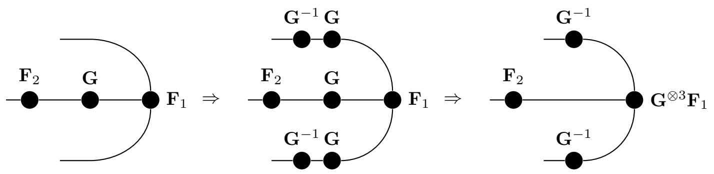


Figure 6 Local holographic transformation.

4.3 Class $\mathscr{E}$

Suppose $\mathcal{F}$ satisfies the conditions of $\mathscr{E}$ by an orthogonal T. Let Ω be a $T\mathcal{F}$ signature grid. If Ω does not use any signature from $\mathcal{R}$, then all signatures in Ω are EBD. Then, whenever there is an edge between two signatures with different support, the edge factors as pinning. For example, if there is an edge e between a signature $\mathbf{F}_{\mathsf{BG}} \in \mathcal{F}_{\mathsf{BG}}$ and $\mathbf{F}_{\mathsf{GR}} \in \mathcal{F}_{\mathsf{GR}}$, we may remove the edge and connect the unary $(0, 1, 0)$ to both $\mathbf{F}_{\mathsf{BG}}$ and $\mathbf{F}_{\mathsf{GR}}$. This is because e can only take value G in any nonzero assignment. Also, if there is a connection between any unary signature $\mathbf{u}$ and a EBD signature supported on $\{i, j\}$, then we may replace $\mathbf{u}$ with $\mathbf{u}^{*\to\{ij\}}$. We obtain a new signature grid Ω' after factoring all such edges, and each of the connected components of Ω' is an instance of $\operatorname{Holant}_2^*(\mathcal{F}_{ij})$ for some $i, j \in \{\mathsf{B}, \mathsf{G}, \mathsf{R}\}$. By assumption, the Holant value of each of the connected components can be computed in polynomial time.

Now, suppose Ω contains signatures from $\mathcal{R}$. We may replace those signatures by the corresponding scalar multiple in $\mathcal{R}' \subseteq O_h$. By the above, we may assume that in Ω, there is no edge between EBD signatures of different supports. First, we reduce any chain of $\mathcal{R}'$ signatures into a single binary signature in $\langle \mathcal{R}' \rangle \subseteq O_h$. Reducing any connection of $\mathcal{R}'$ signature with a unary signature to a unary, we obtain a new grid Ω' in which any $\langle \mathcal{R}' \rangle$ signature is between two signatures from $\mathcal{F}_{ij}$ and $\mathcal{F}_{i'j'}$. We imagine Ω' is composed of clusters C where each cluster is a connected component of $\mathcal{F}_{ij}$ signatures for some $i, j \in \{\mathsf{B}, \mathsf{G}, \mathsf{R}\}$, and each cluster has outgoing edges of $\langle \mathcal{R}' \rangle$ signatures.

Suppose $\mathbf{G} \in \langle \mathcal{R}' \rangle$ is a self loop on a cluster C. As its two end points are from C and all signatures in C are EBD on the same $\{i, j\}$, we may replace $\mathbf{G}$ with $\mathbf{G}^{*\to\{i,j\}}$ which is compatible with signatures in C by assumption, so we may absorb it into C.

Suppose $\mathbf{G} \in \langle \mathcal{R}' \rangle$ connects two clusters C_1 and C_2 with supports $\{i_1, j_1\}$ and $\{i_2, j_2\}$ respectively. Suppose $\mathbf{F}_1$ and $\mathbf{F}_2$ are the signatures in C_1 and C_2 connected by $\mathbf{G}$. Let the arity of $\mathbf{F}_1$ be n. We replace all the edges of $\mathbf{F}_1$, except the one connecting to $\mathbf{G}$, with $\mathbf{G}^{-1}\mathbf{G}$. Clearly, the Holant value is unchanged. The process is described in Figure 6 for arity 3 case. We now have $\mathbf{G}^{\otimes n}\mathbf{F}_1$ connected directly to $\mathbf{F}_2$. If $|\operatorname{supp} \mathbf{G}^{\otimes n}\mathbf{F}_1 \cap \operatorname{supp} \mathbf{F}_2| \leq 1$, the edge factors in to pinning. Otherwise, by the assumption, $\mathbf{G}^{\otimes n}\mathbf{F}_1$ is compatible with $\mathbf{F}_2$, so we may absorb it into the C_2 cluster.

We choose a cluster to begin with, and repeatedly absorb its neighboring signatures or factor the edge into pinning by the above process. Each step of the above process can be done in polynomial time, and the number of steps required is at most the number of edges in Ω'. In the end, we will be left with multiple connected components in which all the signatures are EBD supported on the same domain, and their $\operatorname{Holant}_2^*$ values can be computed in polynomial time by the assumption.

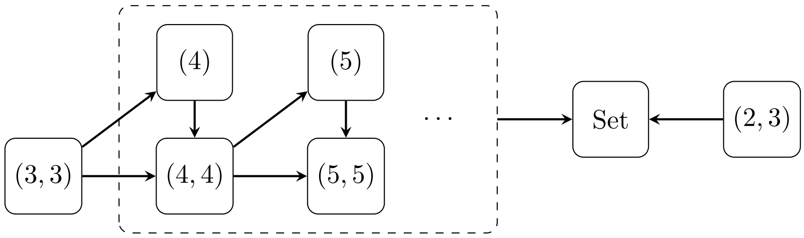


Figure 7 Logical dependency diagram: (n) refers to a dichotomy of $\text{Holant}_3^*(\mathbf{F})$ for an arity-n signature $\mathbf{F}$. (n, m) refers to a dichotomy of $\text{Holant}_3^*(\mathbf{F}, \mathbf{G})$ of an arity-n signature $\mathbf{F}$ and an arity-m signature $\mathbf{G}$. 'Set' refers to the dichotomy of an arbitrary set of signatures.

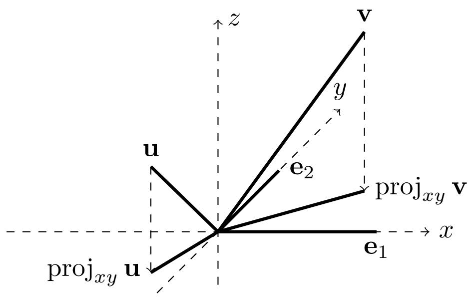


Figure 8 Geometric intuition of domain restriction.

5 Outline of Hardness

After a dichotomy of a single ternary signature [15], a natural next step is proving a dichotomy of a pair of ternary and binary signatures (as binary signatures are the "simplest" signatures after unary signatures), and use it to prove further theorems. However, for domain size 3 in the Holant setting, binary signatures actually allow nontrivial, and somewhat unanticipated, interactions with other signatures. Also, it turns out that a dichotomy for a pair of binary and ternary signatures, while certainly needed on its own, is not easily applicable for showing further dichotomies. We circumvent this difficulty by proving a dichotomy of a pair of ternary signatures directly.

We describe the logical dependency of our proof in Figure 7. We start with the known dichotomy, Theorem 2.3, of a single ternary signature. We use it to show the dichotomy of a pair of ternary and binary signatures and a pair of ternary signatures. These two proofs can be found in the full version [12] and we note that they are independent of each other. To arrive at a dichotomy of a set of signatures, we also need to show a dichotomy for any single signature of arbitrary arity $n \geq 4$, which we show in Section 6. The remaining proof leading to a dichotomy of a set of signatures can be found in the full version [12].

The main intuition behind the proof of hardness is the geometry of the vectors in a tensor decomposition of a signature. We always start with a canonical form of a signature (after a suitable orthogonal transformation), for example, $\mathbf{F} = \mathbf{e}_1^{\otimes 3} + \mathbf{e}_2^{\otimes 3}$. $\mathbf{F}$ can realize $(=_{\mathsf{BG}})$ and take domain restriction to $\{\mathsf{B}, \mathsf{G}\}$. If the other signature is $\mathbf{G} = \mathbf{u}^{\otimes 3} + \mathbf{v}^{\otimes 3}$ for orthogonal $\mathbf{u}, \mathbf{v} \in \mathbb{R}^3$, the domain restriction is $\mathbf{G}^{* \to \{\mathsf{B},\mathsf{G}\}} = (\text{proj}_{xy}\, \mathbf{u})^{\otimes 3} + (\text{proj}_{xy}\, \mathbf{v})^{\otimes 3}$. Essentially, most of the arguments boil down to showing that the angle between $\text{proj}_{xy}\, \mathbf{u}$ and $\text{proj}_{xy}\, \mathbf{v}$ must be $\pi/2$ (with one exception), otherwise, $\mathbf{G}^{* \to \{\mathsf{B},\mathsf{G}\}}$ is #P-hard by the Boolean domain dichotomy. The other tractable possibility is when $\text{proj}_{xy}\, \mathbf{u} \sim \text{proj}_{xy}\, \mathbf{v}$, which can only happen if the plane that contains $\mathbf{u}$ and $\mathbf{v}$ contains the z-axis, which corresponds to the case when $\mathbf{G}^{* \to \{\mathsf{B},\mathsf{G}\}}$ is degenerate.

Using this idea, we can prove the dichotomy of $\text{Holant}_3^*(\mathbf{F}, \mathbf{G})$ in which $\mathbf{F}, \mathbf{G}$ are ternary signatures. The individual statements can be found in the full version [12]. We can see that the statements are more complex when signatures are of rank 2. We define a notion of a *plane* of a rank 2 signature, which formalizes the idea behind Figure 4 and allows us to express the dichotomy succinctly. A rank 2 signature of type $\mathfrak{A}$ or type $\mathfrak{B}$ of Theorem 2.3 has a symmetric tensor decomposition such that the two vectors occurring inside it are orthogonal, i.e. $\mathbf{v}_1^{\otimes 3} + \mathbf{v}_2^{\otimes 3}$ or $(\mathbf{v}_1 + \mathbf{i}\mathbf{v}_2)^{\otimes 3} + (\mathbf{v}_1 - \mathbf{i}\mathbf{v}_2)^{\otimes 3}$. The two vectors are unique up to a scalar, so the plane of a signature $\mathbf{F}$, denoted by $P_\mathbf{F} = \text{span}(\mathbf{v}_1, \mathbf{v}_2)$, is well defined.

▶ **Definition 5.1.** *Let $\mathbf{F}$ be a rank 2 signature over domain $\{\mathsf{B}, \mathsf{G}, \mathsf{R}\}$ of type $\mathfrak{A}$ or type $\mathfrak{B}$. If $\mathbf{F}$ is type $\mathfrak{A}$, we may write $\mathbf{F} = \mathbf{v}_1^{\otimes 3} + \mathbf{v}_2^{\otimes 3}$. If $\mathbf{F}$ is type $\mathfrak{B}$, we may write $\mathbf{F} = (\mathbf{v}_1 + \mathbf{i}\mathbf{v}_2)^{\otimes 3} + (\mathbf{v}_1 - \mathbf{i}\mathbf{v}_2)^{\otimes 3}$. We define the plane of $\mathbf{F}$ as $P_\mathbf{F} = \text{span}(\mathbf{v}_1, \mathbf{v}_2)$.*

Note that $\mathbf{F}$ is EBD if and only if $P_\mathbf{F}$ is one of the coordinate planes, i.e. xy-plane, yz-plane or xz-plane. We will say that two planes are orthogonal if their normal vectors are orthogonal. This notion is well defined because a normal vector to a plane in 3 dimensional space is unique up to a scalar. The following proposition can be shown with elementary linear algebra.

▶ **Proposition 5.2.** *Let $\mathcal{F}$ be a set of rank 2 signatures of type $\mathfrak{A}$ or type $\mathfrak{B}$. The following two statements are equivalent:*

1. *For any orthogonal T, there is some signature $\mathbf{F} \in T\mathcal{F}$ that is not EBD.*
2. *There exist $\mathbf{F}, \mathbf{G} \in \mathcal{F}$ such that $P_\mathbf{F}$ and $P_\mathbf{G}$ are not equal or orthogonal.*

The next lemma encapsulates the essence of the dichotomy theorems of two ternary signatures of rank 2, and extends it to an arbitrary set of rank 2 ternary signatures.

▶ **Lemma 5.3.** *Let $\mathcal{F}$ be a set of rank 2 signatures of type $\mathfrak{A}$ or type $\mathfrak{B}$. Suppose that for every orthogonal matrix T, neither of the following holds:*

1. *for all $\mathbf{F} \in T\mathcal{F}$, $\mathbf{F}$ is $\mathsf{GR|B}$.*
2. *all signatures in $T\mathcal{F}$ are EBD.*

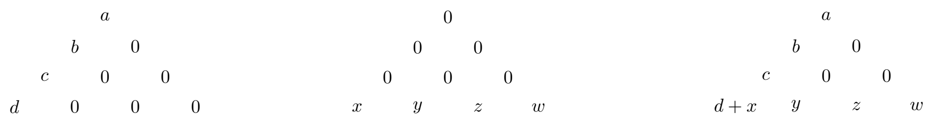
Then, $\text{Holant}_3^*(\mathcal{F})$ is #P-hard.

The idea of viewing binary signatures as transformation on the signatures is naturally captured by the following construction. Let $\mathcal{T}$ be a set of ternary signatures and $\mathcal{B}$ be a set of binary signatures. Let $\mathcal{F} = \mathcal{T} \cup \mathcal{B}$. Let $\langle \mathcal{B} \rangle$ be the monoid generated by $\mathcal{B}$ under multiplication. We define $\mathcal{O}$ to be

$$\mathcal{O} := \{\mathbf{G}^{\otimes 3}\mathbf{F} : \mathbf{F} \in \mathcal{T}, \mathbf{G} \in \langle \mathcal{B} \rangle, \mathbf{G}^{\otimes 3}\mathbf{F} \text{ is non-degenerate}\}\,. \tag{2}$$

Combinatorially, $\mathcal{O}$ is the set of all gadgets constructible from connecting the same chain of binary signatures to the three edges of a ternary signature, $(\mathbf{G}_1\mathbf{G}_2 \cdots \mathbf{G}_k)^{\otimes 3}\mathbf{F}$ for some $\mathbf{G}_i \in \mathcal{B}$ and $\mathbf{F} \in \mathcal{T}$. It can also be viewed as the orbit (ignoring degenerate signatures) of $\mathcal{T}$ under the monoid action of $\langle \mathcal{B} \rangle$, where the action is defined by $\mathbf{G} : \mathbf{F} \mapsto \mathbf{G}^{\otimes 3}\mathbf{F}$ for $\mathbf{G} \in \langle \mathcal{B} \rangle$ and $\mathbf{F} \in \mathcal{T}$. Note that $\mathcal{G}$ is a set of symmetric ternary signatures, and if $\mathbf{G} \in \mathcal{B}$ and $\mathbf{F} \in \mathcal{O}$, then $\mathbf{G}^{\otimes 3}\mathbf{F} \in \mathcal{O}$ as well.

To complete the proof of the dichotomy of $\text{Holant}_3^*(\mathcal{F})$, we use Lemma 5.3 on $\text{Holant}_3^*(\mathcal{O})$ to analyze $\mathcal{B}$. The details can be found in the full version [12].



(a) $\mathbf{F}$ **(b)** $\mathbf{G}$ **(c)** $\mathbf{H}$

Figure 9 $\mathbf{F}$, $\mathbf{G}$, and $\mathbf{H} = \mathbf{F} + \mathbf{G}$.

6 A Single Signature Dichotomy

In this section, we prove a dichotomy of $\text{Holant}_3^*(\mathbf{F})$ for an $\mathbf{F}$ of arbitrary arity n. The case $n \le 3$ was proved in [15]. Let $n \ge 4$. Similar to the Boolean domain case, it is natural to expect that higher arity tractable signatures also have the same tensor decomposition form, i.e. $a\mathbf{e}_1^{\otimes n} + b\mathbf{e}_2^{\otimes n} + c\mathbf{e}_3^{\otimes n}$ or $\boldsymbol{\beta}^{\otimes n} + \overline{\boldsymbol{\beta}}^{\otimes n} + \lambda\mathbf{e}_3^{\otimes n}$ for $\boldsymbol{\beta} = (1, \mathfrak{i}, 0)$ after some orthogonal transformation. We show that indeed this is the case.

The proof is an induction on the arity. Assume $\text{Holant}_3^*(\mathbf{F})$ is not #P-hard. First, we show that $\mathbf{F}$ of arity 4 must be of a suitable form, using the dichotomy of two ternary signatures. Then we show that arity 4 signatures must have a tensor decomposition of at most 3 linearly independent vectors. Then we argue that the dichotomy of two arity 4 signatures must take essentially the same form as the dichotomy of two ternary signatures. Finally, we generalize the proof to higher arities. The logical dependency is visualized in Figure 7.

6.1 Subspace of Signatures

Let $\mathbf{F}$ be a real-valued symmetric signature of arity 4. Consider the set $\mathscr{F} = \{\langle \mathbf{F}, \mathbf{u}\rangle : \mathbf{u} \in \mathbb{R}^3\}$. Note that $\mathbf{u} \mapsto \langle \mathbf{F}, \mathbf{u}\rangle$ is a linear map from $\mathbb{R}^3$ to the space of ternary real-valued symmetric signatures. We show a dichotomy for an arbitrary subspace $\mathscr{F}$ of ternary real-valued signatures. For a nonempty $\mathscr{F}$, each ternary signature must be a tractable signature, otherwise the problem is already #P-hard. We prove this dichotomy for a subspace $\mathscr{F}$ by considering each canonical form such a tractable signature can take under an orthogonal T. Note that $T\mathscr{F}$ is also a subspace. The exact statements can be found in Section B. We summarize the dichotomy statements: there exists an orthogonal T such that $T\mathscr{F}$ is a subset of $\{a\mathbf{e}_1^{\otimes 3} + b\mathbf{e}_2^{\otimes 3} + c\mathbf{e}_3^{\otimes 3} : a, b, c \in \mathbb{R}\}$ or $\{\mathbf{G} + \lambda\mathbf{e}_3^{\otimes 3} : \lambda \in \mathbb{R}, \operatorname{supp} \mathbf{G} \subseteq \{\mathsf{B}, \mathsf{G}\}^*, \mathbf{G}^{*\to\{\mathsf{B},\mathsf{G}\}} \text{ is type II}\}$, *or* it is #P-hard. We may prove the dichotomy of $\mathscr{F}$ using the following three propositions and the dichotomy of two ternary signatures.

▶ **Proposition 6.1.** *Let $\mathbf{u} \in \mathbb{R}^2$ be a nonzero vector. Let $\mathbf{F} = \mathbf{v}_1^{\otimes 3} + \mathbf{v}_2^{\otimes 3}$ be a Boolean domain signature for nonzero $\mathbf{v}_1, \mathbf{v}_2 \in \mathbb{R}^2$ such that $\langle \mathbf{v}_1, \mathbf{v}_2\rangle = 0$. Let $\mathbf{G} = \mathbf{F} + \mathbf{u}^{\otimes 3}$. Then,* $\text{Holant}_2^*(\mathbf{F}, \mathbf{G})$ *is #P-hard unless $\mathbf{u} \sim \mathbf{v}_1$ or $\mathbf{u} \sim \mathbf{v}_2$.*

▶ **Proposition 6.2.** *Let $(a, b) \in \mathbb{R}^2$ be a nonzero vector. Let $\mathbf{F} = \boldsymbol{\beta}^{\otimes 3} + \overline{\boldsymbol{\beta}}^{\otimes 3}$ be a Boolean domain signature where $\boldsymbol{\beta} = \frac{1}{\sqrt{2}}(1, \mathfrak{i})^{\mathsf{T}}$. Let $\mathbf{G} = \mathbf{F} + (a, b)^{\otimes 3}$. Then,* $\text{Holant}_2^*(\mathbf{F}, \mathbf{G})$ *is #P-hard.*

▶ **Proposition 6.3.** *Let $\mathbf{F}$ and $\mathbf{G}$ be nondegenerate, real valued, symmetric, ternary signatures such that $\operatorname{supp} \mathbf{F} \subseteq \{\mathsf{B}, \mathsf{G}\}^*$ and $\operatorname{supp} \mathbf{G} \subseteq \{\mathsf{G}, \mathsf{R}\}^*$. Let $\mathbf{H} = \mathbf{F} + \mathbf{G}$. Then,* $\text{Holant}_3^*(\mathbf{F}, \mathbf{G}, \mathbf{H})$ *is #P-hard unless $\mathbf{F}, \mathbf{G}$ are both GenEQ.*

Proof. Assume $\text{Holant}_3^*(\mathbf{F}, \mathbf{G}, \mathbf{H})$ is not #P-hard. Then, we may assume that $\text{Holant}_3^*(\mathbf{F})$ and $\text{Holant}_3^*(\mathbf{G})$ are tractable. We write $\mathbf{F}, \mathbf{G}, \mathbf{H}$ as in Figure 9. We may realize $(=_{\mathsf{BG}})$ using $\mathbf{F}$ and $(=_{\mathsf{GR}})$ using $\mathbf{G}$.

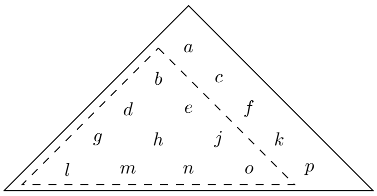


Figure 10 $\mathbf{F}'$ and $\langle \mathbf{F}', \mathbf{e}_2 \rangle$.

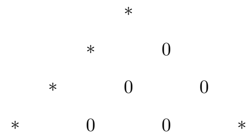


Figure 11 A generic $\mathsf{BG}|\mathsf{R}$ signature.

Suppose $d \neq 0$, then $\text{Holant}_2^*(\mathbf{G}^{*\to\{\mathsf{G},\mathsf{R}\}}, \mathbf{H}^{*\to\{\mathsf{G},\mathsf{R}\}})$ is #P-hard by Proposition 6.1 and Proposition 6.2 unless $\mathbf{G}$ is GenEQ because $\mathbf{H}^{*\to\{\mathsf{G},\mathsf{R}\}} = d\mathbf{e}_1^{\otimes 3} + \mathbf{G}^{*\to\{\mathsf{G},\mathsf{R}\}}$. If $\mathbf{G}$ is GenEQ, then $y, z = 0$ and thus it must be the case that $x \neq 0$ because otherwise $\mathbf{G}$ is degenerate. Then, we may make the same argument on $\text{Holant}_2^*(\mathbf{F}^{*\to\{\mathsf{B},\mathsf{G}\}}, \mathbf{H}^{*\to\{\mathsf{B},\mathsf{G}\}})$, which implies that the only way to escape hardness is for $\mathbf{F}$ to be GenEQ as well.

Therefore, we may assume that $d, x = 0$. Then, since $\mathbf{F}$ and $\mathbf{G}$ are assumed to be nondegenerate, $\mathbf{F}$ and $\mathbf{G}$ cannot be GenEQ, so b and c cannot both be zero and y and z cannot both be zero. Consider a unary signature $\mathbf{u} = (p, q, r)$ for p, q, r to be determined later. Let

$$\mathbf{H}_\mathbf{u} = \langle \mathbf{H}, \mathbf{u} \rangle = \begin{bmatrix} ap + bq & bp + cq & 0 \\ bp + cq & cp + yr & yq + zr \\ 0 & yq + zr & zq + wr \end{bmatrix}.$$

Regardless of the Boolean domain tractable type of $\mathbf{F}^{*\to\{\mathsf{B},\mathsf{G}\}}$, if $y \neq 0$, we can choose some p, q, r such that $(\mathbf{H}_\mathbf{u})^{*\to\{\mathsf{B},\mathsf{G}\}} = [ap + bq, bp + cq, cp + yr]$ is not compatible with $\mathbf{F}^{*\to\{\mathsf{B},\mathsf{G}\}}$ because $bp + cq$ and $cp + yr$ can be made to arbitrary values. Similarly, if $c \neq 0$, we can choose some p, q, r such that $\text{Holant}_2^*(\mathbf{G}^{*\to\{\mathsf{G},\mathsf{R}\}}, (\mathbf{H}_\mathbf{u})^{*\to\{\mathsf{G},\mathsf{R}\}})$ is #P-hard. If $c, y = 0$, then it must be that $b, z \neq 0$. We have $\mathbf{F}^{*\to\{\mathsf{B},\mathsf{G}\}} = [a, b, 0, 0]$ and $\mathbf{G}^{*\to\{\mathsf{G},\mathsf{R}\}} = [0, 0, x, y]$, and neither is a tractable signature. Therefore, $\text{Holant}_2^*(\mathbf{F}^{*\to\{\mathsf{B},\mathsf{G}\}})$ is #P-hard, contrary to the assumption made in the beginning. ◀

6.2 A Single Signature of Arity 4

▶ **Lemma 6.4.** *Let $\mathbf{F}$ be a nondegenerate, real-valued, symmetric signature of arity 4. Then, $\text{Holant}_3^*(\mathbf{F})$ is computable in polynomial time if there exists some real orthogonal matrix T such that one of the following conditions holds. Otherwise, it is #P-hard.*

1. $T^{\otimes 4}\mathbf{F} = a\mathbf{e}_1^{\otimes 4} + b\mathbf{e}_2^{\otimes 4} + c\mathbf{e}_3^{\otimes 4}$ *for some* $a, b, c \in \mathbb{R}$.
2. $T^{\otimes 4}\mathbf{F} = \boldsymbol{\beta}^{\otimes 4} + \overline{\boldsymbol{\beta}}^{\otimes 4} + \lambda\mathbf{e}_3^{\otimes 4}$ *where* $\boldsymbol{\beta} = \frac{1}{\sqrt{2}}(1, \mathfrak{i}, 0)^\intercal$ *for some* $\lambda \in \mathbb{R}$.

Proof. The tractability follows since $\{\mathbf{F}\}$ is in class $\mathscr{D}$ for both cases.

Assume $\text{Holant}_3^*(\mathbf{F})$ is not #P-hard. Let $\mathscr{F} = \{\langle \mathbf{F}, \mathbf{u} \rangle : \mathbf{u} \in \mathbb{R}^3\}$. Then, $\mathscr{F}$ is a vector space and $\text{Holant}_3^*(\mathscr{F}) \leq_T \text{Holant}_3^*(\mathbf{F})$, so we may apply the lemmas in Section B. We may assume that for any $\mathbf{G} \in \mathscr{F}$, $\text{Holant}_3^*(\mathbf{G})$ is tractable. Fix any $\mathbf{G} \in \mathscr{F}$. We may apply a real orthogonal holographic transformation T such that $\mathbf{G}' = T^{\otimes 3}\mathbf{G}$ is of the canonical form. Note that $\mathbf{G}' \in \mathscr{F}' = T\mathscr{F}$, and $\mathscr{F}'$ is a subspace of $\mathsf{S}_\mathbb{R}^3$.

Let $\mathbf{F}' = T^{\otimes 4}\mathbf{F}$. We write $\mathbf{F}'$ by extending the notation of Figure 1 to arity 4 signatures. We categorize the entries in Figure 10 in the following way. A *corner entry* is an entry at the corner of the triangle: a, l, p. An *outer entry* is an entry at the perimeter of the triangle: $a, b, d, g, l, m, n, o, p, k, f, c$. An *inner entry* is an entry at the inside of the triangle: e, h, j. The entries of a subsignature correspond to a smaller triangle. For example, the signature $\langle \mathbf{F}', \mathbf{e}_2 \rangle$ is the triangle given by the dashed lines in Figure 10. Note that $\langle \mathbf{F}', \mathbf{e}_i \rangle \in \mathscr{F}'$ for all $1 \leq i \leq 3$.

Suppose $\mathbf{G}'$ is a signature of rank 3 in type $\mathfrak{A}$. Then, by Lemma B.1, $\mathscr{F}'$ only contains GenEQ signatures. We claim that $\mathbf{F}'$ must be a GenEQ. This can be seen as follows. Note that a GenEQ signature has nonzero value at only the corner entries. The three subsignatures, $\langle \mathbf{F}', \mathbf{e}_i \rangle$ for $1 \le i \le 3$ must all be a GenEQ. For that to be possible, only the corner entries can be nonzero. Therefore, $\mathbf{F}'$ is a GenEQ.

Suppose $\mathbf{G}'$ is a signature of rank 3 in type $\mathfrak{B}$. Then, by Lemma B.2, $\mathscr{F}'$ only contains BG|R signatures that are type II on the domain $\{\mathsf{B}, \mathsf{G}\}$ (see Definition A.1). If an internal entry is nonzero, then there is some subsignature that is not BG|R (see Figure 11). Also, $\mathbf{F}'$ cannot have nonzero entries at the outer entries c, f, k, m, n, o, because that means there is a subsignature that is not BG|R. Therefore, all nonzero entries must be at the domain $\{\mathsf{B}, \mathsf{G}\}$ part a, b, d, g, l and p. Further, $[a, b, d, g, l]$ must satisfy the type II recurrence. This is because $\langle \mathbf{F}', \mathbf{e}_1 \rangle \in \mathscr{F}'$, and thus $(\langle \mathbf{F}', \mathbf{e}_1 \rangle)^{* \to \{\mathsf{B},\mathsf{G}\}} = [a, b, d, g]$ satisfies $a = -d$ and $b = -g$, and similarly, $[b, d, g, f]$ satisfies $b = -g$ and $d = -f$.

Now, suppose $\mathscr{F}'$ does not contain any signature of rank 3. If there is a signature of rank 2, we may apply the same arguments by using Lemma B.3 and Lemma B.4. The only difference is that p must be 0.

Suppose $\mathscr{F}'$ only contains rank 1 signatures. Then, by Lemma B.5, $\mathscr{F}' = \{\lambda \mathbf{v}^{\otimes 3} : \lambda \in \mathbb{R}\}$ for some $\mathbf{v} \in \mathbb{R}^3$. Then, $\langle \mathbf{F}', \mathbf{e}_i \rangle = \lambda_i \mathbf{v}^{\otimes 3}$ for some $\lambda_i \in \mathbb{R}$, which can be used to show that $\mathbf{F}' = \lambda \mathbf{v}^{\otimes 4}$ for some $\lambda \in \mathbb{R}$ by looking at the trinagular representation. ◀

References

1 Andrei Bulatov and Martin Grohe. The complexity of partition functions. *Theor. Comput. Sci.*, 348(2):148–186, December 2005. doi:10.1016/j.tcs.2005.09.011.

2 Andrei A. Bulatov. A dichotomy theorem for constraint satisfaction problems on a 3-element set. *J. ACM*, 53(1):66–120, January 2006. doi:10.1145/1120582.1120584.

3 Andrei A. Bulatov. The complexity of the counting constraint satisfaction problem. *J. ACM*, 60(5), October 2013. doi:10.1145/2528400.

4 Andrei A. Bulatov and Víctor Dalmau. Towards a dichotomy theorem for the counting constraint satisfaction problem. *Inf. Comput.*, 205(5):651–678, May 2007. doi:10.1016/j.ic.2006.09.005.

5 Jin-Yi Cai and Xi Chen. Complexity of counting CSP with complex weights. *Journal of the ACM (JACM)*, 64:1–39, 2011.

6 Jin-Yi Cai and Xi Chen. *Complexity Dichotomies for Counting Problems*. Cambridge University Press, 1 edition, 2017. doi:10.1017/9781107477063.

7 Jin-Yi Cai, Xi Chen, and Pinyan Lu. Graph homomorphisms with complex values: A dichotomy theorem: (extended abstract). In Samson Abramsky, Cyril Gavoille, Claude Kirchner, Friedhelm Meyer Auf Der Heide, and Paul G. Spirakis, editors, *Automata, Languages and Programming*, volume 6198, pages 275–286. Springer Berlin Heidelberg, 2010. Series Title: Lecture Notes in Computer Science. doi:10.1007/978-3-642-14165-2_24.

8 Jin-Yi Cai, Xi Chen, and Pinyan Lu. Nonnegative weighted #CSP: An effective complexity dichotomy. *SIAM J. Comput.*, 45(6):2177–2198, January 2016. doi:10.1137/15M1032314.

9 Jin-Yi Cai and Artem Govorov. Perfect matchings, rank of connection tensors and graph homomorphisms. *Comb. Probab. Comput.*, 31(2):268–303, 2022. doi:10.1017/S0963548321000286.

10 Jin-Yi Cai, Heng Guo, and Tyson Williams. A complete dichotomy rises from the capture of vanishing signatures. *SIAM Journal on Computing*, 45(5):1671–1728, 2016. doi:10.1137/15M1049798.

11 Jin-Yi Cai, Heng Guo, and Tyson Williams. The complexity of counting edge colorings and a dichotomy for some higher domain Holant problems. *Research in the Mathematical Sciences*, 3(1):18, 2016. doi:10.1186/s40687-016-0067-8.

12 Jin-Yi Cai and Jin Soo Ihm. Holant* dichotomy on domain size 3: A geometric perspective, 2025. arXiv:2504.14074.

13 Jin-Yi Cai, Pinyan Lu, and Mingji Xia. Holant problems and counting CSP. In *Proceedings of the Forty-First Annual ACM Symposium on Theory of Computing*, STOC '09, pages 715–724, New York, NY, USA, 2009. Association for Computing Machinery. doi:10.1145/1536414.1536511.

14 Jin-Yi Cai, Pinyan Lu, and Mingji Xia. Computational complexity of Holant problems. *SIAM Journal on Computing*, 40(4):1101–1132, 2011. doi:10.1137/100814585.

15 Jin-Yi Cai, Pinyan Lu, and Mingji Xia. Dichotomy for Holant problems with a function on domain size 3. In *Proceedings of the Twenty-Fourth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 1278–1295. Society for Industrial and Applied Mathematics, 2013. doi:10.1137/1.9781611973105.93.

16 Jin-Yi Cai, Pinyan Lu, and Mingji Xia. Dichotomy for Holant* problems on the Boolean domain. *Theory of Computing Systems*, 64(8):1362–1391, 2020. doi:10.1007/s00224-020-09983-8.

17 Pierre Comon, Gene Golub, Lek-Heng Lim, and Bernard Mourrain. Symmetric tensors and symmetric tensor rank. arXiv:0802.1681[math].

18 Nadia Creignou and Miki Hermann. Complexity of generalized satisfiability counting problems. *Inf. Comput.*, 125:1–12, 1996. doi:10.1006/INCO.1996.0016.

19 Martin Dyer, Leslie Ann Goldberg, and Mark Jerrum. The complexity of weighted Boolean #CSP. *SIAM Journal on Computing*, 38(5):1970–1986, 2009. doi:10.1137/070690201.

20 Martin Dyer and Catherine Greenhill. The complexity of counting graph homomorphisms. *Random Struct. Algorithms*, 17(3–4):260–289, October 2000. doi:10.1002/1098-2418(200010/12)17:3/4\%3C260::AID-RSA5\%3E3.0.CO;2-W.

21 Martin E. Dyer and David Richerby. An effective dichotomy for the counting constraint satisfaction problem. *SIAM J. Comput.*, 42(3):1245–1274, 2013. doi:10.1137/100811258.

22 Michael H. Freedman, László Lovász, and Alexander Schrijver. Reflection positivity, rank connectivity, and homomorphism of graphs. *Journal of the American Mathematical Society*, 20:37–51, 2004.

23 Leslie Ann Goldberg, Martin Grohe, Mark Jerrum, and Marc Thurley. A complexity dichotomy for partition functions with mixed signs. *SIAM Journal on Computing*, 39(7):3336–3402, 2010. doi:10.1137/090757496.

24 Pavol Hell and Jaroslav Nesetril. Graphs and homomorphisms. In *Oxford lecture series in mathematics and its applications*, 2004.

25 Yin Liu, Austen Z. Fan, and Jin-Yi Cai. Restricted holant dichotomy on domain sizes 3 and 4. *Theoretical Computer Science*, 1023:114931, 2025. doi:10.1016/j.tcs.2024.114931.

26 László Lovász. Operations with structures. *Acta Mathematica Academiae Scientiarum Hungarica*, 18:321–328, 1967.

27 Shuai Shao and Jin-Yi Cai. A dichotomy for real Boolean Holant problems. In *2020 IEEE 61st Annual Symposium on Foundations of Computer Science (FOCS)*, pages 1091–1102, 2020. doi:10.1109/FOCS46700.2020.00105.

28 Leslie G. Valiant. Accidental algorthims. In *Proceedings of the 47th Annual IEEE Symposium on Foundations of Computer Science*, FOCS '06, pages 509–517, USA, 2006. IEEE Computer Society. doi:10.1109/FOCS.2006.7.

A Boolean Domain Dichotomy Theorem

▶ **Definition A.1** (Definition 2.9 in [6]). *A signature* $[x_0, x_1, \ldots, x_n]$*, where* $n \geq 2$*, has type* $\mathrm{I}(a, b)$*, if there exist* a *and* b *(not both* 0*), such that* $ax_k + bx_{k+1} = ax_{k+2}$ *for* $0 \leq k \leq n-2$*. We say it it is of type* II*, if* $x_k = -x_{k+2}$ *for* $0 \leq k \leq n-2$*.*

▶ **Theorem A.2** (Theorem 2.12 in [6]). *Let $\mathcal{F}$ be a set of nondegenerate symmetric signatures over $\mathbb{C}$ in Boolean variables. Then* $\operatorname{Holant}_2^*(\mathcal{F})$ *is computable in polynomial time for the following three classes of $\mathcal{F}$. In all other cases,* $\operatorname{Holant}_2^*(\mathcal{F})$ *is #P-hard.*

1. *Every signature in $\mathcal{F}$ is of arity ≤ 2.*
2. *There exists a and b (not both 0, depending only on $\mathcal{F}$), such that every signature in $\mathcal{F}$ either (1) has type $\mathrm{I}(a, b)$ or (2) has arity 2 and is of the form $[2a\lambda, b\lambda, -2a\lambda]$.*
3. *Every signature in $\mathcal{F}$ either (1) has type II or (2) has arity 2 and is of the form $[\lambda, 0, \lambda]$.*

B Subspace of Signatures

Let $\mathsf{S}_{\mathbb{R}}^3$ denote the set of all real-valued symmetric ternary signatures. We list the dichotomy statements of $\operatorname{Holant}_3^*(\mathscr{F})$ for a subspace $\mathscr{F} \subseteq \mathsf{S}_{\mathbb{R}}^3$. The complete proof can be found in the full version [12].

▶ **Lemma B.1.** *Let $\mathscr{F}$ be a subspace of $\mathsf{S}_{\mathbb{R}}^3$. Suppose $\mathbf{F} = c_1\mathbf{e}_1^{\otimes 3} + c_2\mathbf{e}_2^{\otimes 3} + c_3\mathbf{e}_3^{\otimes 3} \in \mathscr{F}$ for nonzero $c_1, c_2, c_3 \in \mathbb{R}$. Then,* $\operatorname{Holant}_3^*(\mathscr{F})$ *is #P-hard unless every $\mathbf{G} \in \mathscr{F}$ is a GenEQ.*

▶ **Lemma B.2.** *Let $\mathscr{F}$ be a subspace of $\mathsf{S}_{\mathbb{R}}^3$. Suppose $\mathbf{F} = c(\boldsymbol{\beta}^{\otimes 3} + \overline{\boldsymbol{\beta}}^{\otimes 3}) + \lambda\mathbf{e}_3^{\otimes 3} \in \mathscr{F}$ where $\boldsymbol{\beta} = \frac{1}{\sqrt{2}}(1, \mathfrak{i}, 0)^\mathsf{T}$ and nonzero $c, \lambda \in \mathbb{R}$. Then,* $\operatorname{Holant}_3^*(\mathscr{F})$ *is #P-hard unless every $\mathbf{G} \in \mathscr{F}$ is such that $\mathbf{G}$ is $\mathsf{BG|R}$ and $\mathbf{G}^{*\to\{\mathsf{B},\mathsf{G}\}} = [x, y, -x, -y]$ for some $x, y \in \mathbb{R}$.*

▶ **Lemma B.3.** *Let $\mathscr{F}$ be a subspace of $\mathsf{S}_{\mathbb{R}}^3$ such that all signatures in $\mathscr{F}$ are of rank at most 2. Suppose $\mathbf{F} = c_1\mathbf{e}_1^{\otimes 3} + c_2\mathbf{e}_2^{\otimes 3} \in \mathscr{F}$ for nonzero $c_1, c_2 \in \mathbb{R}$. Then,* $\operatorname{Holant}_3^*(\mathscr{F})$ *is #P-hard unless every $\mathbf{G} \in \mathscr{F}$ is $d_1\mathbf{e}_1^{\otimes 3} + d_2\mathbf{e}_2^{\otimes 3}$ for some $d_1, d_2 \in \mathbb{R}$.*

▶ **Lemma B.4.** *Let $\mathscr{F}$ be a subspace of $\mathsf{S}_{\mathbb{R}}^3$ such that all signatures in $\mathscr{F}$ are of rank at most 2. Suppose $\mathbf{F} = \boldsymbol{\beta}^{\otimes 3} + \overline{\boldsymbol{\beta}}^{\otimes 3} \in \mathscr{F}$ where $\boldsymbol{\beta} = \frac{1}{\sqrt{2}}(1, \mathfrak{i}, 0)^\mathsf{T}$. Then,* $\operatorname{Holant}_3^*(\mathscr{F})$ *is #P-hard unless every $\mathbf{G} \in \mathscr{F}$ is such that $\operatorname{supp}\mathbf{G} \subseteq \{\mathsf{B}, \mathsf{G}\}^*$ and $\mathbf{G}^{*\to\{\mathsf{B},\mathsf{G}\}} = [x, y, -x, -y]$ for some $x, y \in \mathbb{R}$.*

▶ **Lemma B.5.** *Let $\mathscr{F}$ be a subspace of $\mathsf{S}_{\mathbb{R}}^3$ such that all signatures in $\mathscr{F}$ are of rank at most 1. Then, there exists some $\mathbf{v} \in \mathbb{R}^3$ such that $\mathbf{F} = \{\lambda\mathbf{v}^{\otimes 3} : \lambda \in \mathbb{R}\}$.*