Restricted Holant Dichotomy on Domains 3 and 4

Yin Liu, Austen Z. Fan, and Jin-Yi Cai

University of Wisconsin-Madison, Madison WI 53715, USA
{yinl,afan,jyc}@cs.wisc.edu

Abstract. Holant$^*(f)$ denotes a class of counting problems specified by a constraint function f. We prove complexity dichotomy theorems for Holant$^*(f)$ in two settings: (1) f is any arity-3 real-valued function on input of domain size 3. (2) f is any arity-3 $\{0,1\}$-valued function on input of domain size 4.

Keywords: Holant problem · Dichotomy · Higher domain.

1 Introduction and background

Counting problems arise in many branches in computer science, machine learning and statistical physics. Holant problems encompass a broad class of counting problems [1,2,7,8,10,12,16,17,19,22,23,24]. For symmetric constraint functions (a.k.a. signatures) this is also equivalent to edge-coloring models [20,21]. These problems extend counting constraint satisfaction problems. Freedman, Lovász and Schrijver proved that some prototypical Holant problems, such as counting perfect matchings, cannot be expressed as vertex-coloring models known as graph homomorphisms [14,18]. The complexity classification program of counting problems is to classify the computational complexity of these problems.

Formally, a Holant problem on domain D is defined on a graph $G = (V, E)$ where edges are variables and vertices are constraint functions. Given a set of constraint functions $\mathcal{F}$ defined on D, a *signature grid* $\Omega = (G, \pi)$ assigns to each vertex $v \in V$ an $f_v \in \mathcal{F}$. The aim is to compute the following partition function

$$\text{Holant}_\Omega = \sum_{\sigma:E \to D} \prod_{v \in V} f_v \left(\sigma|_{E(v)} \right).$$

The computational problem is denoted by Holant($\mathcal{F}$). E.g., on the Boolean domain, it is over all $\{0,1\}$-edge assignments. On domain size 3, it is over all $\{R, G, B\}$-edge assignments, signifying three colors Red, Green and Blue. On domain size 4, it is over all $\{R, G, B, W\}$-edge assignments. On the Boolean domain, if every vertex has the EXACT-ONE function (which evaluates to 1 if exactly one incident edge is 1, and evaluates to 0 otherwise), then the partition function gives the number of perfect matchings. On domain size k, if every vertex has the ALL-DISTINCT function, then the partition function gives the number of valid k-edge colorings.

A *symmetric* signature is a function that is invariant under any permutation of its variables. The value of such a signature depends only on the numbers of each color assigned to its input variables. The number of variables is its arity; unary, binary, ternary signatures have arities 1, 2, 3. We denote a symmetric ternary signature g on domain size 3 by a "triangle" consisting of 10 numbers:

$$
\begin{array}{ccccccc}
& & & g_{3,0,0} & & & \\
& & g_{2,1,0} & & g_{2,0,1} & & \\
& g_{1,2,0} & & g_{1,1,1} & & g_{1,0,2} & \\
g_{0,3,0} & & g_{0,2,1} & & g_{0,1,2} & & g_{0,0,3}
\end{array}
$$

where $g_{i,j,k}$ is the value on inputs having i Red, j Green and k Blue. Similarly, we denote a symmetric ternary signature g on domain size 4 by a "tetrahedron":

While much progress has been made for the classification of counting CSP [3] [4,5,13], and some progress for Holant problems [6,11,15], classifying Holant problems on higher domains is particularly challenging. One of the few existing work on a higher domain is [11], in which a dichotomy for Holant$^*(f)$ is proved where f is a ternary complex symmetric function on domain size 3 and the $*$ means all unary functions are available. (Note that Holant problems with signatures of arity ≤ 2 are all P-time tractable; the interesting case where both tractable and #P-hardness occur starts with ternary signatures.)

In this work, we attempt to extend this to Holant problems on domain size 4. Our effort only met with partial success. We are able to prove a complexity dichotomy for Holant$^*(f)$ for any $\{0,1\}$-valued symmetric ternary constraint function f (see Theorem 4).

Our technique is to try to reduce a domain 4 problem to a domain 3 problem, and then analyze the situation using the existing domain 3 dichotomy [11]. To do so, we will need to be able to construct (or interpolate) a suitable constraint function that allows us to effectively restrict the problem to a domain 3 problem, where the new domain elements are superpositions of old domain elements under a holographic transformation. This turns out to be a nontrivial task. And one

reason that we cannot extend to a more general domain 4 dichotomy is that for some real-valued signatures, it is impossible to construct such a constraint function. On the other hand, for $\{0,1\}$-valued domain 4 signatures, we are able to succeed in this plan (using several different constructions).

To carry out this plan, we use the domain 3 dichotomy [11] extensively. This motivates us to examine the domain 3 dichotomy more closely, when the constraint function f is real-valued. Since the domain 3 dichotomy [11] applies to all complex-valued functions it certainly also applies to real-valued functions[⋆]. However, we found out that applying the dichotomy for complex-valued functions directly is very cumbersome, so much so that the attempt to use it for our exploration in domain 4 grinds to a halt.

So, we return to domain size 3, and found that there is a cleaner form of the dichotomy of Holant$^*(f)$ on domain size 3 for real-valued f. This turns out to be a non-trivial adaptation, as we prove that certain tractable forms in the complex case *cannot* occur for real-valued signatures (see Theorem 2). In the proof of this real Holant* dichotomy, orthogonal holographic transformations are heavily used.

Armed with this more effective form of the domain 3 dichotomy, we return to Holant problems on domain 4, and prove a dichotomy for Holant$^*(f)$ where f is $\{0,1\}$-valued (see Theorem 4). We use several strategies that are more generally applicable and are worth mentioning (see subsection 3.1). For general real-valued ternary symmetric signatures on domain 4, we prove Theorem 3. It gives some broad classes of constraint functions that define P-time tractable Holant problems. We conjecture that this is actually a complexity dichotomy.

Some preliminaries We can picture a signature as a vertex with several dangling edges as its input variables. Connecting a unary signature u to another signature f of arity $r \geq 1$ creates a signature of arity $r - 1$. If f is symmetric then this does not depend on which variable (dangling edge of f) u is connected to, and the resulting signature is denoted by $\langle f, u \rangle$. In particular, if f is also a unary, then $\langle f, u \rangle$ is a scalar value, equal to the dot product of their signature entries. One should note that, for complex unary signatures, this dot product (without conjugation) is not the usual inner product and it is possible that $\langle u, u \rangle = 0$ for $u \neq 0$. We call a vector u *isotropic* if $\langle u, u \rangle = 0$ (including $u = 0$).

We now introduce *holographic transformations*. We use Holant $(\mathcal{R} \mid \mathcal{G})$ to denote bipartite Holant problems on bipartite graphs $H = (U, V, E)$, where each signature on a vertex in U or V is from $\mathcal{R}$ or $\mathcal{G}$, respectively. Suppose T is an invertible matrix of the same size as that of the domain. We say that there is a holographic transformation from Holant $(\mathcal{R} \mid \mathcal{G})$ to Holant $(\mathcal{R}' \mid \mathcal{G}')$ by T, if $\mathcal{R}' = \mathcal{R}T^{-1}$ and $\mathcal{G}' = T\mathcal{G}$, where $\mathcal{R}T^{-1} = \{f(T^{-1})^{\otimes r(f)} \mid f \in \mathcal{R}\}$, $T\mathcal{G} = \{T^{\otimes r(f)}f \mid f \in \mathcal{G}\}$ and $r(f)$ is the arity of f. Here each signature is writ-

---

[⋆] There is a slight issue that for a real-valued f, Holant$^*(f)$ naturally refers to having free real-valued unary functions, while the existing Holant* dichotomy for complex-valued f assumes all complex-valued unary functions are available for free. In Lemma 9 we address this technical difficulty.

ten as a column/row vector in lexicographical order as a truth-table. We also write Tf for $T^{\otimes r(f)}f$ when the arity $r(f)$ is clear.

Theorem 1 (Valiant's Holant Theorem [23]). *Suppose there is a holographic transformation from* Holant $(\mathcal{R} \mid \mathcal{G})$ *to* Holant $(\mathcal{R}' \mid \mathcal{G}')$, *then* Holant $(\mathcal{R} \mid \mathcal{G})$ $\equiv_T$ Holant $(\mathcal{R}' \mid \mathcal{G}')$, *where* $\equiv_T$ *means equivalence up to a P-time reduction.*

Therefore, if there is a holographic transformation from Holant $(\mathcal{G} \mid \mathcal{R})$ to Holant $(\mathcal{G}' \mid \mathcal{R}')$, then one problem is in P iff the other one is, and similarly one problem is #P-hard iff the other one is. For any general graph, we can make it bipartite by adding an additional vertex on each edge (thus forming the vertex-edge incidence graph), and assigning those new vertices the binary EQUALITY signature $(=_2)$ of the corresponding domain size. Note that for the binary EQUALITY, if T is orthogonal, then it is unchanged under the holographic transformation by T. Hence, Holant$(\mathcal{F}) \equiv_T$ Holant $(=_2\mid \mathcal{F}) \equiv_T$ Holant$(T\mathcal{F})$ for any orthogonal matrix T.

2   A real dichotomy for Holant$^*(f)$ on domain 3

In this section we prove a complexity dichotomy of Holant$^*(f)$ for any *real-valued* symmetric ternary function f over the domain $\{R, G, B\}$. We investigate the three tractable forms of Theorems 3.1 and 3.2 in [11] when f is real-valued. It turns out that they take more special forms, and one tractable case for complex-valued f does not occur for real-valued f. However, complex tensors are still needed to express one of the two tractable forms. Lemmas 2, 3 and 4 address each of the three tractable families.

Lemma 1. *For all $\beta \in \mathbb{C}^3$, if $\langle \beta, \beta \rangle = 0$ then there exists a 3-by-3 real orthogonal matrix T, such that $T\beta = c(1, i, 0)^T$ where $c \in \mathbb{R}$.*

Proof. Write $\beta = \gamma + \delta i$, $\gamma, \delta \in \mathbb{R}^3$. Then $0 = \langle \beta, \beta \rangle = \langle \gamma, \gamma \rangle - \langle \delta, \delta \rangle + 2\langle \gamma, \delta \rangle i$. Considering its real and imaginary parts separately, $\|\gamma\| = \|\delta\|$ and $\gamma \perp \delta$. Then there exists a real orthogonal T, such that $T\gamma = ce_1$ and $T\delta = ce_2$, where $c = \|\gamma\| \in \mathbb{R}$. It follows that $T\beta = T(\gamma + \delta i) = c(1, i, 0)^T$.

Lemma 2. *If there exist $\alpha, \beta, \gamma \in \mathbb{C}^3$, s.t., $f = \alpha^{\otimes 3} + \beta^{\otimes 3} + \gamma^{\otimes 3}$, $\langle \alpha, \beta \rangle = \langle \beta, \gamma \rangle = \langle \gamma, \alpha \rangle = 0$, and f is real-valued, then there exist $\alpha', \beta', \gamma' \in \mathbb{R}^3$, s.t., $f = \alpha'^{\otimes 3} + \beta'^{\otimes 3} + \gamma'^{\otimes 3}, \langle \alpha', \beta' \rangle = \langle \beta', \gamma' \rangle = \langle \gamma', \alpha' \rangle = 0$. Thus, there is a real orthogonal transformation T, such that $Tf = ae_1^{\otimes 3} + be_2^{\otimes 3} + ce_3^{\otimes 3}$, for some $a, b, c \in \mathbb{R}$.*

Proof. Let $M_i = \langle f, e_i \rangle = \alpha_i \alpha^{\otimes 2} + \beta_i \beta^{\otimes 2} + \gamma_i \gamma^{\otimes 2}, i = 1, 2, 3$, then M_i is a real symmetric matrix.

If there is any v among $\{\alpha, \beta, \gamma\}$ that is non-isotropic, i.e., $\langle v, v \rangle \neq 0$, then by symmetry, assume it is α. Then $\alpha \neq 0$. At least one of $\alpha_i \neq 0$. Say $\alpha_1 \neq 0$.

We have $M_i \alpha = \lambda_i \alpha$, where $\lambda_i = \alpha_i \langle \alpha, \alpha \rangle$. So λ_i is an eigenvalue of a real symmetric matrix M_i, therefore it is real ($i = 1, 2, 3$). As $\langle \alpha, \alpha \rangle \neq 0$ and $\alpha_1 \neq 0$,

$\frac{\alpha_i}{\alpha_1} = \frac{\lambda_i}{\lambda_1}$ is real and is well defined $(i = 1, 2, 3)$. We can then write $\alpha = \mu u$, where $\mu \in \mathbb{C}, u \in \mathbb{R}^3$ and $\|u\| = 1$. As $\alpha \neq 0$, we have $\mu \neq 0$ and $\langle u, \beta \rangle = \langle u, \gamma \rangle = 0$.

Thus $f = \mu^3 u^{\otimes 3} + \beta^{\otimes 3} + \gamma^{\otimes 3}$. We have $\langle f, u \rangle = \mu^3 u^{\otimes 2}$. Since f and u are both real, and $u \neq 0$, we have $\mu^3 \in \mathbb{R}$. Thus, $\mu^3 = t^3$ for some real $t \in \mathbb{R}$. It follows that $\alpha^{\otimes 3} = t^3 u^{\otimes 3} = (tu)^{\otimes 3}$.

Then we replace α with tu. Similarly, if β or γ is a non-isotropic vector, we can replace it with a real vector, without changing f. Thus we get a new form $f = \alpha^{\otimes 3} + \beta^{\otimes 3} + \gamma^{\otimes 3}$ where α, β, γ are either real or isotropic (and since the zero vector is real we may further assume α, β, γ are either real or nonzero isotropic.)

If they are all real, then we can use a real orthogonal matrix T to transform f, i.e., $Tf = ae_1^{\otimes 3} + be_2^{\otimes 3} + ce_3^{\otimes 3}$, for some $a, b, c \in \mathbb{R}$. Then we are done.

Now suppose there is at least one nonzero isotropic vector among $\{\alpha, \beta, \gamma\}$. W.o.l.o.g., we can assume γ is nonzero and isotropic. By Lemma 1, there exists a real orthogonal T, s.t., $T\gamma = r(1, i, 0)^T, r \in \mathbb{R} \setminus \{0\}$. Then since $\langle T\alpha, T\gamma \rangle = \langle T\beta, T\gamma \rangle = 0$, $T\alpha, T\beta$ each must have the form $(c, ci, d)^T$. If, in addition, α is also isotropic, then $T\alpha$ must have the form $(c, ci, 0)^T$, which is a multiple of $T\gamma$. As γ is nonzero, α is hence a multiple of γ. Then α can be absorbed into γ and form a new isotropic vector γ', and $\langle \gamma', \beta \rangle = 0$, $f = \beta^{\otimes 3} + \gamma'^{\otimes 3}$. We have the same argument for β. So w.o.l.o.g., we can write $f = \alpha^{\otimes 3} + \beta^{\otimes 3} + \gamma^{\otimes 3}$, and there is *at most one*, and therefore, *exactly one*, nonzero isotropic vector among $\{\alpha, \beta, \gamma\}$, and the others are real vectors and could be zero. We have $f = \gamma^{\otimes 3} + R$, where R is some real-valued tensor. Hence $\gamma^{\otimes 3}$ is real. Since $\gamma \neq 0$, we have $\gamma^{\otimes 3} \neq 0$. But then, $\langle \langle \langle \gamma^{\otimes 3}, \gamma \rangle, \gamma \rangle, \gamma \rangle = \langle \gamma^{\otimes 3}, \gamma^{\otimes 3} \rangle \neq 0$. In particular, $0 \neq \langle \gamma^{\otimes 3}, \gamma \rangle = \langle \gamma, \gamma \rangle \gamma^{\otimes 2} = 0$. This is a contradiction.

Lemma 3. *If there exist $\alpha, \beta_1, \beta_2 \in \mathbb{C}^3, s.t., f = \alpha^{\otimes 3} + \beta_1^{\otimes 3} + \beta_2^{\otimes 3}, \langle \alpha, \beta_i \rangle = \langle \beta_i, \beta_i \rangle = 0, i = 1, 2$, and f is real-valued, then there is a real orthogonal transformation T, such that $cTf = \epsilon(\beta_0^{\otimes 3} + \overline{\beta_0}^{\otimes 3}) + \lambda e_3^{\otimes 3}$, where $\beta_0 = (1, i, 0)^T, \epsilon \in \{0, 1\}$ and $c, \lambda \in \mathbb{R}, c \neq 0$. Thus, there exist $\alpha \in \mathbb{R}^3$ and $\beta \in \mathbb{C}^3, s.t., f = \alpha^{\otimes 3} + \beta^{\otimes 3} + \overline{\beta}^{\otimes 3}$, $\langle \alpha, \beta \rangle = \langle \beta, \beta \rangle = 0$.*

Proof. First we assume $\beta_1 = \beta_2 = 0$. As f is real, by a similar argument as in the proof of Lemma 2, we know that there exists an $\alpha' \in \mathbb{R}^3, f = \alpha^{\otimes 3} = \alpha'^{\otimes 3}$. Thus there is a real orthogonal T, $T\alpha' = (0, 0, d)^T, d \in \mathbb{R}$, so $Tf = (T\alpha')^{\otimes 3} = d^3 e_3^{\otimes 3}$.

Now w.o.l.o.g., we can assume $\beta_1 \neq 0$.

Suppose $\langle \alpha, \alpha \rangle = 0$. By Lemma 1, there exists a real orthogonal transformation T, such that $T\beta_1 = c(1, i, 0)^T$ for some $c \neq 0$. Since $\langle T\alpha, T\beta_1 \rangle = 0$, and that α is also isotropic, $T\alpha$ must have the form $(a, ai, 0)^T$, thus a multiple of $T\beta_1$. So, α is a multiple of β_1. Then α can be absorbed into β_1 and form a new isotropic vector $\mu\beta_1$ for some μ. Then we replace β_1 with $\mu\beta_1$ and $f = \beta_1^{\otimes 3} + \beta_2^{\otimes 3}$.

Else, $\langle \alpha, \alpha \rangle \neq 0$, then $\alpha \neq 0$. Let $M_i = \langle f, e_i \rangle, i = 1, 2, 3$, then M_i is a real symmetric matrix. $M_i \alpha = \alpha_i \langle \alpha, \alpha \rangle \alpha$. By the same argument as in the proof of Lemma 2, there exists an $\alpha' \in \mathbb{R}^3, s.t., \alpha^{\otimes 3} = \alpha'^{\otimes 3}$.

Hence in both cases, we get a new form $f = \alpha^{\otimes 3} + \beta_1^{\otimes 3} + \beta_2^{\otimes 3}$, where $\alpha \in \mathbb{R}^3$ is real (possibly 0), $\langle \alpha, \beta_i \rangle = \langle \beta_i, \beta_i \rangle = 0, i = 1, 2$. Since α is real, there is some

real orthogonal T, $T\alpha = te_3 = (0,0,t)^T, t \in \mathbb{R}$. If $\beta_1 = \beta_2 = 0$, we are done. Thus we may assume $\beta_1 \neq 0$.

Case 1: $\alpha \neq 0$, i.e., $t \neq 0$. Since $\langle T\beta_i, T\alpha \rangle = 0, i = 1,2$ and β_i is isotropic, we have $T\beta_1 = u(1, \pm i, 0)^T$, $T\beta_2 = v(1, \pm i, 0)^T, u \neq 0$. If β_2 is a multiple of β_1, it can be absorbed into β_1 and form a new isotropic vector $s(1, \pm i, 0)^T$ for some s. Then we get $Tf = (te_3)^{\otimes 3} + s^3((1, \pm i, 0)^T)^{\otimes 3}$. As Tf is real and t is real, we get $s = 0$ and we are done. Else, β_1, β_2 are independent, i.e., $Tf = (te_3)^{\otimes 3} + u^3((1, \pm i, 0)^T)^{\otimes 3} + v^3((1, \mp i, 0)^T)^{\otimes 3}$ where $uv \neq 0$. As Tf is real, it follows that $u^3 + v^3 \in \mathbb{R}$ and $u^3 - v^3 = 0$, and thus $u^3 = v^3 \in \mathbb{R}$. So $\frac{1}{u^3}Tf = \beta_0^{\otimes 3} + \overline{\beta_0}^{\otimes 3} + \frac{t^3}{u^3}e_3^{\otimes 3}$, where $\beta_0 = (1, i, 0)^T$.

Case 2: $\alpha = 0$, i.e., $t = 0$. We have $f = \beta_1^{\otimes 3} + \beta_2^{\otimes 3}$, $\langle \beta_i, \beta_i \rangle = 0$, $i = 1,2$, and $\beta_1 \neq 0$. By Lemma 1, there exists a real orthogonal T, $T\beta_1 = u(1, i, 0)^T, u \in \mathbb{R} \setminus \{0\}$. If $\beta_2 = 0$, then $Tf = T\beta_1^{\otimes 3} = u^3((1, i, 0)^T)^{\otimes 3}$, where the LHS is real but the RHS is not, which is a contradiction. So we have $\beta_2 \neq 0$. Let $\beta_i' = T\beta_i$, and we have $Tf = \beta_1'^{\otimes 3} + \beta_2'^{\otimes 3}$. Let $M_i = \langle Tf, e_i \rangle = \beta_{1i}'\beta_1'^{\otimes 2} + \beta_{2i}'\beta_2'^{\otimes 2}, i = 1,2,3$. Both $\beta_1', \beta_2' \neq 0$. Then,

$$\begin{cases} M_i\beta_1' = \lambda_{2i}\beta_2' \\ M_i\beta_2' = \lambda_{1i}\beta_1', \end{cases} \quad \text{where} \quad \begin{cases} \lambda_{1i} = \beta_{1i}'\langle \beta_1', \beta_2' \rangle \\ \lambda_{2i} = \beta_{2i}'\langle \beta_1', \beta_2' \rangle. \end{cases}$$

Applying M_i twice, we get $M_i^2\beta_1' = \lambda_{2i}\lambda_{1i}\beta_1'$. Since M_i is real symmetric, so is M_i^2. As $\beta_1' \neq 0$, $\lambda_{2i}\lambda_{1i}$ is an eigenvalue of M_i^2, therefore real, i.e., $\beta_{1i}'\beta_{2i}'\langle \beta_1', \beta_2' \rangle^2$ is real, $i = 1,2,3$. Recall $\beta_1' = u(1, i, 0)^T$, and now let $\beta_2' = (x, y, z)^T \in \mathbb{C}^3$. Let $\tau = x+yi$, and $\mu = \langle \beta_1', \beta_2' \rangle = u\tau$. Then we have $\begin{cases} \beta_{11}'\beta_{21}'\mu^2 = ux \cdot u^2\tau^2 = u^3\tau^2 x \\ \beta_{12}'\beta_{22}'\mu^2 = uiy \cdot u^2\tau^2 = u^3\tau^2 yi, \end{cases}$
both of which are real. Since $u \in \mathbb{R} \setminus \{0\}$, we know $\tau^2 x, \tau^2 yi \in \mathbb{R}$.

If $\tau = 0$, then $y = xi$ and hence $z = 0$ (because β_2' is isotropic). It follows that β_2' can be absorbed into β_1'. We can then rewrite $Tf = \lambda((1, i, 0)^T)^{\otimes 3}$ for some λ. As Tf is real, we know $\lambda = 0$ and we are done.

Now we can assume $\tau \neq 0$. Since both $\tau^2 x, \tau^2 yi \in \mathbb{R}$, adding them we know τ^3 is real. Then for some $k \in \{0,1,2\}$, $\omega^k\tau \in \mathbb{R}$, where $\omega^3 = 1$. Replacing β_2' by $\omega^k\beta_2'$, which satisfies $(\omega^k\beta_2')^{\otimes 3} = \beta_2'^{\otimes 3}$, we may assume $\tau \in \mathbb{R} \setminus \{0\}$.

We have $y = (x - \tau)i$. Since $\tau^2 x \in \mathbb{R}$, we know $x \in \mathbb{R}$. Because $Tf = \beta_1'^{\otimes 3} + \beta_2'^{\otimes 3}$ is real, we know $\begin{cases} (Tf)_{2,1,0} = u^3 i + x^2 y = (u^3 + x^2(x - \tau))i \in \mathbb{R} \\ (Tf)_{0,3,0} = y^3 - u^3 i = -(u^3 + (x - \tau)^3)i \in \mathbb{R}. \end{cases}$
From the fact that $x, \tau \in \mathbb{R}$ and the above relations, we know $x - \tau = -u$ (which gives $y = -ui$), and $x = \pm u$. As β_2' is isotropic, it follows that $z = 0$ and hence $\beta_2' = \pm u(1, \mp i, 0)^T$. If it is $+u(1, -i, 0)^T$, then $\frac{1}{u^3}Tf = \beta_0^{\otimes 3} + \overline{\beta_0}^{\otimes 3}$. If it is $-u(1, +i, 0)^T$, then $f = 0$.

Lemma 4. *If f is a real-valued signature, s.t., there exist $\beta, \gamma \in \mathbb{C}^3$, $f = f_\beta + \beta^{\otimes 2} \otimes \gamma + \beta \otimes \gamma \otimes \beta + \gamma \otimes \beta^{\otimes 2}$, where $\beta \neq 0$, $\langle \beta, \beta \rangle = 0$, and f_β is a complex ternary signature satisfying $\langle f_\beta, \beta \rangle = 0$, then there exists a real orthog-*

onal transformation T such that $Tf = \lambda e_3^{\otimes 3}$, $\lambda \in \mathbb{R}$. It implies that there is an $\alpha \in \mathbb{R}^3$, $f = \alpha^{\otimes 3}$.

Proof. Let $\beta = (\beta_1, \beta_2, \beta_3)^T$, we have $\langle f, \beta \rangle = \langle \gamma, \beta \rangle \beta^{\otimes 2}$. Let $M_i = \langle f, e_i \rangle$, for $i = 1, 2, 3$, so M_i is real symmetric. Then $M_i \beta = \langle \langle f, e_i \rangle, \beta \rangle = \langle \langle f, \beta \rangle, e_i \rangle = \beta_i \langle \gamma, \beta \rangle \beta$, for $i = 1, 2, 3$. As $\beta \neq 0$, we know $\beta_i \langle \gamma, \beta \rangle$ is a real eigenvalue of M_i.

Suppose $\langle \gamma, \beta \rangle \neq 0$. Since $\beta \neq 0$, we can w.o.l.o.g. assume $\beta_1 \neq 0$. Then $\frac{\beta_i}{\beta_1} = \frac{\beta_i \langle \gamma, \beta \rangle}{\beta_1 \langle \gamma, \beta \rangle}$ is real and well defined ($i = 1, 2, 3$). We can then write $\beta = \lambda u$, where $\lambda \in \mathbb{C}, u \in \mathbb{R}^3$ and $\lambda \neq 0, u \neq 0$. In particular, $0 \neq \lambda^2 \langle u, u \rangle = \langle \beta, \beta \rangle = 0$. This is a contradiction.

So we have $\langle \gamma, \beta \rangle = 0$. Then, $\langle f, \beta \rangle = 0$. From Lemma 1, we know there exists some real orthogonal matrix T such that $T\beta = t\beta_0$ where $t \in \mathbb{R}, \beta_0 = (1, i, 0)^T$. Since $\beta \neq 0$, we have $t \neq 0$. Then $0 = \langle f, \beta \rangle = \langle T\beta, Tf \rangle$. So Tf has the form

$$
\begin{matrix}
& & a & & \\
& ai & & b & \\
-a & & bi & & c \\
-ai & & -b & & ci & & d
\end{matrix}
$$

Since Tf is real, we get $a = b = c = 0$. Thus, $Tf = de_3^{\otimes 3}$ for some $d \in \mathbb{R}$.

Theorem 2. *Let f be a* real-valued *symmetric ternary function over domain $\{R, G, B\}$. Then* $\mathrm{Holant}^*(f)$ *is #P-hard unless the function f in expressible as one of the following two forms, in which case the problem is in FP.*

1. $f = \alpha^{\otimes 3} + \beta^{\otimes 3} + \gamma^{\otimes 3}$ where $\alpha, \beta, \gamma \in \mathbb{R}^3$ and $\langle \alpha, \beta \rangle = \langle \beta, \gamma \rangle = \langle \gamma, \alpha \rangle = 0$.
2. $f = \alpha^{\otimes 3} + \beta^{\otimes 3} + \overline{\beta}^{\otimes 3}$ where $\alpha \in \mathbb{R}^3$, $\langle \alpha, \beta \rangle = \langle \beta, \beta \rangle = 0$.

This is equivalent to the existence of a real orthogonal transformation T, s.t.,

1. $Tf = ae_1^{\otimes 3} + be_2^{\otimes 3} + ce_3^{\otimes 3}$ for some $a, b, c \in \mathbb{R}$.
2. $cTf = \epsilon(\beta_0^{\otimes 3} + \overline{\beta_0}^{\otimes 3}) + \lambda e_3^{\otimes 3}$ where $\beta_0 = (1, i, 0)^T$, $\epsilon \in \{0, 1\}$, and for some $c, \lambda \in \mathbb{R}$ and $c \neq 0$.

Proof. This follows from Theorems 3.1 and 3.2 in [11] and Lemmas 2, 3, and 4.

Theorem 2 is the adapted real dichotomy for $\mathrm{Holant}^*(f)$ with any real-valued signature f of arity 3 on domain 3. We showed that in the real case, we can take a real orthogonal transformation to the corresponding canonical forms. Also the third tractable case in the scenario of complex dichotomy does not exist anymore in the real case, which will simplify and expedite the analysis of further exploration of real dichotomies on domain 4.

3 $\mathrm{Holant}^*(f)$ dichotomy for $\{0, 1\}$-valued f on domain 4

A dichotomy for $\mathrm{Holant}^*(f)$ on domain 2 has been known [9]. With the real dichotomy for $\mathrm{Holant}^*(f)$ on domain 3, we will now explore the situation on domain 4. In this section we give a dichotomy for a single $\{0, 1\}$-valued arity-3 signature f taking input values on domain 4.

Theorem 3. *Let f be a real symmetric ternary function defined on a domain of size 4. If there is a real orthogonal transformation T such that Tf has one of the following forms, then $\mathrm{Holant}^*(f)$ is P-time computable, where $\beta_0 = (1, i, 0, 0)^T$ and $\beta_1 = (0, 0, 1, i)^T$.*

1. *For some $a, b, c, d \in \mathbb{R}$, $Tf = ae_1^{\otimes 3} + be_2^{\otimes 3} + ce_3^{\otimes 3} + de_4^{\otimes 3}$.*
2. *For some $c, \lambda_1, \lambda_2 \in \mathbb{R}$, and $c \neq 0$, $cTf = \beta_0^{\otimes 3} + \overline{\beta_0}^{\otimes 3} + \lambda_1 e_3^{\otimes 3} + \lambda_2 e_4^{\otimes 3}$.*
3. *For some $\lambda_1, \lambda_2 \in \mathbb{R}$, $Tf = \lambda_1(\beta_0^{\otimes 3} + \overline{\beta_0}^{\otimes 3}) + \lambda_2(\beta_1^{\otimes 3} + \overline{\beta_1}^{\otimes 3})$.*

Proof. The proof is by a holographic transformation. We omit the details.

Theorem 4. *Let f be a $\{0, 1\}$-valued symmetric ternary function defined on a domain of size 4. If f is not among the P-time computable cases in Theorem 3, then the problem $\mathrm{Holant}^*(f)$ is #P-hard. Moreover, for $\{0, 1\}$-valued f, only cases 1 and 2 in Theorem 3 occur.*

We remark that for $\{-1, 1\}$-valued symmetric ternary functions, the third tractable case of Theorem 3 does occur. The following is an example:

$$Q = \begin{bmatrix} \frac{1+\sqrt{2}}{2\sqrt{\sqrt{2}+2}} & -\frac{1}{2\sqrt{\sqrt{2}+2}} & \frac{-\sqrt{2}-1}{2\sqrt{\sqrt{2}+2}} & \frac{1}{2\sqrt{\sqrt{2}+2}} \\ \frac{-1+\sqrt{2}}{2\sqrt{2-\sqrt{2}}} & \frac{1}{2\sqrt{2-\sqrt{2}}} & \frac{-1+\sqrt{2}}{2\sqrt{2-\sqrt{2}}} & \frac{1}{2\sqrt{2-\sqrt{2}}} \\ \frac{-3+2\sqrt{2}}{2\sqrt{10-7\sqrt{2}}} & \frac{1-\sqrt{2}}{2\sqrt{10-7\sqrt{2}}} & \frac{3-2\sqrt{2}}{2\sqrt{10-7\sqrt{2}}} & \frac{-1+\sqrt{2}}{2\sqrt{10-7\sqrt{2}}} \\ \frac{-3-2\sqrt{2}}{2\sqrt{7\sqrt{2}+10}} & \frac{1+\sqrt{2}}{2\sqrt{7\sqrt{2}+10}} & \frac{-3-2\sqrt{2}}{2\sqrt{7\sqrt{2}+10}} & \frac{1+\sqrt{2}}{2\sqrt{7\sqrt{2}+10}} \end{bmatrix}$$

On the left is a signature g. On the right is an orthogonal matrix Q. In fact, under the transformation, $Qg = \sqrt{2 - \sqrt{2}}(\beta_0^{\otimes 3} + \overline{\beta_0}^{\otimes 3}) - \sqrt{2 + \sqrt{2}}(\beta_1^{\otimes 3} + \overline{\beta_1}^{\otimes 3})$, which is one example of the third tractable case of Theorem 3.

There are only a finite (albeit a large) number of $\{0, 1\}$-valued symmetric ternary signatures on domain 4. We will prove Theorem 4 by going through all signatures using five general strategies. When one signature could not be identified as #P-hard by any of the five strategies in section 3.1, it is shown that it actually satisfies the first or second tractable conditions in Theorem 3.

3.1 Strategies

There are five different strategies we use to identify #P-hard signatures.

1. Use gadgets to form a binary symmetric signature which when written as a matrix M has rank 2. Then apply an orthogonal holographic transformation T, which transforms M to the form $\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & \lambda & 0 \\ 0 & 0 & 0 & \mu \end{bmatrix}$, $\lambda\mu \neq 0$. We then interpolate $\mathrm{diag}(0, 0, 1, 1)$, a Boolean equality on the last two (new) domain elements by Lemma 6. Finally apply the Boolean domain dichotomy.

2. Similarly, we form a binary symmetric signature which when written as a matrix has rank 3. Then apply an orthogonal transformation T and get $\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & \lambda_1 & 0 & 0 \\ 0 & 0 & \lambda_2 & 0 \\ 0 & 0 & 0 & \lambda_3 \end{bmatrix}$, $\lambda_1 \lambda_2 \lambda_3 \neq 0$, and interpolate $\mathrm{diag}(0,1,1,1)$ by Lemma 5. Finally, apply Theorem 2 to Tf on the last three (new) domain elements.

3. Find a nonzero unary signature $u \in \mathbb{R}^4$, such that $\langle f, u \rangle = 0$. Then define an orthogonal matrix T with (normalized) u as the first row. T transforms f to a signature supported on a lower domain (all 0's except the bottom face of the signature tetrahedron). Then apply the corresponding dichotomy.

4. Find some nonzero unary signature $u \in \mathbb{R}^4$, and nonzero $c \in \mathbb{R}$, such that $\langle f, u \rangle = cu \cdot u^T$. Define an orthogonal T using (normalized) u to be the first row. T will transform f to be domain separated (where the first new domain element R' is separated from the rest $\{G', B', W'\}$, i.e., Tf evaluates to 0, when R' is among its input, *except* possibly on (R', R', R')). Then apply the domain 3 dichotomy Theorem 2.

5. Use gadgets to construct a symmetric binary signature M which when written as a matrix has rank 4, and its 4 eigenvalues $\lambda_1, \lambda_2, \lambda_3, \lambda_4$ satisfy some condition under which we can interpolate (by Lemma 7 and Lemma 8) either $\mathrm{diag}(0,0,1,1)$ or $\mathrm{diag}(0,1,1,1)$ from $\mathrm{diag}(\lambda_1, \lambda_2, \lambda_3, \lambda_4)$. Form the orthogonal matrix Q such that $QMQ^T = \mathrm{diag}(\lambda_1, \lambda_2, \lambda_3, \lambda_4)$. Finally apply the corresponding lower domain dichotomy to the corresponding part of Qf.

Below we show several examples using some of the strategies above. Consider the tetrahedron on the left. We use strategy 5:

$$\begin{array}{cccc} & & -\dfrac{41}{24\sqrt{3}} & \\ & \dfrac{1}{4\sqrt{2}} & -\dfrac{1}{12\sqrt{6}} & \\ \dfrac{2}{\sqrt{3}} & 0 & -\dfrac{5}{3\sqrt{3}} & \\ -\dfrac{2}{\sqrt{2}} & \dfrac{\sqrt{2}}{\sqrt{3}} & \dfrac{1}{\sqrt{2}} & -\dfrac{\sqrt{2}}{3\sqrt{3}} \end{array}$$

We call this signature g, and use a unary $e_4 = (0,0,0,1)^T$ to connect to g to produce a binary symmetric function $M = \begin{bmatrix} 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 \end{bmatrix}$ whose eigenvalues are $\{3, 1, -1, -1\}$. We then use an orthogonal matrix $Q = \dfrac{1}{2\sqrt{3}} \begin{bmatrix} \sqrt{3} & \sqrt{3} & \sqrt{3} & \sqrt{3} \\ -1 & -1 & -1 & 3 \\ -\sqrt{6} & 0 & \sqrt{6} & 0 \\ \sqrt{2} & -2\sqrt{2} & \sqrt{2} & 0 \end{bmatrix}$ to diagonalize M, i.e., $QMQ^T = \mathrm{diag}(3, -1, 1, -1)$. It follows that we are able to interpolate $\mathrm{diag}(0,1,1,1)$ (by Lemma 7). We build a new signature grid where we add a binary vertex on each edge between two Qg's, and assign $\mathrm{diag}(0,1,1,1)$ on all the new degree 2 vertices. This restricts all edges in the new signature grid to be assigned a color only from $\{G', B', W'\}$ (no R') in order the evaluation of any product term in the partition function to be nonzero. It follows that we have a problem on domain size 3, which is defined by the ternary signature on

domain 3 shown on the right as a "triangle". We can apply Theorem 2, and find that it is #P-hard. Therefore, the problem Holant$^*(g)$ is #P-hard.

Next we show another example using strategy number 3 above.

Let g be the signature on the left. Connecting a unary $u = (0, 1, -1, 0)^T$ to it, we get a zero binary function. We construct an orthogonal matrix Q with the normalized u as its first row, $Q = \begin{bmatrix} 0 & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 \\ 1 & 0 & 0 & 0 \\ 0 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$. Then the transformed signature Qg is shown as the tetrahedron on the right, which contains only nonzero elements on its bottom face. This signature is effectively a ternary signature on domain size 3, and the original problem is equivalent to this problem on domain size 3. Then we apply the domain 3 dichotomy. By Theorem 2, the problem is #P-hard.

Consider another example:

On the left is g, a ternary signature on domain size 4. There is a unary $u = (1, -1, 0, 0)^T$ such that $\langle u, g \rangle = 0$. Hence, we construct an orthogonal $Q = \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 & 0 \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$ (where its first row is the normalized u), such that under the transformation, g becomes Qg (on the right) and $Qg = \alpha^{\otimes 3} + \beta^{\otimes 3}$ where $\alpha = (0, -2^{\frac{1}{6}}, 2^{-\frac{1}{3}}, 2^{-\frac{1}{3}})^T$ and $\beta = (0, 2^{\frac{1}{6}}, 2^{-\frac{1}{3}}, 2^{-\frac{1}{3}})^T$. Here, after we use strategy number 3 to transform the problem to one on domain 3, we find it is tractable.

For all $\{0, 1\}$-valued ternary signatures on domain size 4, we went through them using one of the above 5 strategies. It is found that we either can prove it is #P-hard, or when it fails to do so, it is tractable by being in one of the tractable forms in Theorem 4.

Tricks in applying the domain 3 dichotomy To apply the domain 3 dichotomy, there are also several tricks that can help simplify the calculation.

1. When checking whether a real ternary domain 3 signature satisfies the first tractable form $f = \alpha^{\otimes 3} + \beta^{\otimes 3} + \gamma^{\otimes 3}$, we can use different unary signatures u to connect to it and get different binary functions. Written as a matrix form $M = \langle f, u \rangle$, it's obviously symmetric. Write its eigen-decomposition as $M = Q\Lambda Q^T$, where Q is an orthogonal matrix $Q = [q_1, q_2, q_3]$ where q_i $(i = 1, 2, 3)$ are column vectors, and $\Lambda = \mathrm{diag}(\lambda_1, \lambda_2, \lambda_3)$. If $\lambda_1, \lambda_2, \lambda_3$ are all distinct (at most one of them can be 0), and if f falls into the first tractable case, then the set $\{\pm q_1, \pm q_2, \pm q_3\}$ is independent of u.

   Proof. If f satisfies the first tractable form $f = \alpha^{\otimes 3} + \beta^{\otimes 3} + \gamma^{\otimes 3}$, then $M = \langle u, f \rangle = \langle u, \alpha \rangle \alpha \alpha^T + \langle u, \beta \rangle \beta \beta^T + \langle u, \gamma \rangle \gamma \gamma^T$. Then $M\alpha = \langle u, \alpha \rangle \langle \alpha, \alpha \rangle \alpha$, so α is an eigenvector of M (if nonzero). Similarly, nonzero β, γ are both eigenvectors of M. As rank $M \geq 2$, at most one of α, β, γ can be 0. So at least two of them are scalar multiples of q_i $(i = 1, 2, 3)$. So, at least two of $\pm q_i$ $(i = 1, 2, 3)$ do not depend on the choice of u, and the third is uniquely determined by those two up to a $\pm$ factor.

2. When checking whether a real ternary domain 3 signature is in the first tractability case, we can search for whether there exists a unary u such that $\langle u, f \rangle = \langle u, u \rangle uu^T$. If a nonzero f does not have such a nonzero unary, it cannot be in the first tractable case since α, β, γ satisfy this relation.

3. To check whether a real ternary domain 3 signature is in the second tractability case, assume it has the form $f = \alpha^{\otimes 3} + \beta^{\otimes 3} + \overline{\beta}^{\otimes 3}$. If $\beta = 0$, then it also falls into the first tractability category. Now we assume $\beta \neq 0$. Then we have a nonzero β such that $\langle \beta, f \rangle = \langle \beta, \overline{\beta} \rangle \overline{\beta}\ \overline{\beta}^T$ and $\langle \beta, \beta \rangle = 0$. So if there does not exist such a nonzero isotropic β, it is not in the second tractable case.

3.2   Interpolate restricted equalities

Suppose for some unary u, the 4-by-4 matrix $A = \langle u, f \rangle$ has rank 3. Because A is a real symmetric matrix, we can construct an orthogonal matrix T so that

$$TAT^T = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & \lambda_1 & 0 & 0 \\ 0 & 0 & \lambda_2 & 0 \\ 0 & 0 & 0 & \lambda_3 \end{bmatrix} \tag{1}$$

where $\lambda_1 \lambda_2 \lambda_3 \neq 0$. We denote by $=_{G,B,W}$ the binary function in (1) if all $\lambda_i = 1$.

Lemma 5. *Let $H : \{R, G, B, W\}^2 \to \mathbb{R}$ be a rank 3 binary function of the form (1). Then for any $\mathcal{F}$ containing H, we have*

$$\mathrm{Holant}(\mathcal{F} \cup \{=_{G,B,W}\}) \leq_T \mathrm{Holant}(\mathcal{F}).$$

Similarly, we denote by $=_{B,W}$ the binary function in (1) if $\lambda_1 = 0$ and $\lambda_2 = \lambda_3 = 1$. We have

Lemma 6. *Let $H : \{R, G, B, W\}^2 \to \mathbb{R}$ be a rank 2 binary function of the form*
$\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & \lambda & 0 \\ 0 & 0 & 0 & \mu \end{bmatrix}$ *where $\lambda\mu \neq 0$. Then for any signature set $\mathcal{F}$ containing H, we have*

$$\mathrm{Holant}(\mathcal{F} \cup \{=_{B,W}\}) \leq_T \mathrm{Holant}(\mathcal{F}).$$

Lemma 5 and 6 enable us to construct instances on a lower domain which can help establish #P-hardness. In the next two lemmas we interpolate a lower domain equality directly from a rank 4 real symmetric matrix.

Lemma 7. *Let $H : \{R, G, B, W\}^2 \to \mathbb{R}$ be a rank 4 symmetric binary function. Let Q be the orthogonal matrix such that $QHQ^T = \begin{bmatrix} \lambda_1 & 0 & 0 & 0 \\ 0 & \lambda_2 & 0 & 0 \\ 0 & 0 & \lambda_3 & 0 \\ 0 & 0 & 0 & \lambda_4 \end{bmatrix}$ where $\lambda_1\lambda_2\lambda_3\lambda_4 \neq 0$. If the four eigenvalues $\lambda_1, \lambda_2, \lambda_3, \lambda_4$ further satisfy the condition: for all $s, a, b, c \in \mathbb{Z}$, $s > 0$, if $a + b + c = s$ then $\lambda_1^s \neq \lambda_2^a \lambda_3^b \lambda_4^c$. Then for any $\mathcal{F}$ containing H, we have*

$$\mathrm{Holant}(Q\mathcal{F} \cup \{=_{G,B,W}\}) \leq_T \mathrm{Holant}(\mathcal{F}).$$

Similarly, we have Lemma 8.

Lemma 8. *Let $H : \{R, G, B, W\}^2 \to \mathbb{R}$ be a rank 4 symmetric binary function. Let Q be the orthogonal matrix such that $QHQ^T = \begin{bmatrix} \lambda_1 & 0 & 0 & 0 \\ 0 & \lambda_2 & 0 & 0 \\ 0 & 0 & \lambda_3 & 0 \\ 0 & 0 & 0 & \lambda_4 \end{bmatrix}$ where $\lambda_1\lambda_2\lambda_3\lambda_4 \neq 0$. If the four eigenvalues $\lambda_1, \lambda_2, \lambda_3, \lambda_4$ further satisfy the condition: for all $s, t, a, b \in \mathbb{Z}$, $s, t \geq 0$, $s + t > 0$, if $a + b = s + t$ then $\lambda_1^s \lambda_2^t \neq \lambda_3^a \lambda_4^b$. Then for any $\mathcal{F}$ containing H, we have*

$$\mathrm{Holant}(Q\mathcal{F} \cup \{=_{B,W}\}) \leq_T \mathrm{Holant}(\mathcal{F}).$$

Let $\mathcal{U}$ be the set of all complex unaries and f be a possibly complex signature, both over domain 3. Then by definition, $\mathrm{Holant}^*(f) = \mathrm{Holant}(f \cup \mathcal{U})$. When f is a real-valued signature, we temporarily define $\mathrm{Holant}^{r*}(f)$ as the Holant problem where only all the **real** unaries are available. In Lemma 9, we prove that in fact the complexity of the problem $\mathrm{Holant}^{r*}(f)$ is the same as $\mathrm{Holant}^*(f)$. (Hence, after Lemma 9, this new notation $\mathrm{Holant}^{r*}(f)$ will be seen as unimportant, as far as the complexity of the problem is concerned.) The same lemma also holds for any domain k; the proof of Lemma 9 is easily adapted.

Lemma 9. *For any real signature f over domain 3, we have*

$$\mathrm{Holant}^*(f) \leq_T \mathrm{Holant}^{r*}(f).$$

Proof. By Theorem 3.1 in [11], the problem $\mathrm{Holant}^*(f)$ is either polynomial-time solvable, or it is #P-hard. If it is polynomial-time solvable, then the reduction trivially holds. Otherwise, it is #P-hard. This means that there exist finitely many unaries $u_1, u_2, \ldots, u_k$, possibly complex-valued, such that the

problem $\text{Holant}(\{f, u_1, \ldots, u_k\})$ is #P-hard. We prove that $\text{Holant}(\{f, u\}) \leq_T \text{Holant}^{r*}(f)$, for any complex-valued u, and then use induction, as k is a constant.

Now consider an arbitrary instance I of $\text{Holant}(\{f, u\})$ and suppose the unary function u appears m times in I. We write $u = [x, y, z]$ where $x, y, z \in \mathbb{C}$. We may assume $u \neq 0$, and by symmetry we may assume $z \neq 0$. We stratify the edge assignments according to the number of Red, Green and Blue assigned to the u's. Specifically, let ρ_{ij} denote the sum of products of the evaluations of all f's, where the sum is over all assignments with exactly i many times Red, j many times Green, and $(m - i - j)$ many times Blue are assigned to the u's. Then the Holant value on the instance I can be written as $\sum_{i=0}^{m} \sum_{j=0}^{m-i} \rho_{ij} x^i y^j z^{m-i-j} = z^m \sum_{i=0}^{m} \sum_{j=0}^{m-i} \rho_{ij} (\frac{x}{z})^i (\frac{y}{z})^j$. Observe that once we know all the values of ρ_{ij}, we can compute the Holant value of I in P-time. We now interpolate the values of ρ_{ij}.

We construct the instances I_k, $1 \leq k \leq \binom{m+2}{2}$, for the problem $\text{Holant}^{r*}(f)$. For any such k, let I_k be the same signature grid as I except we replace every appearance of u by the real unary $[3^k, 2^k, 1]$. Then, the Holant value for I_k equals to $\sum_{i=0}^{m} \sum_{j=0}^{m-i} \rho_{ij} 3^{ki} 2^{kj}$. Therefore, we can write a non-degenerate Vandermonde system, where the matrix has entries $3^{ki} 2^{kj}$ with the columns indexed by lexicographical order of the tuples (i, j) and the rows indexed by k, and the unknown variables are ρ_{ij}. We can then solve for ρ_{ij} and therefore compute the Holant value of I in polynomial time. This finishes the proof of the lemma.

References

1. Backens, M.: A full dichotomy for Holant^c, inspired by quantum computation. SIAM Journal on Computing **50**(6), 1739–1799 (2021), https://doi.org/10.1137/20M1311557

2. Backens, M., Goldberg, L.A.: Holant clones and the approximability of conservative Holant problems. ACM Trans. Algorithms **16**(2), 23:1–23:55 (2020), https://doi.org/10.1145/3381425

3. Bulatov, A.A.: A dichotomy theorem for constraint satisfaction problems on a 3-element set. Journal of the ACM (JACM) **53**(1), 66–120 (2006), https://doi.org/10.1145/1120582.1120584

4. Cai, J.Y., Chen, X.: Complexity of counting CSP with complex weights. JACM **64**(3), 19:1–19:39 (2017), https://doi.org/10.1145/2822891

5. Cai, J.Y., Chen, X., Lu, P.: Nonnegative weighted# CSP: An effective complexity dichotomy. SIAM Journal on Computing **45**(6), 2177–2198 (2016), https://doi.org/10.1137/15M1032314

6. Cai, J., Guo, H., Williams, T.: The complexity of counting edge colorings and a dichotomy for some higher domain holant problems. In: 55th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2014, Philadelphia, PA, USA, October 18-21, 2014. pp. 601–610. IEEE Computer Society (2014), https://doi.org/10.1109/FOCS.2014.70

7. Cai, J.Y., Guo, H., Williams, T.: A complete dichotomy rises from the capture of vanishing signatures. SIAM J. Comput. **45**(5), 1671–1728 (2016), https://doi.org/10.1137/15M1049798

8. Cai, J.Y., Lu, P.: Holographic algorithms: From art to science. J. Comput. Syst. Sci. **77**(1), 41–61 (2011), https://doi.org/10.1016/j.jcss.2010.06.005

9. Cai, J.Y., Lu, P., Xia, M.: Holographic algorithms by Fibonacci gates and Holographic reductions for hardness. In: 2008 49th Annual IEEE Symposium on Foundations of Computer Science. pp. 644–653 (2008). https://doi.org/10.1109/FOCS.2008.34

10. Cai, J.Y., Lu, P., Xia, M.: Holant problems and counting CSP. In: Proceedings of the forty-first annual ACM symposium on Theory of computing. pp. 715–724 (2009), https://doi.org/10.1145/1536414.1536511

11. Cai, J.Y., Lu, P., Xia, M.: Dichotomy for Holant* problems with domain size 3. In: Proceedings of the Twenty-Fourth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2013, New Orleans, Louisiana, USA, January 6-8, 2013. pp. 1278–1295. SIAM (2013), https://doi.org/10.1137/1.9781611973105.93

12. Cai, J.Y., Lu, P., Xia, M.: Holographic algorithms with matchgates capture precisely tractable planar #CSP. SIAM Journal on Computing **46**(3), 853–889 (2017), https://doi.org/10.1137/16M1073984

13. Dyer, M., Richerby, D.: An effective dichotomy for the counting constraint satisfaction problem. SIAM Journal on Computing **42**(3), 1245–1274 (2013), https://doi.org/10.1137/100811258

14. Freedman, M., Lovász, L., Schrijver, A.: Reflection positivity, rank connectivity, and homomorphism of graphs. J. Amer. Math. Soc. **20**(1), 37–51 (2007)

15. Fu, Z., Yang, F., Yin, M.: On blockwise symmetric matchgate signatures and higher domain #csp. Inf. Comput. **264**, 1–11 (2019), https://doi.org/10.1016/j.ic.2018.09.012

16. Guo, H., Huang, S., Lu, P., Xia, M.: The complexity of weighted Boolean #CSP modulo k. In: 28th International Symposium on Theoretical Aspects of Computer Science, STACS. pp. 249–260 (2011), https://doi.org/10.4230/LIPIcs.STACS.2011.249

17. Guo, H., Lu, P., Valiant, L.G.: The complexity of symmetric Boolean parity Holant problems. SIAM J. Comput. **42**(1), 324–356 (2013), https://doi.org/10.1137/100815530

18. Hell, P., Nešetřil, J.: Graphs and homomorphisms, Oxford lecture series in mathematics and its applications, vol. 28. Oxford University Press (2004)

19. Kowalczyk, M., Cai, J.Y.: Holant problems for 3-regular graphs with complex edge functions. Theory Comput. Syst. **59**(1), 133–158 (2016), https://doi.org/10.1007/s00224-016-9671-7

20. Szegedy, B.: Edge coloring models and reflection positivity. Journal of the American mathematical Society **20**(4), 969–988 (2007)

21. Szegedy, B.: Edge coloring models as singular vertex coloring models. Bolyai Society Mathematical Studies. Springer Berlin Heidelberg pp. 327–336 (2010)

22. Valiant, L.G.: Accidental algorithms. In: 47th Annual IEEE Symposium on Foundations of Computer Science, FOCS. pp. 509–517. IEEE (2006), https://doi.org/10.1109/FOCS.2006.7

23. Valiant, L.G.: Holographic algorithms. SIAM Journal on Computing **37**(5), 1565–1594 (2008), https://doi.org/10.1137/070682575

24. Xia, M.: Holographic reduction: A domain changed application and its partial converse theorems. Int. J. Softw. Informatics **5**(4), 567–577 (2011), http://www.ijsi.org/ch/reader/view_abstract.aspx?file_no=i109

Appendix

We start with the proof of Lemma 5.

Proof. Fix an instance I of $\mathrm{Holant}(\mathcal{F}\cup\{=_{G,B,W}\})$. Suppose the signature $=_{G,B,W}$ appears m times in I. To have a nonzero contribution to the Holant sum, the assignments given to any occurrence of the $=_{G,B,W}$ signatures must be (G,G), (B,B) or (W,W). We can stratify the edge assignments in the Holant sum of I according to the number of (G,G), (B,B) and (W,W) assigned to the occurrences of $=_{G,B,W}$. Specifically, let ρ_{ij} denote the sum of products of the evaluations of all other signatures, where the sum is over all assignments with exactly i many times (G,G), j many times (B,B) and $(m-i-j)$ many times (W,W) assigned to those $=_{G,B,W}$ signatures. Then the Holant value of the instance I can be written as $\sum_{i=0}^{m}\sum_{j=0}^{m-i}\rho_{ij}$.

Let $K=\binom{m+2}{2}$. Now we construct from I a sequence of instances I_k of $\mathrm{Holant}(\mathcal{F})$ indexed by k, $1\le k\le K$, by replacing each occurrence of $(=_{G,B,W})$ in I with a chain of k copies of the function H. The signature of the k-chain is

$H^k = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & \lambda_1^k & 0 & 0 \\ 0 & 0 & \lambda_2^k & 0 \\ 0 & 0 & 0 & \lambda_3^k \end{bmatrix}$. This effectively replaces each occurrence of $(=_{G,B,W})$ in I

by H^k. Let x_{ij} be $(\frac{\lambda_1}{\lambda_3})^i(\frac{\lambda_2}{\lambda_3})^j$. A moment of reflection shows that the value of the instance I_k is

$$\sum_{i=0}^{m}\sum_{j=0}^{m-i}\rho_{ij}\lambda_1^{ki}\lambda_2^{kj}\lambda_3^{k(m-i-j)} = \lambda_3^{mk}\sum_{i=0}^{m}\sum_{j=0}^{m-i}\rho_{ij}x_{ij}^k.$$

Then we can view the above as a linear system where the coefficients are x_{ij}^k and the variables are ρ_{ij}. Two columns indexed by $(i,j)\ne(i',j')$ are identical

in the matrix $\begin{bmatrix} 1 & \cdots & x_{ij} & \cdots & x_{i'j'} & \cdots \\ 1 & \cdots & x_{ij}^2 & \cdots & x_{i'j'}^2 & \cdots \\ \vdots & \ddots & \vdots & \ddots & \vdots & \ddots \\ 1 & \cdots & x_{ij}^K & \cdots & x_{i'j'}^K & \cdots \end{bmatrix}$, iff $x_{ij}=x_{i'j'}$. We can combine these two

columns and create a new variable representing the sum of ρ_{ij} and $\rho_{i'j'}$. Repeating in this way until no two columns are the same, we will arrive at a matrix of $K'\le K$ columns such that the first K' rows form a square Vandermonde matrix with full rank. We can then solve the Vandermonde system and take the sum of all the variables. This is the sum of the original ρ_{ij}'s, which is exactly the Holant value on I'. The proof is now complete.

The proof of Lemma 6 can be easily adapted from the proof of Lemma 5. We now continue with the proof of Lemma 7.

Proof. We have
$$\mathrm{Holant}(Q\mathcal{F}) \equiv_T \mathrm{Holant}(\mathcal{F}),$$

and the binary function $QHQ^T = \begin{bmatrix} \lambda_1 & 0 & 0 & 0 \\ 0 & \lambda_2 & 0 & 0 \\ 0 & 0 & \lambda_3 & 0 \\ 0 & 0 & 0 & \lambda_4 \end{bmatrix} \in Q\mathcal{F}$ because $H \in \mathcal{F}$.

Consider an instance I of $\mathrm{Holant}(Q\mathcal{F} \cup \{=_{G,B,W}\})$. Suppose the function $(=_{G,B,W}) = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$ appears m times in I. Similar to the proof of Lemma 5, we can stratify the edge assignments in the Holant sum according to the number of (R, R), (G, G), (B, B) and (W, W) assigned to the occurrences of $(=_{G,B,W})$. Specifically, let $\rho_{\ell ij}$ denote the sum of products of the evaluations of all other signatures, where the sum is over all assignments with exactly ℓ many times (R, R), i many times (G, G), j many times (B, B) and $(m - \ell - i - j)$ many times (W, W) assigned to those $(=_{G,B,W})$ signatures. Then, the Holant value on the instance I can be written as $\sum_{i=0}^{m} \sum_{j=0}^{m-i} \rho_{0ij}$, since any edge assignment that makes nonzero contribution in the Holant sum must assign 0 time (R, R) to $(=_{G,B,W})$.

Now we construct from I a sequence of instances I_k for $\mathrm{Holant}(Q\mathcal{F})$ indexed by k, $1 \le k \le \binom{m+3}{3}$, by replacing each occurrence of $(=_{G,B,W})$ with a chain of k copies of the function $\begin{bmatrix} \lambda_1 & 0 & 0 & 0 \\ 0 & \lambda_2 & 0 & 0 \\ 0 & 0 & \lambda_3 & 0 \\ 0 & 0 & 0 & \lambda_4 \end{bmatrix}$. That is, each occurrence of $\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$ is replaced by $\begin{bmatrix} \lambda_1^k & 0 & 0 & 0 \\ 0 & \lambda_2^k & 0 & 0 \\ 0 & 0 & \lambda_3^k & 0 \\ 0 & 0 & 0 & \lambda_4^k \end{bmatrix}$. Let $x_{\ell ij}$ be $(\frac{\lambda_1}{\lambda_4})^{\ell}(\frac{\lambda_2}{\lambda_4})^{i}(\frac{\lambda_3}{\lambda_4})^{j}$. A moment of reflection shows that the Holant value of the instance I_k is

$$\sum_{\ell=0}^{m} \sum_{i=0}^{m-\ell} \sum_{j=0}^{m-\ell-i} \rho_{\ell ij} \lambda_1^{\ell k} \lambda_2^{ik} \lambda_3^{jk} \lambda_4^{(m-\ell-i-j)k} = \lambda_4^{mk} \sum_{\ell=0}^{m} \sum_{i=0}^{m-\ell} \sum_{j=0}^{m-\ell-i} \rho_{\ell ij} x_{\ell ij}^{k}.$$

Then we can view the above as a linear system where the coefficient matrix has entries $x_{\ell ij}^{k}$ with the columns indexed by lexicographical order of the tuples (ℓ, i, j) and the rows indexed by $k \in \mathbb{Z}_+$, and the unknown variables are $\rho_{\ell ij}$. We will take $k = 1, \dots, K$, where $K = \binom{m+3}{3}$. We combine two columns indexed by $(\ell, i, j) \neq (\ell', i', j')$ if the two columns are equal. This happens iff $\lambda_1^{\ell} \lambda_2^{i} \lambda_3^{j} \lambda_4^{m-\ell-i-j} = \lambda_1^{\ell'} \lambda_2^{i'} \lambda_3^{j'} \lambda_4^{m-\ell'-i'-j'}$. In that case we create a new variable ρ' to be the sum of the two variables $\rho_{\ell ij}$ and $\rho_{\ell'i'j'}$. We keep combining the columns in this way until no two columns are the same, at which point we arrive at a Vandermonde matrix with full rank. Notice that what we want is $\mathrm{Holant}(I) = \sum_{i=0}^{m} \sum_{j=0}^{m-i} \rho_{0ij}$. Therefore, as long as $(0, i, j)$ and (ℓ', i', j') for $\ell' > 0$ are never combined, the sum of ρ_{0ij}'s and the sum of $\rho_{\ell'i'j'}$'s are not "mixed". This condition is expressed as: For all $i, j, i', j' \ge 0$, and $\ell' > 0$, $\lambda_2^{i} \lambda_3^{j} \lambda_4^{m-i-j} \neq \lambda_1^{\ell'} \lambda_2^{i'} \lambda_3^{j'} \lambda_4^{m-\ell'-i'-j'}$. This is guaranteed by the assumption of the lemma.

Therefore, we can solve the Vandermonde system, summing up the ones coming from the form ρ_{0ij} where $i + j \le m$, to get $\mathrm{Holant}(I)$. Therefore, $\mathrm{Holant}(Q\mathcal{F} \cup \{=_{G,B,W}\}) \le_T \mathrm{Holant}(\mathcal{F})$ and the lemma is proved.

The proof of Lemma 8 can be easily adapted from the proof of Lemma 7. We will define $\rho_{\ell i j}$'s similarly. We want to compute the Holant value $\sum_{j=0}^{m} \rho_{00j}$. The sum of ρ_{00j}'s and the sum of $\rho_{\ell' i' j'}$'s are not "mixed" as long as for all $\ell', i' \geq 0$, with $\ell' + i' > 0$, and all $j, j' \geq 0$, we have $\lambda_3^j \lambda_4^{m-j} \neq \lambda_1^{\ell'} \lambda_2^{i'} \lambda_3^{j'} \lambda_4^{m-\ell'-i'-j'}$, which is guaranteed by the assumption of Lemma 8.