

## COMPUTATIONAL COMPLEXITY OF HOLANT PROBLEMS\*

JIN-YI CAI<sup>†</sup>, PINYAN LU<sup>‡</sup>, AND MINGJI XIA<sup>§</sup>

**Abstract.** We propose and explore a novel alternative framework to study the complexity of counting problems, called Holant problems. Compared to counting constraint satisfaction problems (#CSP), it is a refinement with a more explicit role for the constraint functions. Both graph homomorphism and #CSP can be viewed as special cases of Holant problems. We prove complexity dichotomy theorems in this framework. Our dichotomy theorems apply to local constraint functions, which are symmetric functions on Boolean input variables and evaluate to arbitrary real or complex values. We discover surprising tractable subclasses of counting problems, which could not easily be specified in the #CSP framework. When all unary functions are assumed to be free (Holant\* problems), the tractable ones consist of functions that are degenerate, or of arity at most two, or holographic transformations of Fibonacci gates. When only two special unary functions, the constant zero and constant one functions, are assumed to be free (Holant<sup>c</sup> problems), we further identify three special families of tractable cases. Then we prove that all other cases are #P-hard. The main technical tool we use and develop is holographic reductions. Another technical tool used in combination with holographic reductions is polynomial interpolations.

**Key words.** Holant problems, holographic reduction, polynomial interpolation, #P-hardness

**AMS subject classifications.** 68Q17, 68Q25

**DOI.** 10.1137/100814585

**1. Introduction.** In order to study the complexity of counting problems, several interesting frameworks have been proposed. One is called counting constraint satisfaction problems (#CSP) [4, 2, 18, 3]. Another well-studied framework is called graph homomorphisms or  $H$ -coloring problems, which can be viewed as a special case of #CSP problems [5, 6, 21, 22, 19, 20, 24, 7]. One reason such frameworks are interesting is because the language is *expressive* enough so that they can express many natural counting problems, while *specific* enough so that it is possible to prove complete classification theorems on their complexity [15]. Natural counting problems that can be expressed as graph homomorphism problems include counting the number of vertex covers, the number of  $k$ -colorings in a graph, and many others. However, there are some natural and important counting problems, which cannot be expressed as a graph homomorphism problem. In [23], it is proved that counting the number of perfect matchings in a graph *cannot* be expressed as a graph homomorphism function. Additionally, sometimes a problem can be expressed in the existing framework, such as #CSP, but only with some contrived restrictions.

---

\*Received by the editors November 11, 2010; accepted for publication (in revised form) May 5, 2011; published electronically July 19, 2011. Some results in this paper have been reported in “Holant problems and counting CSP,” in *Proceedings of STOC 2009 the 41st Annual ACM Symposium on Theory of Computing*, 2009.

<http://www.siam.org/journals/sicomp/40-4/81458.html>

<sup>†</sup>Computer Sciences Department, University of Wisconsin, Madison, WI 53706, and Beijing University, Beijing, People’s Republic of China (jyc@cs.wisc.edu). The first author’s work was supported by NSF CCF-0830488, CCF-0914969, and CCF-0511679.

<sup>‡</sup>Microsoft Research Asia, Beijing, People’s Republic of China (pinyanl@microsoft.com).

<sup>§</sup>State Key Laboratory of Computer Science, Institute of Software, Chinese Academy of Sciences, Beijing, People’s Republic of China (mingji@ios.ac.cn). The third author’s work was supported by Hundred Talent Program of Chinese Academy of Sciences Under Angsheng Li, and the Grand Challenge Program “Network Algorithms and Digital Information” of ISCAS, and by NSFC 61003030 and 60970003.

In this paper, we propose and explore an alternative framework to study the complexity of counting problems, called Holant problems. This notion is motivated by holographic reductions proposed by Valiant [34, 33]. Compared to #CSP, it is a generalization with a more explicit role for the constraint functions, without assuming implicitly that EQUALITY functions of all arities are freely available. Both graph homomorphism and #CSP can be viewed as special cases of Holant problems. We give a brief description here, and a more formal definition is given in section 2. A *signature grid*  $\Omega = (G, \mathcal{F}, \pi)$  is a tuple, where  $G = (V, E)$  is an undirected graph,  $\mathcal{F}$  is a set of functions, and  $\pi$  maps each  $v \in V(G)$  to a function  $f_v \in \mathcal{F}$ , with input variables associated with the incident edges at  $v$ . Edges are treated as variables that take values from a finite domain set  $[q]$ . An assignment  $\sigma$  gives each edge  $e \in E$  a value from  $[q]$  and produces an evaluation  $\prod_{v \in V} f_v(\sigma|_{E(v)})$ , where  $E(v)$  denotes the incident edges of  $v$ , and  $f_v$  is evaluated on the restriction of  $\sigma$  on  $E(v)$ . The counting problem on an input instance  $\Omega$  is to compute

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

For example, consider the PERFECT MATCHING problem on  $G$ . This problem corresponds to attaching the EXACT-ONE function at every vertex of  $G$ . Consider all 0-1 edge assignments  $\sigma$ . The product  $\prod_{v \in V} f_v(\sigma|_{E(v)})$  evaluates to 0 or 1, and is 1 iff  $\sigma^{-1}(1) \subseteq E$  is a perfect matching. Hence in this case,  $\text{Holant}_{\Omega}$  counts the number of perfect matchings. If we use the AT-MOST-ONE function at every vertex, then we are counting all (not necessarily perfect) matchings. So this new framework can express some natural counting problems that are not expressible as graph homomorphisms.

To see that Holant is a more expressive framework, we show that every #CSP problem can be simulated by a Holant problem. Represent an instance of a #CSP problem by a bipartite graph where the left-hand sides (LHS) are labeled by variables and the right-hand sides (RHS) are labeled by constraints. Now the signature grid  $\Omega$  on this bipartite graph is as follows: Every variable node on the LHS is attached to an EQUALITY function, and every constraint node on the RHS has the given constraint function. Then  $\text{Holant}_{\Omega}$  is exactly the answer to the #CSP problem. In effect, the EQUALITY function on each variable node forces the incident edges to take the same value; this effectively reduces edge assignments to vertex assignments assigning values to each variable on the LHS as in #CSP. It follows that #CSP problems are precisely the special case of Holant problems on bipartite graphs where every node on LHS is attached to an EQUALITY function. It is easy to show that the class of #CSP problems is equivalent to Holant problems where all EQUALITY functions (of arbitrary arities) are always assumed to be freely available, and implicitly so. Graph homomorphism is a further special case where not only all EQUALITY functions are freely (and implicitly) available, but also the function set  $\mathcal{F}$  in our signature grid  $\Omega$  contains exactly one binary function (other than the EQUALITY functions). It turns out that allowing EQUALITY functions has a major influence on the computational complexity of the problems. By making the presence of these EQUALITY functions explicit, the Holant framework of counting problems can make a finer complexity classification, which is difficult to do in #CSP.

Our Holant problem framework is strongly influenced by the development of holographic algorithms and holographic reductions [34, 33, 9, 13]. Indeed, we will use and develop holographic reductions in this paper as a primary technique. One advantage of our new framework is that one can naturally consider new subclasses of counting

problems as special cases of Holant problems other than  $\#CSP$  problems. It turns out that a special class of constraint functions, called *degenerate* functions, plays a basic role in tractable cases. These are functions that can be expressed as tensor products of unary functions (a unary function is a function of arity one, i.e., on one variable). Degenerate functions are particularly weak. By assuming all unary functions are freely available, we propose an interesting class of counting problems called Holant\* problems. Our first main result is a complexity dichotomy theorem for all Holant\* problems for arbitrary complex valued symmetric functions over Boolean variables: Each problem in the class is either  $\#P$ -hard or solvable in polynomial time. In this dichotomy theorem, essentially all tractable cases are accomplished by holographic algorithms with Fibonacci gates [13, 12]. And what is more interesting and surprising is that the key technique used in the hardness proof is also holographic reductions. Furthermore, we prove that the dichotomy theorem also holds for planar graphs.

Our second main result is a dichotomy theorem for an even more appealing class of counting problems, called Holant<sup>c</sup> problems, where we assume only two special unary functions  $\Delta_0$  and  $\Delta_1$  are available. These two unary functions simply set a particular edge (variable) to a constant value 0 or 1, respectively. We can prove again that every problem in this class is either  $\#P$ -hard or solvable in  $P$ . However, here we can only prove it for all real valued symmetric functions over Boolean variables, as local constraint functions. Note that when we assume fewer functions are freely available in the framework, it makes the specification of the family more stringent. It delineates more precisely what functions and what combinations lead to  $\#P$ -hardness or to tractability, respectively. However, the fewer functions are assumed free, the more challenging it is to prove  $\#P$ -hardness. We make *essential* use of the dichotomy theorem just proved for Holant\* problems, as a launching station to prove our dichotomy theorem for Holant<sup>c</sup> problems.

Holant<sup>c</sup> problems are basically generic Holant problems with the ability to fix the assignments of some edges to constants. In many cases this is indeed very natural. By the Pinning Lemma in [18], in any  $\#CSP$  problem,  $\Delta_0$  and  $\Delta_1$  can be simulated (using EQUALITY functions), and as a result can be viewed as freely available. In other words EQUALITY functions are stronger than  $\Delta_0$  and  $\Delta_1$ . Therefore Holant<sup>c</sup> problems already subsume  $\#CSP$ , and meanwhile provide a way for a more exacting account of what makes a problem tractable or  $\#P$ -hard.

The main technique for the proof of the second dichotomy theorem is polynomial interpolation. Once we can interpolate all unary functions, we can apply the dichotomy theorem for Holant\* problems. Our dichotomy theorems have already paid dividends in the study of classifications of  $\#CSP$  problems. Since  $\#CSP$  can be viewed as a special case of Holant<sup>c</sup> problems, the dichotomy theorem for Holant<sup>c</sup> problems automatically implies a dichotomy theorem for Boolean  $\#CSP$  problems with real symmetric constraints. Motivated by this, we investigated how one might generalize the tractable cases to asymmetric ones. Surprisingly it turns out that the symmetric tractable cases already supplied the essential ingredients for all possible (including asymmetric) tractable ones. This led us to a dichotomy theorem for the whole family of complex weighted Boolean  $\#CSP$ . However, the proof requires substantial new techniques, and we report it in a separate paper [11, 14].

**2. Definitions and background.** Our functions take values in  $\mathbb{C}$  by default. In this paper, we will mostly be concerned with symmetric functions on Boolean variables as local constraint functions; however, the framework of Holant problems is defined for functions mapping any  $[q]^k \rightarrow \mathbb{F}$  for a finite  $q$  and a field  $\mathbb{F}$ . Results in this

paper are for the case  $q = 2$  and  $\mathbb{F} = \mathbb{C}$ , as well as  $\mathbb{F} = \mathbb{R}$  when specified. In order to avoid any difficulty with respect to models of computation, we will assume the function values are algebraic numbers. However, for convenience we will still use the symbol  $\mathbb{C}$ . (If we use the model of computation over  $\mathbb{C}$  from [1], then there is no need for any restriction.) We say a problem is tractable if it is computable in polynomial time. Since these are functions, this means computable in the class FP (we sometimes say simply in P, although strictly speaking the latter is a language class).

As stated, a *signature grid*  $\Omega = (H, \mathcal{F}, \pi)$  consists of a graph  $H = (V, E)$  with each vertex labeled by a function  $f_v \in \mathcal{F}$  of arity  $\deg(v)$ , where  $\deg(v)$  denotes the degree of  $v$ , and each edge is treated as a variable. We use the symbol  $\mathcal{F}_q$  when variables range over  $[q]$ . The Holant problem on instance  $\Omega$  is that of computing  $\text{Holant}_\Omega = \sum_\sigma \prod_{v \in V} f_v(\sigma|_{E(v)})$ , a sum over all edge assignments  $\sigma : E \rightarrow [q]$ . When  $q = 2$ , a function  $f_v$  can be represented as a vector of dimension  $2^{\deg(v)}$  listing all its values, or as a tensor in  $(\mathbb{C}^2)^{\otimes \deg(v)}$ . This is called a *signature*. We denote by  $=_k$  the EQUALITY signature of arity  $k$ .  $\Delta_0$  (respectively,  $\Delta_1$ ) denotes the unary signature that takes value 1 on input 0 (respectively, 1), and 0 otherwise. A symmetric function  $f$  on  $k$  Boolean variables can be denoted by  $[f_0, f_1, \dots, f_k]$ , where  $f_j$  is the value of  $f$  on inputs of Hamming weight  $j$ . Thus,  $(=_k) = [1, 0, \dots, 0, 1]$  (with  $k - 1$  many 0's),  $\Delta_0 = [1, 0]$ , and  $\Delta_1 = [0, 1]$ . A Holant problem is parameterized by a set of signatures.

**DEFINITION 2.1.** *Given a set of signatures  $\mathcal{F}$ , we define a counting problem  $\text{Holant}(\mathcal{F})$ :*

**Input:** A signature grid  $\Omega = (G, \mathcal{F}, \pi)$ ;

**Output:**  $\text{Holant}_\Omega$ .

We would like to characterize the complexity of Holant problems in terms of their signature sets. Some special families of Holant problems have already been widely studied. For example, if  $\mathcal{F}_q$  contains all EQUALITY signatures  $\{=1, =2, =3, \dots\}$ , then this is exactly the weighted #CSP problem. #CSP problems are a special family of Holant problems, in which we assume that all EQUALITY functions are freely available. Graph homomorphism is a further special case, where we allow only a single binary function in  $\mathcal{F}_q$  other than all the EQUALITY functions.

We now define two more special families of Holant problems by assuming some signatures are freely available. We define them for  $q = 2$ ; they can easily be extended to arbitrary  $[q]$ .

**DEFINITION 2.2.** *Let  $\mathcal{U}$  denote the set of all unary signatures. Given a set of signatures  $\mathcal{F}$ , we use  $\text{Holant}^*(\mathcal{F})$  to denote  $\text{Holant}(\mathcal{F} \cup \mathcal{U})$ .*

**DEFINITION 2.3.** *Given a set of signatures  $\mathcal{F}$ , we use  $\text{Holant}^c(\mathcal{F})$  to denote  $\text{Holant}(\mathcal{F} \cup \{\Delta_0, \Delta_1\})$ .*

Replacing a signature  $f \in \mathcal{F}$  by a constant multiple  $cf$ , where  $c \neq 0$ , does not change the complexity of  $\text{Holant}(\mathcal{F})$ . It only contributes an easily computable power of  $c$  as a global factor to  $\text{Holant}_\Omega$ . So we view  $f$  and  $cf$  as the same signature. An important property of a signature is whether it is degenerate.

**DEFINITION 2.4.** *A signature is degenerate iff it is a tensor product of unary signatures.*

In particular, a symmetric signature in  $\mathcal{F}$  is degenerate iff it can be expressed as  $\lambda[x, y]^{\otimes n}$ . Also a symmetric signature  $[x_0, x_1, \dots, x_n]$  is nondegenerate iff  $\text{rank} \begin{bmatrix} x_0 & \dots & x_{n-1} \\ x_1 & \dots & x_n \end{bmatrix} = 2$ . To see this, suppose  $[x_0, x_1, \dots, x_n] = \lambda[x, y]^{\otimes n}$ , and then clearly  $\text{rank} \begin{bmatrix} x_0 & \dots & x_{n-1} \\ x_1 & \dots & x_n \end{bmatrix} \leq 1$ . Conversely, suppose this rank is at most 1. If  $x_0 \neq 0$ , then there exists a  $c$ , such that  $x_i = cx_{i-1}$ , for  $1 \leq i \leq n$ , and we have  $x_0[1, c]^{\otimes n}$ . If  $x_0 = 0$ , since the rank is at most 1, an easy induction shows that  $x_1 = \dots = x_{n-1} = 0$ . Then it is  $x_n[0, 1]^{\otimes n}$ .

To introduce the idea of holographic reductions, it is convenient to consider bipartite graphs. For any general graph as a signature grid, we can make it bipartite by adding an additional vertex on each edge and giving each new vertex the EQUALITY function  $=_2$  on two inputs.

We use  $\#\mathcal{R}_q|\mathcal{G}_q$  to denote all counting problems, expressed as Holant problems on bipartite graphs  $H = (U, V, E)$ , where each signature for a vertex in  $U$  (respectively,  $V$ ) is from  $\mathcal{R}_q$  (respectively,  $\mathcal{G}_q$ ). An input instance for the bipartite Holant problem is a bipartite signature grid and is denoted as  $\Omega = (H, \mathcal{R}_q|\mathcal{G}_q, \pi)$ . Signatures in  $\mathcal{G}_q$  are denoted by column vectors (or contravariant tensors); signatures in  $\mathcal{R}_q$  are denoted by row vectors (or covariant tensors) [17, 8].

One can perform (contravariant and covariant) tensor transformations on the signatures. We will define a simple version of holographic reductions, which are invertible. Suppose  $\#\mathcal{R}_q|\mathcal{G}_q$  and  $\#\mathcal{R}'_q|\mathcal{G}'_q$  are two Holant problems defined for the same family of graphs, and  $T \in \mathbf{GL}_q(\mathbb{C})$ . We will use  $T$  to carry out a basis transformation. We say that there is a holographic reduction from  $\#\mathcal{R}_q|\mathcal{G}_q$  to  $\#\mathcal{R}'_q|\mathcal{G}'_q$ , if the *contravariant* transformation  $G' = T^{\otimes g}G$  and the *covariant* transformation  $R = R'T^{\otimes r}$  map  $G \in \mathcal{G}_q$  to  $G' \in \mathcal{G}'_q$  and  $R \in \mathcal{R}_q$  to  $R' \in \mathcal{R}'_q$ , where  $G$  and  $R$  have arity  $g$  and  $r$ , respectively. (Notice the reversal of direction when the transformations  $T^{\otimes g}$  versus  $T^{\otimes r}$  are applied. This is the meaning of *contravariance* and *covariance*.)

**THEOREM 2.5** (Valiant’s Holant theorem [34]). *Suppose there is a holographic reduction from  $\#\mathcal{R}_q|\mathcal{G}_q$  to  $\#\mathcal{R}'_q|\mathcal{G}'_q$  mapping signature grid  $\Omega$  to  $\Omega'$ ; then  $\text{Holant}_\Omega = \text{Holant}_{\Omega'}$ .*

In particular, if  $T$  is an invertible holographic reduction from  $\#\mathcal{R}_q|\mathcal{G}_q$  to  $\#\mathcal{R}'_q|\mathcal{G}'_q$ , and  $T^{-1}$  is a holographic reduction from  $\#\mathcal{R}'_q|\mathcal{G}'_q$  to  $\#\mathcal{R}_q|\mathcal{G}_q$ , then one problem is in P iff the other one is, and similarly one problem is #P-hard iff the other one is also.

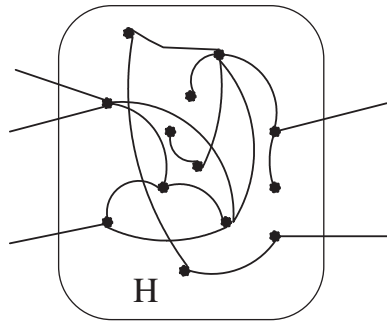
**THEOREM 2.6.** *Let  $\mathcal{F}_q$  be a set of signatures and  $M$  be a  $q \times q$  orthogonal matrix, i.e.,  $MM^T = I$ , the identity matrix. For any signature grid  $\Omega = (G, \mathcal{F}_q, \pi)$ , replacing every signature  $F \in \mathcal{F}_q$  by  $M^{\otimes n}F$ , where  $n$  is the arity of  $F$ , we can get a new signature grid  $\Omega'$ . Then  $\text{Holant}_\Omega = \text{Holant}_{\Omega'}$ .*

*Proof.* First we reformulate the signature grid  $\Omega = (G, \mathcal{F}_q, \pi)$ . We insert a new vertex on each edge of  $G$  with signature  $=_2$ ; the original edge becomes a path of length 2. This will not change the Holant value. Then for the new bipartite signature grid  $(G', \{=_2\}|\mathcal{F}_q, \pi')$ , we apply a holographic reduction with  $M$ . This will map a signature  $F \in \mathcal{F}_q$  to  $M^{\otimes n}F$ , where  $n$  is the arity of  $F$ . It is a direct consequence of  $MM^T = I$  that  $=_2$  will map to itself. Indeed, the covariant transformation maps  $=_2$  to  $(=_2)(M^{-1})^{\otimes 2}$  as a  $q^2$ -dimensional row vector. A binary function  $f$  can be denoted by a matrix, where the  $(i, j)$  entry is  $f(i, j)$ , for  $i, j \in [q]$ . Then  $=_2$  is the identity matrix  $I$ . In matrix notation,  $(=_2)(M^{-1})^{\otimes 2}$  is precisely  $(M^{-1})^T I M^{-1} = I$ . Now we can replace each new  $=_2$  node back to an edge to revert back to  $G$ . This gives the signature grid  $\Omega'$  as required. By the Holant theorem, its value is the same as  $\Omega$ .  $\square$

This theorem is very useful as a way to normalize a given signature set  $\mathcal{F}_q$ .

There is a technical issue with models of computation for  $\mathbb{C}$  [1, 27]. Strictly speaking, in the Turing model, we must only use computable numbers. We will state our results for all  $\mathbb{C}$ , assuming all numbers and their arithmetic operations  $(+, \cdot)$  in a particular instance (signature entries) are computable in polynomial time in the Turing sense. This is certainly the case when  $\mathcal{F}_q$  is a finite set and function values are all algebraic numbers.

**3. Some preliminary results.** A signature from  $\mathcal{F}_q$  at a vertex is considered a basic realizable function. Instead of a single vertex, we can use a graph fragment

FIG. 3.1. An  $\mathcal{F}$ -gate with five dangling edges.

to generalize this notion. An  $\mathcal{F}_q$ -gate  $\Gamma$  is a tuple  $(H, \mathcal{F}_q, \pi)$ , where  $H = (V, E, D)$  is a graph with regular edges  $E$  and some dangling edges  $D$ . (See Figure 3.1 for one example.) Other than these dangling edges, an  $\mathcal{F}_q$ -gate is the same as a signature grid. The role of dangling edges is to provide input/output variables. This is similar to that of external nodes in Valiant's notion for matchgates [32, 34]; however, we allow more than one dangling edge for a node. In  $H = (V, E, D)$  each node in  $V$  is assigned a function in  $\mathcal{F}_q$  by the mapping  $\pi$  (we do not consider "dangling" leaf nodes at the end of a dangling edge among these),  $E$  is the set of regular edges, denoted as  $1, 2, \dots, m$ , and  $D$  is the set of dangling edges, denoted as  $m + 1, m + 2, \dots, m + n$ . Then we can define a function for this  $\mathcal{F}$ -gate  $\Gamma = (H, \mathcal{F}_q, \pi)$ ,

$$\Gamma(y_1, y_2, \dots, y_n) = \sum_{x_1, x_2, \dots, x_m \in [q]} H(x_1, x_2, \dots, x_m, y_1, y_2, \dots, y_n),$$

where  $(y_1, y_2, \dots, y_n) \in [q]^n$  denotes an assignment on the dangling edges and  $H(x_1, x_2, \dots, x_m, y_1, y_2, \dots, y_n)$  denotes the value of the signature grid on an assignment of all edges. We will also call this function the signature of the  $\mathcal{F}_q$ -gate  $\Gamma$ . An  $\mathcal{F}_q$ -gate can be used in a signature grid as if it is just a single node with the particular signature. We note that even for a very simple signature set  $\mathcal{F}_q$ , the signatures for all  $\mathcal{F}_q$ -gates can be quite complicated and expressive. Matchgate signatures are an example [32].

Using the idea of  $\mathcal{F}_q$ -gates, we can reduce one Holant problem to another. Let  $g$  be the signature of some  $\mathcal{F}_q$ -gate  $\Gamma$ . Then  $\text{Holant}(\mathcal{F}_q \cup \{g\})$  is polynomial-time Turing reducible to  $\text{Holant}(\mathcal{F}_q)$ . The reduction is simple. Given an instance of  $\text{Holant}(\mathcal{F}_q \cup \{g\})$ , replacing every appearance of  $g$  by the  $\mathcal{F}_q$ -gate  $\Gamma$ , we get an instance of  $\text{Holant}(\mathcal{F}_q)$ . Since the signature of  $\Gamma$  is  $g$ , the Holant values for these two signature grids are identical.

We give some propositions that are useful in the proof in the next section. We first give one more definition.

**DEFINITION 3.1.** *Given a symmetric signature  $[x_0, x_1, \dots, x_n]$  and any  $l, r$  where  $0 \leq l < r \leq n$ , we call  $[x_l, x_{l+1}, \dots, x_r]$  a subsignature of  $[x_0, x_1, \dots, x_n]$  with arity  $r - l$ .*

**PROPOSITION 3.2.** *If  $[x_l, x_{l+1}, \dots, x_r]$  is a subsignature of  $[x_0, x_1, \dots, x_n]$  and  $\text{Holant}^*([x_l, x_{l+1}, \dots, x_r])$  is  $\#P$ -hard, then  $\text{Holant}^*([x_0, x_1, \dots, x_n])$  is  $\#P$ -hard. This is also true for  $\text{Holant}^c$ .*

*Proof.* We use unary signatures  $\Delta_0 = [1, 0]$ ,  $\Delta_1 = [0, 1]$ , and  $[x_0, x_1, \dots, x_n]$  to simulate a subsignature  $[x_l, x_{l+1}, \dots, x_r]$ . We connect  $l$  dangling edges of  $[x_0, x_1, \dots, x_n]$  to the unary signature  $[0, 1]$ , and connect another  $n - r$  dangling edges to the unary signature  $[1, 0]$ .  $\square$

PROPOSITION 3.3. *Let  $n \geq 3$  and let  $[x_0, x_1, \dots, x_n]$  be a nondegenerate symmetric signature. Then for any  $m = 2, 3, \dots, n - 1$ , there exists a nondegenerate subsignature of arity  $m$ , unless the signature is of the form  $[x_0, 0, \dots, 0, x_n]$ .*

*Proof.* Since  $\text{rank} \begin{bmatrix} x_0 & \dots & x_{n-1} \\ x_1 & \dots & x_n \end{bmatrix} = 2$ , there must be nonzero entries among  $x_0, x_1, \dots, x_n$ . If all entries are nonzero, then either  $\text{rank} \begin{bmatrix} x_0 & \dots & x_{n-2} \\ x_1 & \dots & x_{n-1} \end{bmatrix} = 2$  or  $\text{rank} \begin{bmatrix} x_1 & \dots & x_{n-1} \\ x_2 & \dots & x_n \end{bmatrix} = 2$ . Otherwise they are both of rank 1, and being nonzero, the second row is a multiple of the first row in both matrices. Since they share at least one column, and being nonzero, this multiplier must be the same, which says that  $\text{rank} \begin{bmatrix} x_0 & \dots & x_{n-1} \\ x_1 & \dots & x_n \end{bmatrix} = 1$ , a contradiction. Then we use induction to complete the proof.

Now suppose there are zero entries. Consider  $x_1, \dots, x_{n-1}$ . Since  $[x_0, \dots, x_n]$  is not of the form  $[x_0, 0, \dots, 0, x_n]$ , there must be some  $1 \leq i \leq n - 1$  such that  $x_i \neq 0$ . Find an  $x_i \neq 0$ , for some  $1 \leq i \leq n - 1$ , such that a neighbor  $x_{i-1} = 0$  or  $x_{i+1} = 0$ . Now any submatrix containing  $\begin{bmatrix} \dots & x_{i-1} & x_i & \dots \\ \dots & x_i & x_{i+1} & \dots \end{bmatrix}$  has rank 2.  $\square$

PROPOSITION 3.4. *Let  $\mathcal{F}$  be a set of signatures and  $M$  be a  $2 \times 2$  orthogonal matrix. We define a new set of signatures  $\mathcal{F}_M$  as follows:*

$$\mathcal{F}_M = \{Y \mid Y = M^{\otimes n} X, \text{ where } X \text{ is a signature in } \mathcal{F} \text{ with arity } n\}.$$

*Then  $\text{Holant}^*(\mathcal{F})$  and  $\text{Holant}^*(\mathcal{F}_M)$  have the same complexity; i.e., one problem is in  $P$  (respectively,  $\#P$ -hard) iff the other problem is also.*

*Proof.* Because  $M$  is nonsingular, the set of unary signatures  $\mathcal{U}$  is transformed to the same set  $\mathcal{U}$ . By Theorem 2.6,  $\text{Holant}^*(\mathcal{F})$  and  $\text{Holant}^*(\mathcal{F}_M)$  have the same value on the corresponding signature grids. This is clearly a two-way reduction between them.  $\square$

PROPOSITION 3.5. *Let  $[x_0, x_1, x_2, x_3]$  be a symmetric signature of arity three, expressed as row vectors*

$$[x_0, x_1, x_2, x_3] = c(\alpha_1, \alpha_2)^{\otimes 3} + d(\beta_1, \beta_2)^{\otimes 3}.$$

*Then*

$$\det \begin{bmatrix} x_0 & x_1 \\ x_1 & x_2 \end{bmatrix} = cd \det \begin{bmatrix} \alpha_1 & \beta_1 \\ \alpha_2 & \beta_2 \end{bmatrix}^2 \alpha_1 \beta_1, \quad \det \begin{bmatrix} x_1 & x_2 \\ x_2 & x_3 \end{bmatrix} = cd \det \begin{bmatrix} \alpha_1 & \beta_1 \\ \alpha_2 & \beta_2 \end{bmatrix}^2 \alpha_2 \beta_2,$$

*and*

$$\det \begin{bmatrix} x_0 - x_2 & x_1 \\ x_1 - x_3 & x_2 \end{bmatrix} = cd \det \begin{bmatrix} \alpha_1 & \beta_1 \\ \alpha_2 & \beta_2 \end{bmatrix}^2 (\alpha_1 \beta_1 + \alpha_2 \beta_2).$$

*Suppose  $cd \neq 0$  and  $\det \begin{bmatrix} \alpha_1 & \beta_1 \\ \alpha_2 & \beta_2 \end{bmatrix} \neq 0$ . Then  $[x_0, x_1, x_2, x_3]$  is nondegenerate. Furthermore,  $\det \begin{bmatrix} x_0 - x_2 & x_1 \\ x_1 - x_3 & x_2 \end{bmatrix} = 0$  iff  $\alpha_1 \beta_1 + \alpha_2 \beta_2 = 0$ . In the latter case, the unique (up to a scalar multiple) nonzero solution  $(a, b)$  to*

$$\begin{bmatrix} x_0 - x_2 & x_1 \\ x_1 - x_3 & x_2 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = 0$$

*is  $a = \alpha_1 \beta_1 = -\alpha_2 \beta_2$  and  $b = \alpha_1 \beta_2 + \alpha_2 \beta_1$ .*

*Proof.* A straightforward calculation.  $\square$

PROPOSITION 3.6. *Let  $[x_0, x_1, x_2, x_3]$  be a symmetric signature of arity three, expressed as*

$$[x_0, x_1, x_2, x_3] = c(\alpha_1, \alpha_2)^{\otimes 3} + d(\beta_1, \beta_2)^{\otimes 3},$$

where  $cd \neq 0$  and  $\det \begin{bmatrix} \alpha_1 & \beta_1 \\ \alpha_2 & \beta_2 \end{bmatrix} \neq 0$ . Then,  $[x_0, x_1, x_2, x_3]$  is of the form  $[x, y, -x, -y]$  iff  $\alpha_1^2 + \alpha_2^2 = \beta_1^2 + \beta_2^2 = 0$ .

*Proof.* The equations  $x_0 + x_2 = 0$  and  $x_1 + x_3 = 0$  are, respectively,  $c\alpha_1(\alpha_1^2 + \alpha_2^2) + d\beta_1(\beta_1^2 + \beta_2^2) = 0$  and  $c\alpha_2(\alpha_1^2 + \alpha_2^2) + d\beta_2(\beta_1^2 + \beta_2^2) = 0$ . Viewed as a linear equation system on  $c$  and  $d$ , there are nonzero solutions iff its determinant is zero:

$$\det \begin{bmatrix} \alpha_1 & \beta_1 \\ \alpha_2 & \beta_2 \end{bmatrix} (\alpha_1^2 + \alpha_2^2)(\beta_1^2 + \beta_2^2) = 0.$$

It follows that at least one of the factors  $\alpha_1^2 + \alpha_2^2 = 0$  or  $\beta_1^2 + \beta_2^2 = 0$ . In the first case, by the linear equation system, since  $(\beta_1, \beta_2)$  is a nonzero vector, we get the second factor  $\beta_1^2 + \beta_2^2 = 0$  as well. Similarly starting with  $\beta_1^2 + \beta_2^2 = 0$  we also get  $\alpha_1^2 + \alpha_2^2 = 0$ .

Conversely, if  $\alpha_2 = \pm i\alpha_1$  and  $\beta_2 = \pm i\beta_1$ , then  $c(\alpha_1, \alpha_2)^{\otimes 3} + d(\beta_1, \beta_2)^{\otimes 3}$  has the form  $[x, y, -x, -y]$ .  $\square$

**4. A dichotomy theorem for Holant\*( $\mathcal{F}$ ).** Our first dichotomy theorem is for Holant\*( $\mathcal{F}$ ), where  $\mathcal{F}$  is an arbitrary set of symmetric signatures on Boolean variables.

Since all unary signatures can be used for free in Holant\*( $\mathcal{F}$ ), we always assume the arity of every signature in  $\mathcal{F}$  is greater than one. And since any degenerate signature can be decomposed to unary signatures, we also assume that every signature in  $\mathcal{F}$  is nondegenerate.

**THEOREM 4.1.** *Let  $\mathcal{F}$  be a set of nondegenerate symmetric signatures over  $\mathbb{C}$ . Then Holant\*( $\mathcal{F}$ ) is computable in polynomial time in the following three classes. In all other cases, Holant\*( $\mathcal{F}$ ) is #P-hard.*

1. Every signature in  $\mathcal{F}$  is of arity no more than two.
2. There exist two constants  $a$  and  $b$  (not both zero, depending only on  $\mathcal{F}$ ), such that for all signatures  $[x_0, x_1, \dots, x_n] \in \mathcal{F}$  one of the two conditions is satisfied: (1) for every  $k = 0, 1, \dots, n-2$ , we have  $ax_k + bx_{k+1} - ax_{k+2} = 0$ ; (2)  $n = 2$  and the signature  $[x_0, x_1, x_2]$  is of the form  $[2a\lambda, b\lambda, -2a\lambda]$ .
3. For every signature  $[x_0, x_1, \dots, x_n] \in \mathcal{F}$  one of the two conditions is satisfied: (1) For every  $k = 0, 1, \dots, n-2$ , we have  $x_k + x_{k+2} = 0$ ; (2)  $n = 2$  and the signature  $[x_0, x_1, x_2]$  is of the form  $[\lambda, 0, \lambda]$ .

The dichotomy is still valid even if the inputs are restricted to planar graphs.

*Remark.* In order that Holant\*( $\mathcal{F}$ ) be a finitely specifiable problem with parameter  $\mathcal{F}$ , we may require  $\mathcal{F}$  to be a finite set. However, our dichotomy theorem is stronger, and applies to an infinite set  $\mathcal{F}$  in the following sense: When  $\mathcal{F}$  falls in one of the tractable classes, the Holant problem is computable in polynomial time even when  $\mathcal{F}$  is infinite, but the input size of the signature grid includes a description (in symmetric signature notation) of the functions at each node. On the other hand, when  $\mathcal{F}$  does not belong to one of the tractable classes, then there is a finite subset  $\mathcal{F}' \subseteq \mathcal{F}$  for which Holant\*( $\mathcal{F}'$ ) is #P-hard.

The statement of this dichotomy theorem essentially says that the only symmetric signature set  $\mathcal{F}$  for which Holant\*( $\mathcal{F}$ ) is tractable is when  $\mathcal{F}$  is degenerate, or of arity at most two, or most interestingly  $\mathcal{F}$  consists of Fibonacci gates under a holographic transformation [13, 12].

*Proof outline.* The first class to be computable in P is easy. One can compute the signature of a path by matrix multiplication and the signature of a cycle by taking trace. The other two polynomial time computable classes use holographic algorithms and mainly follow from our previous work on Fibonacci gates [13, 12]. See subsection 4.5.



Now for the proof of hardness, we first prove in Lemma 4.2 that the theorem holds if  $\mathcal{F}$  contains a single symmetric signature of arity three. The main technique is holographic reductions. The choice of specific reductions is informed by our previous work on signature theory in holographic algorithms [10, 9]. This theory gives us three categories in a certain parameterization for a signature of arity three according to its eigenvalues. For each category, we choose one #P-hard problem to reduce from, all using holographic reductions. In Lemma 4.3, we prove that if one signature of arity three has the form in class 2 of Theorem 4.1, and we combine it with another signature of arity two that is not in class 2, then the Holant\* problem is #P-hard. The main idea of the proof of Lemma 4.3 is to reduce it to Lemma 4.2 with holographic reductions. In Lemma 4.4, we prove the same thing for class 3. In Lemma 4.5 we extend it to a single signature of arbitrary arity. Finally in subsection 4.5 we extend the above proofs to a set of signatures of arbitrary arities and finish the proof of Theorem 4.1.

**4.1. The first lemma: A single ternary signature.** The following lemma is an important first step toward the proof of Theorem 4.1. It says that Theorem 4.1 holds if  $\mathcal{F}$  contains only one signature and it has arity three. Holographic reductions play a decisive role in the proof. This lemma serves as the foundation for all subsequent lemmas.

LEMMA 4.2. *Let  $[x_0, x_1, x_2, x_3]$  be a nondegenerate symmetric signature with arity three, then  $\text{Holant}^*([x_0, x_1, x_2, x_3])$  is #P-hard unless one of the following two statements is true: (1) there exist two constants  $a, b$  (not both zero) such that  $ax_0 + bx_1 - ax_2 = 0$  and  $ax_1 + bx_2 - ax_3 = 0$ ; (2)  $x_0 + x_2 = 0$  and  $x_1 + x_3 = 0$ .*

*Proof.* Assuming  $[x_0, x_1, x_2, x_3]$  does not satisfy either of the two statements, we prove that  $\text{Holant}^*([x_0, x_1, x_2, x_3])$  is #P-hard. Our starting point is that  $\#[0, 1, 1][1, 0, 0, 1]$  and  $\#[1, 0, 1][1, 1, 0, 0]$  are both #P-complete [35]. The first problem is to count the number of vertex covers for 3-regular graphs; while the second is to count the number of (not necessarily perfect) matchings for 3-regular graphs. We remark that both of them remain #P-complete even for planar graphs [16, 35].

First we give a useful parameterization. Given a nondegenerate signature  $[x_0, x_1, x_2, x_3]$ , there are three categories: For  $0 \leq k \leq 3$ ,

- Category 1.  $x_k = \alpha_1^{3-k} \alpha_2^k + \beta_1^{3-k} \beta_2^k$ , where  $\det \begin{bmatrix} \alpha_1 & \beta_1 \\ \alpha_2 & \beta_2 \end{bmatrix} \neq 0$ ;
- Category 2.  $x_k = ck\alpha^{k-1} + d\alpha^k$ , where  $c \neq 0$ ; or
- Category 3.  $x_k = c(3-k)\alpha^{2-k} + d\alpha^{3-k}$ , where  $c \neq 0$ .

This parameterization can be obtained by considering the rank 1 solution to

$$\begin{bmatrix} x_0 & x_1 & x_2 \\ x_1 & x_2 & x_3 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = 0.$$

If  $a = c = 0$  we have  $x_1 = x_2 = 0$  and a diagonal matrix  $\begin{bmatrix} \sqrt[3]{x_0} & 0 \\ 0 & \sqrt[3]{x_3} \end{bmatrix}$  can be used in Category 1. If  $a$  and  $c$  are not both 0, then we may consider  $[x_0, x_1, x_2, x_3]$  satisfies a second order linear recurrence relation (either forward or backward). Depending on whether the characteristic equation has two distinct roots (or, respectively, a double root), we have a case in Category 1 (or, respectively, Category 2 or 3). Category 3 can be viewed as the reversal of Category 2, so we will omit the proof for Category 3. The choices made here in this particular parameterization are informed by the signature theory [10, 9] that we had developed in the previous work. But one can directly check that for any nondegenerate signature  $[x_0, x_1, x_2, x_3]$ , one of these three

parameterizations is always possible. Note that, if  $\alpha = 0$ , then we take the convention that the expression  $k\alpha^{k-1} = 0, 1, 0, 0$  for  $k = 0, 1, 2, 3$ , respectively.

For Category 1, we have

$$X = [x_0, x_1, x_2, x_3] = \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix}^{\otimes 3} + \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}^{\otimes 3}.$$

We restate our conditions from the lemma’s statement in this new parameterization. The fact that  $X$  is nondegenerate implies that  $\alpha_1\beta_2 - \alpha_2\beta_1 \neq 0$ . The fact that  $X$  is not in the case indicated in statement (1) of Lemma 4.2 implies that  $\alpha_1\beta_1 + \alpha_2\beta_2 \neq 0$ . This follows from Proposition 3.5. The fact that  $X$  is not in the case indicated in statement (2) implies that  $\alpha_1^2 + \alpha_2^2 \neq 0$  or  $\beta_1^2 + \beta_2^2 \neq 0$ . This follows from Proposition 3.6. By symmetry, we can assume that  $\alpha_1^2 + \alpha_2^2 \neq 0$ .

Under the condition  $\alpha_1^2 + \alpha_2^2 \neq 0$ , we can apply an orthogonal transformation to map the vector  $(\alpha_1, \alpha_2)^T$  to  $(\alpha'_1, 0)^T$ , for some  $\alpha'_1 \neq 0$ . We may use the (complex orthogonal) matrix  $(1/\sqrt{\alpha_1^2 + \alpha_2^2}) \begin{bmatrix} \alpha_1 & \alpha_2 \\ -\alpha_2 & \alpha_1 \end{bmatrix}$  for this purpose. Then under this orthogonal transformation, the signature becomes

$$X' = [x'_0, x'_1, x'_2, x'_3] = \begin{bmatrix} \alpha'_1 \\ 0 \end{bmatrix}^{\otimes 3} + \begin{bmatrix} \beta'_1 \\ \beta'_2 \end{bmatrix}^{\otimes 3}.$$

By Proposition 3.4, this transformation does not change the complexity of the Holant problem. So it suffices to prove #P-hardness for this signature. By a scalar multiplication we assume  $\alpha'_1 = 1$ . So, reusing the notation  $X$ , we can assume the given signature is of this form

$$X = [x_0, x_1, x_2, x_3] = \begin{bmatrix} 1 \\ 0 \end{bmatrix}^{\otimes 3} + \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}^{\otimes 3}.$$

The two conditions from the statement of Lemma 4.2 become simply  $\beta_1\beta_2 \neq 0$ .

Now under the basis transformation  $T = \begin{bmatrix} 1 & \beta_1 \\ 0 & \beta_2 \end{bmatrix}$ , the EQUALITY signature  $(=_3) = [1, 0, 0, 1]$  becomes  $[x_0, x_1, x_2, x_3]$ . This is the result of the contravariant transformation  $(x_0, x_1, x_1, x_2, x_1, x_2, x_2, x_3)^T = T^{\otimes 3}(1, 0, 0, 0, 0, 0, 0, 1)^T$ , namely,  $X = T^{\otimes 3}(\begin{bmatrix} 1 \\ 0 \end{bmatrix}^{\otimes 3} + \begin{bmatrix} 0 \\ 1 \end{bmatrix}^{\otimes 3})$ . By the same matrix,  $[0, 1, 1]$  undergoes a covariant transformation. We have

$$(0, 1, 1, 1)(T^{-1})^{\otimes 2} = \frac{1}{\beta_2^2}(0, \beta_2, \beta_2, 1 - 2\beta_1).$$

Again, we can ignore the scalar factor  $1/\beta_2^2$ . So by the holographic reduction defined by  $T$ , the complexity of the problem  $\#[0, \beta_2, 1 - 2\beta_1][x_0, x_1, x_2, x_3]$  is the same as  $\#[0, 1, 1][1, 0, 0, 1]$ , which is #P-complete (vertex cover). In order to prove that  $\text{Holant}^*([x_0, x_1, x_2, x_3])$  is #P-hard, we only need to show that the signature  $[0, \beta_2, 1 - 2\beta_1]$  can be realized by  $[x_0, x_1, x_2, x_3]$  with some unary signatures.

For a binary signature  $F$  we can write it in a matrix form  $\begin{bmatrix} F(00) & F(01) \\ F(10) & F(11) \end{bmatrix}$ . We use the gadget in Figure 4.1 to realize  $[0, \beta_2, 1 - 2\beta_1]$ , where the two unary signatures  $(t_0, t_1)$  and  $(s_0, s_1)$  will be determined later. Let

$$A = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \quad B = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} \begin{bmatrix} \beta_1 & \beta_2 \end{bmatrix} = \begin{bmatrix} \beta_1^2 & \beta_1\beta_2 \\ \beta_1\beta_2 & \beta_2^2 \end{bmatrix}.$$

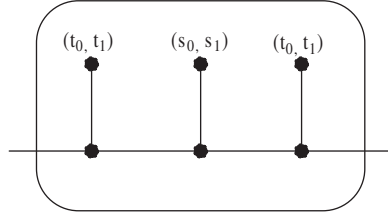


FIG. 4.1. We use this gadget to realize the signature  $[0, \beta_2, 1 - 2\beta_1]$ . All (three) nodes of degree 3 in this gadget have the signature  $[x_0, x_1, x_2, x_3]$ .

In  $X$ , if one input is 0, the induced binary signature has its matrix form  $A + \beta_1 B$ . If one input is 1, the induced binary signature has matrix form  $\beta_2 B$ . It follows that the signature of the above gadget, as a binary function in matrix form, is

$$(t_0(A + \beta_1 B) + t_1 \beta_2 B)(s_0(A + \beta_1 B) + s_1 \beta_2 B)(t_0(A + \beta_1 B) + t_1 \beta_2 B) \\ = (t_0 A + (t_0 \beta_1 + t_1 \beta_2) B)(s_0 A + (s_0 \beta_1 + s_1 \beta_2) B)(t_0 A + (t_0 \beta_1 + t_1 \beta_2) B).$$

Now we use a new set of variables

$$(4.1) \quad x = t_0, \quad y = t_0 \beta_1 + t_1 \beta_2, \quad z = s_0, \quad w = s_0 \beta_1 + s_1 \beta_2,$$

and write the above matrix as  $(xA + yB)(zA + wB)(xA + yB)$ . We note that, since  $\beta_2 \neq 0$ , for any given  $x, y, z, w$ , we can find  $t_0, t_1, s_0, s_1$  to satisfy the relationships (4.1). Then, to realize  $[0, \beta_2, 1 - 2\beta_1]$ , we just want to choose some  $x, y, z$ , and  $w$  such that

$$(xA + yB)(zA + wB)(xA + yB) = \begin{bmatrix} 0 & \beta_2 \\ \beta_2 & 1 - 2\beta_1 \end{bmatrix}.$$

We show that we can find some  $x, y, z$ , and  $w$  to satisfy the above condition.

Substituting  $A$  and  $B$ , and denoting  $\beta_1^2 + \beta_2^2$  by  $\gamma$ , we have the following:

$$(xA + yB)(zA + wB)(xA + yB) \\ = w \begin{bmatrix} \beta_1^2(x + y\gamma)^2 & y\beta_1\beta_2\gamma(x + y\gamma) \\ y\beta_1\beta_2\gamma(x + y\gamma) & y^2\beta_2^2\gamma^2 \end{bmatrix} + z \begin{bmatrix} (x + y\beta_1^2)^2 & y\beta_1\beta_2(x + y\beta_1^2) \\ y\beta_1\beta_2(x + y\beta_1^2) & y^2\beta_1^2\beta_2^2 \end{bmatrix}.$$

We may choose  $w = (x + y\beta_1^2)^2$  and  $z = -\beta_1^2(x + y\gamma)^2$  to make the (1, 1) entry zero. The (1, 2) (and (2, 1)) entry is

$$g_1 = xy\beta_1\beta_2^3(x + y\beta_1^2)(x + y\gamma);$$

and the (2, 2) entry is

$$g_2 = xy^2\beta_2^4(x(2\beta_1^2 + \beta_2^2) + 2y(\beta_1^4 + \beta_1^2\beta_2^2)).$$

We want to choose some  $x$  and  $y$  such that  $[g_1, g_2] = [\beta_2, 1 - 2\beta_1]$ . We have  $\beta_2 \neq 0$ . We will choose  $xy \neq 0$ . As both  $g_1$  and  $g_2$  are homogeneous (of degree 4) in  $x$  and  $y$ , we can ignore (and remove) the common factor  $xy\beta_2^3$  of  $g_1$  and  $g_2$ . It follows that we only have to satisfy that  $g_2/g_1 = (1 - 2\beta_1)/\beta_2$  and  $g_1 \neq 0$ , with  $y = 1$ . Setting  $\beta_2 g_2 = (1 - 2\beta_1)g_1$ , we will satisfy the following:

$$(4.2) \quad \beta_1(2\beta_1 - 1)x^2 + (2\beta_1^2 - \beta_1 + \beta_2^2)(2\beta_1^2 + \beta_2^2)x + \beta_1^2(\beta_1^2 + \beta_2^2)(2\beta_1^2 - \beta_1 + 2\beta_2^2) = 0.$$

What we have to prove is that at least one of the roots  $x$  to the equation in (4.2) is not a root of  $g_1 = g_1(x, 1) = 0$ . The roots of  $g_1 = 0$  are  $x = 0$ ,  $x = -\beta_1^2$ , and  $x = -\gamma$ . First we can verify that  $x = -\beta_1^2$  is not a root of (4.2). This is because when  $x = -\beta_1^2$ , the expression in (4.2) can be simplified to  $\beta_1^2\beta_2^4 \neq 0$ . Second when  $x = -\gamma$  the expression in (4.2) can be simplified to  $-\beta_2^4\gamma$ , and if  $x = -\gamma$  is a root, this would force  $\gamma = 0$ . So, assuming the expression in (4.2) as a polynomial in  $x$  is indeed of degree 2, then the only case we need to worry about is that  $x = 0$  is a double root of (4.2). Indeed, suppose (4.2) is of degree 2, and  $x = 0$  is not a double root, then we may let  $\xi \neq 0$  be a root of (4.2). This implies  $\xi \neq -\beta_1^2$ , because  $-\beta_1^2$  is not a root of (4.2);  $\xi$  cannot be  $-\gamma$  either, for otherwise  $-\gamma$  would be a root of (4.2), which we had proved would force  $\gamma = 0$ , and thus  $\xi = -\gamma = 0$ , a contradiction. Thus  $\xi$  is a root of (4.2) but not a root of  $g_1$ , as is needed.

Now let us consider the exceptional cases: either  $x = 0$  is a double root of (4.2) or (4.2) has degree less than 2. If  $x = 0$  is a double root of (4.2), we have

$$(2\beta_1^2 - \beta_1 + \beta_2^2)(2\beta_1^2 + \beta_2^2) = \beta_1^2(\beta_1^2 + \beta_2^2)(2\beta_1^2 - \beta_1 + 2\beta_2^2) = 0.$$

To satisfy this, since  $\beta_1\beta_2 \neq 0$ , there are only four exceptional cases (A1 to A4):  $\beta_1 = 1, \beta_2 = \pm i$  or  $\beta_1 = -\frac{1}{2}, \beta_2 = \pm \frac{i}{\sqrt{2}}$ . On the other hand, if the polynomial in (4.2) has degree less than 2, by  $\beta_1 \neq 0$ , we get  $\beta_1 = \frac{1}{2}$ . In this case, the polynomial becomes

$$(1/2 + \beta_2^2)x + (1/4 + \beta_2^2)/2 = 0.$$

This gives us four additional exceptional cases (B1 to B4):  $\beta_1 = \frac{1}{2}, \beta_2 = \pm \frac{i}{2}$ , in which case the polynomial is linear with root  $x = 0$ ; or  $\beta_1 = \frac{1}{2}, \beta_2 = \pm \frac{i}{\sqrt{2}}$ , in which case the polynomial degenerates to a (nonzero) constant. In all other cases, there is a root of (4.2) which is not a root of  $g_1(x, 1)$ , which completes the #P-hardness proof.

For the cases A1 and A2, we use a new starting problem  $\#[1, 1, 0][1, 0, 0, 1]$ , which is the reversal of the previous problem, and therefore it is also #P-complete. Then all previous parts of the proof are still valid, except that the signature of arity two we would like to realize is

$$(1, 1, 1, 0)(T^{-1})^{\otimes 2} = \left(1, \frac{1 - \beta_1}{\beta_2}, \frac{1 - \beta_1}{\beta_2}, \frac{\beta_1^2 - 2\beta_1}{\beta_2^2}\right).$$

Substituting  $\beta_1 = 1, \beta_2 = \pm i$ , the signature is  $[1, 0, 1]$ , which is trivially realizable by one edge. So we have proved that it is #P-hard in the cases A1 and A2. Now consider the cases A3 and A4:  $\beta_1 = -\frac{1}{2}, \beta_2 = \pm \frac{i}{\sqrt{2}}$ . We will give a different parameterization.

For the case A3, we apply an orthogonal transformation  $M = \begin{bmatrix} -i & -\sqrt{2} \\ \sqrt{2} & -i \end{bmatrix}$  and a scalar multiplier  $(2i)^3$  on the signature and it becomes  $[\frac{1}{0}]^{\otimes 3} + [\frac{2}{2\sqrt{2}i}]^{\otimes 3}$ . This is not one of the exceptional cases, and we have proved that it is #P-hard. For the case A4, we apply another orthogonal transformation  $M' = \begin{bmatrix} i & -\sqrt{2} \\ \sqrt{2} & i \end{bmatrix}$  and a scalar multiplier  $(-2i)^3$  on the signature, and it becomes  $[\frac{1}{0}]^{\otimes 3} + [\frac{2}{-2\sqrt{2}i}]^{\otimes 3}$ .

The cases B3 and B4 can be shown by the same method as in A4 and A3, using  $M'$  and  $M$ , respectively. The only remaining cases are B1 and B2. Here we will use another gadget similar to the one in Figure 4.1 except we remove the middle edge (including the node labeled  $(s_0, s_1)$  and the middle node of degree 3). For B1,  $\beta_1 = \frac{1}{2}, \beta_2 = \frac{i}{2}$ , the signature of this gadget is

$$(t_0(A + \beta_1 B) + t_1\beta_2 B)^2 = (xA + yB)^2,$$

where  $A$  and  $B$  are as before, and with the specific values of  $\beta_1, \beta_2$ ,  $B = \frac{1}{4} \begin{bmatrix} 1 & i \\ i & -1 \end{bmatrix}$ . By setting  $x = i$  and  $y = -2i$ , we have  $(xA + yB)^2 = \begin{bmatrix} 0 & i/2 \\ i/2 & 0 \end{bmatrix}$ , which is the matrix form of the target signature  $[0, \beta_2, 1 - 2\beta_1] = [0, \frac{i}{2}, 0]$ . This finishes case B1. The case B2 can be proved the same way with  $x = 1$  and  $y = -2$ .

Now we prove for Category 2. In this case  $x_k = ck\alpha^{k-1} + d\alpha^k$ , the condition that it does not satisfy statement (1) in Lemma 4.2 implies that  $\alpha \neq \pm i$ . Indeed,  $x_k$  satisfies the recurrence  $x_{k+2} = 2\alpha x_{k+1} - \alpha^2 x_k$ , for  $k = 0, 1$ . (Note that the characteristic polynomial  $\lambda^2 - 2\alpha\lambda + \alpha^2$  has a double root  $\alpha$ .) If  $\alpha = \pm i$ ,  $x_k$  would satisfy statement (1) in Lemma 4.2 with  $(a, b) = (1, 2\alpha)$ . Since  $\alpha \neq \pm i$ , we can choose some orthogonal matrix to transform the signature to the form  $[x, y, 0, 0]$  where  $y \neq 0$ . In fact, if we let  $T = \begin{bmatrix} 1 & \frac{d-1}{3} \\ \alpha & c + \frac{d-1}{3}\alpha \end{bmatrix}$  (which has  $\det T = c \neq 0$ ), then the signature  $[x_0, x_1, x_2, x_3]$  can be expressed as

$$(x_0, x_1, x_1, x_2, x_1, x_2, x_2, x_3)^T = T^{\otimes 3}(1, 1, 1, 0, 1, 0, 0, 0)^T.$$

(We chose these transformations based on an underlying signature theory of holographic algorithms [9], not “out of blue.” But for brevity of exposition we state these transformations *as is* without discussing the background. They can be directly verified, albeit a bit tedious.) Let  $T = QR$  be its  $QR$ -factorization, i.e., where  $Q$  is orthogonal and  $R$  is upper triangular. In fact, for  $T = \begin{bmatrix} 1 & * \\ \alpha & * \end{bmatrix}$ , we can choose our  $Q$  as the (complex) orthogonal matrix  $Q = Q^T = \frac{1}{\sqrt{1+\alpha^2}} \begin{bmatrix} 1 & \\ \alpha & -1 \end{bmatrix}$ . Then  $QT = R = \begin{bmatrix} u & w \\ 0 & v \end{bmatrix}$  is upper triangular, where  $u = \sqrt{1 + \alpha^2}$ . Because  $\det Q = -1$  and  $\det R = -\det T = -c \neq 0$ , we have  $uv \neq 0$ . This  $Q$  is our choice of the orthogonal transformation. It follows that

$$\begin{aligned} & Q^{\otimes 3}(x_0, x_1, x_1, x_2, x_1, x_2, x_2, x_3)^T \\ &= (QT)^{\otimes 3}(1, 1, 1, 0, 1, 0, 0, 0)^T \\ &= R^{\otimes 3}(1, 1, 1, 0, 1, 0, 0, 0)^T \\ &= R^{\otimes 3} \left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}^{\otimes 3} + \begin{bmatrix} 1 \\ 0 \end{bmatrix} \otimes \begin{bmatrix} 1 \\ 0 \end{bmatrix} \otimes \begin{bmatrix} 0 \\ 1 \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \end{bmatrix} \otimes \begin{bmatrix} 0 \\ 1 \end{bmatrix} \otimes \begin{bmatrix} 1 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} \otimes \begin{bmatrix} 1 \\ 0 \end{bmatrix} \otimes \begin{bmatrix} 1 \\ 0 \end{bmatrix} \right\} \\ &= \begin{bmatrix} u \\ 0 \end{bmatrix}^{\otimes 3} + \begin{bmatrix} u \\ 0 \end{bmatrix} \otimes \begin{bmatrix} u \\ 0 \end{bmatrix} \otimes \begin{bmatrix} w \\ v \end{bmatrix} + \begin{bmatrix} u \\ 0 \end{bmatrix} \otimes \begin{bmatrix} w \\ v \end{bmatrix} \otimes \begin{bmatrix} u \\ 0 \end{bmatrix} + \begin{bmatrix} w \\ v \end{bmatrix} \otimes \begin{bmatrix} u \\ 0 \end{bmatrix} \otimes \begin{bmatrix} u \\ 0 \end{bmatrix}. \end{aligned}$$

In the notation for symmetric signatures, this signature is  $[u^3 + 3u^2w, u^2v, 0, 0]$ . Note that the entry  $u^2v \neq 0$ .

By a scalar multiplication, we can make the entry  $u^2v$  equal to 1, so we only have to deal with a signature of the form  $[v, 1, 0, 0]$  for an arbitrary given  $v$ .

For this signature, we can apply a holographic transformation defined by the matrix  $T' = \begin{bmatrix} 1 & (v-1)/3 \\ 0 & 1 \end{bmatrix}$  with inverse  $T'^{-1} = \begin{bmatrix} 1 & -(v-1)/3 \\ 0 & 1 \end{bmatrix}$ . To prove #P-hardness, we will reduce from the matching problem  $\#[1, 0, 1] \mid [1, 1, 0, 0]$ . (Note that this is counting all, not necessarily perfect, matchings on 3-regular graphs, which remains #P-complete even for planar graphs [16]; counting perfect matchings on planar graphs is in P.) Under a contravariant transformation  $(v, 1, 1, 0, 1, 0, 0, 0)^T = T'^{\otimes 3}(1, 1, 1, 0, 1, 0, 0, 0)^T$ , the signature  $[1, 1, 0, 0]$  becomes  $[v, 1, 0, 0]$ . Under the covariant transformation by the same matrix,  $[1, 0, 1]$  becomes

$$(1, 0, 0, 1)(T'^{-1})^{\otimes 2} = ((1, 0)^{\otimes 2} + (0, 1)^{\otimes 2})(T'^{-1})^{\otimes 2} = \left( 1, \frac{1-v}{3}, \frac{1-v}{3}, 1 + \frac{(1-v)^2}{9} \right);$$

i.e., the signature  $[1, 0, 1]$  becomes a new symmetric signature  $[1, \frac{1-v}{3}, 1 + \frac{(1-v)^2}{9}]$ . We complete this proof by using the same gadget as in Figure 4.1 to realize this signature, using unary signatures and  $[v, 1, 0, 0]$ .

We will rename the values  $x = t_0$ ,  $y = t_1$ ,  $z = s_0$ , and  $w = s_1$  in Figure 4.1. The signature of this gadget in matrix form is  $(xA + yB)(zA + wB)(xA + yB)$ , where  $A = \begin{bmatrix} v & 1 \\ 1 & 0 \end{bmatrix}$  and  $B = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$ . After some calculations we found that this signature in matrix form is

$$\begin{bmatrix} w(xv + y)^2 + z(x^2(v^3 + 2v) + 2xy(v^2 + 1) + y^2v) & wx(xv + y) + z(x^2(v^2 + 1) + xyv) \\ w(x^2v + xy) + z(x^2(v^2 + 1) + xyv) & wx^2 + zx^2v \end{bmatrix}.$$

Our goal is to choose  $x$ ,  $y$ ,  $z$  and  $w$  such that it is equal to  $\begin{bmatrix} 1 & (1-v)/3 \\ (1-v)/3 & 1 + ((1-v)^2)/9 \end{bmatrix}$ . We can write this as a system of three linear equations in  $z$  and  $w$ , whose coefficient matrix depends on  $x$  and  $y$ . Then we can complete the proof, if we can choose  $x$  and  $y$  such that the following matrix has determinant 0, yet the first two columns have rank 2.

$$\begin{bmatrix} x^2v^2 + 2xyv + y^2 & x^2(v^3 + 2v) + 2xy(v^2 + 1) + y^2v & 1 \\ x^2v + xy & x^2(v^2 + 1) + xyv & \frac{1-v}{3} \\ x^2 & x^2v & 1 + \frac{(1-v)^2}{9} \end{bmatrix}.$$

After some elementary row operations it becomes

$$\begin{bmatrix} y^2 & 2xy + y^2v & f_3 \\ xy & x^2 + xyv & f_2 \\ x^2 & x^2v & f_1 \end{bmatrix},$$

where  $f_1, f_2, f_3$  are polynomials in  $v$ . Explicitly, let  $f_1 = (10 - 2v + v^2)/9$  and  $f_2 = (3 - 13v + 2v^2 - v^3)/9$ . Subtracting from the second column the first column multiplied by  $v$ , we get

$$\begin{bmatrix} y^2 & 2xy & f_3 \\ xy & x^2 & f_2 \\ x^2 & 0 & f_1 \end{bmatrix}.$$

We will set  $x = 1$ ; this guarantees that the first two columns have rank 2, and gives the matrix

$$\begin{bmatrix} y^2 & 2y & f_3 \\ y & 1 & f_2 \\ 1 & 0 & f_1 \end{bmatrix}.$$

Now the determinant is easily calculated (subtract the first row by the second row multiplied by  $y$ , and the second from the third multiplied by  $y$ ). The determinant is  $-(f_1y^2 - 2f_2y + f_3)$ . As long as  $f_1$  and  $f_2$  are not simultaneously 0, we can always choose a  $y$  to make this determinant 0.

However, it is easy to show that  $f_1$  and  $f_2$  have no common zero in  $v$ , as  $3(f_2 + vf_1) = 1 - v$  and  $v = 1$  is not a zero of either  $f_1$  or  $f_2$ . This completes the proof of Lemma 4.2.  $\square$

**4.2. A pair of signatures: Case (1).** Lemma 4.2 shows what happens when there is a single nondegenerate symmetric signature of arity three. It explicitly lists

two exceptional cases for being *not* #P-hard. The next lemma addresses what happens when we have more than one signature, where one signature of arity three happens to be in the exceptional case (1) of Lemma 4.2, but some other signature *does not fit*.

LEMMA 4.3. *Let  $[x_0, x_1, x_2, x_3]$  and  $[y_0, y_1, y_2]$  be nondegenerate symmetric signatures with arity three and two, respectively. Suppose there exist two constants  $a, b$  (not both zero), such that  $ax_0 + bx_1 - ax_2 = 0$  and  $ax_1 + bx_2 - ax_3 = 0$ , but  $ay_0 + by_1 - ay_2 \neq 0$  and  $[y_0, y_1, y_2]$  is not of the form  $[2a\lambda, b\lambda, -2a\lambda]$ . Then  $\text{Holant}^*([x_0, x_1, x_2, x_3], [y_0, y_1, y_2])$  is #P-hard. In fact, it is #P-hard when restricted to bipartite graphs, where  $[x_0, x_1, x_2, x_3]$  is on one side and  $[y_0, y_1, y_2]$  is on another.*

We remark that, since  $[x_0, x_1, x_2, x_3]$  is nondegenerate, the constant pair  $(a, b)$  is unique up to a scalar factor.

*Proof.* Our proof plan is as follows: We will show that the counting problem  $\text{Holant}^*([x_0, x_1, x_2, x_3], [y_0, y_1, y_2])$  is still #P-hard when restricted to instances where the input graph is bipartite, where other than degree one nodes, all nodes on one side are of degree three and are given the signature  $[x_0, x_1, x_2, x_3]$ , and all nodes on the other side are of degree two and are given the signature  $[y_0, y_1, y_2]$ . (There can be any number of degree one nodes, on either side of the bipartite graph, assigned any unary signatures.) We may denote this problem  $\text{Holant}^*([x_0, x_1, x_2, x_3] \mid [y_0, y_1, y_2])$ . We show it is #P-hard by a holographic reduction, where  $[y_0, y_1, y_2]$  is (contravariantly) transformed to  $[1, 0, 1]$  and  $[x_0, x_1, x_2, x_3]$  is (covariantly) transformed to some  $[u_0, u_1, u_2, u_3]$ . Note that  $[1, 0, 1]$  can be replaced by an edge, and the unary signatures are transformed to unary signatures. Thus the complexity of the problem  $\text{Holant}^*([x_0, x_1, x_2, x_3] \mid [y_0, y_1, y_2])$  is the same as  $\text{Holant}^*([u_0, u_1, u_2, u_3])$ . We then apply Lemma 4.2 to  $[u_0, u_1, u_2, u_3]$ . Note that  $[u_0, u_1, u_2, u_3]$  is nondegenerate, since otherwise  $[u_0, u_1, u_2, u_3]$  is a tensor product of unary signatures, and then  $[x_0, x_1, x_2, x_3]$  would also be a tensor product of unary signatures, and thus degenerate.

We want an invertible matrix  $T = \begin{bmatrix} \alpha_1 & \beta_1 \\ \alpha_2 & \beta_2 \end{bmatrix}$  such that  $(T^{-1})^{\otimes 2} [y_0, y_1, y_2]^T = [1, 0, 1]^T$ , but  $[u_0, u_1, u_2, u_3] = [x_0, x_1, x_2, x_3]T^{\otimes 3}$  is not of either of the two exceptional forms in Lemma 4.2.  $T^{\otimes 2} [1, 0, 1]^T = [y_0, y_1, y_2]^T$  can be written as

$$T^{\otimes 2} (1, 0, 0, 1)^T = T^{\otimes 2} \left( \begin{bmatrix} 1 \\ 0 \end{bmatrix}^{\otimes 2} + \begin{bmatrix} 0 \\ 1 \end{bmatrix}^{\otimes 2} \right) = \left( T \begin{bmatrix} 1 \\ 0 \end{bmatrix} \right)^{\otimes 2} + \left( T \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right)^{\otimes 2} = (y_0, y_1, y_1, y_2)^T.$$

This is equivalent to  $\begin{bmatrix} y_0 & y_1 \\ y_1 & y_2 \end{bmatrix} = TT^T$ , in matrix notation. Such a factorization for a symmetric matrix is always possible for complex matrices. Since  $[y_0, y_1, y_2]$  is nondegenerate, we have a full rank matrix  $T = \begin{bmatrix} \alpha_1 & \beta_1 \\ \alpha_2 & \beta_2 \end{bmatrix}$  such that  $(y_0, y_1, y_1, y_2)^T = T^{\otimes 2} (1, 0, 0, 1)^T$ .

Let  $[u_0, u_1, u_2, u_3] = [x_0, x_1, x_2, x_3]T^{\otimes 3}$ . We want to show that it does not fall into either of the two exceptional cases of Lemma 4.2.

There are two cases. First we assume for a contradiction that  $[u_0, u_1, u_2, u_3]$  is of the form  $[x, y, -x, -y]$ . Then  $\{u_k\}$  satisfies the recurrence  $u_{k+2} + u_k = 0$ , for  $k = 0, 1$ , with characteristic polynomial  $\lambda^2 + 1$ . Therefore there are scalars  $c$  and  $d$ , such that

$$[u_0, u_1, u_2, u_3] = c(1, i)^{\otimes 3} + d(1, -i)^{\otimes 3}.$$

Let  $\Delta = \det(T) \neq 0$ , we have  $T^{-1} = \Delta^{-1} \begin{bmatrix} \beta_2 & -\beta_1 \\ -\alpha_2 & \alpha_1 \end{bmatrix}$ . Then it follows that

$$\begin{aligned} [x_0, x_1, x_2, x_3] &= [u_0, u_1, u_2, u_3](T^{-1})^{\otimes 3} \\ &= c\Delta^{-3}(\beta_2 - \alpha_2 i, -\beta_1 + \alpha_1 i)^{\otimes 3} + d\Delta^{-3}(\beta_2 + \alpha_2 i, -\beta_1 - \alpha_1 i)^{\otimes 3}. \end{aligned}$$

Since  $[x_0, x_1, x_2, x_3]$  is nondegenerate,  $cd \neq 0$ . By the assumption on  $[x_0, x_1, x_2, x_3]$  and Proposition 3.5, we have

$$0 = (\beta_2 - \alpha_2 i)(\beta_2 + \alpha_2 i) + (-\beta_1 + \alpha_1 i)(-\beta_1 - \alpha_1 i) = \alpha_1^2 + \alpha_2^2 + \beta_1^2 + \beta_2^2.$$

And after a scaling, we also have  $a = \alpha_2^2 + \beta_2^2 = -\alpha_1^2 - \beta_1^2$  and  $b = -2(\alpha_1\alpha_2 + \beta_1\beta_2)$ . Now

$$(y_0, y_1, y_1, y_2)^T = T^{\otimes 2}(1, 0, 0, 1)^T = (\alpha_1^2 + \beta_1^2, \alpha_1\alpha_2 + \beta_1\beta_2, \alpha_1\alpha_2 + \beta_1\beta_2, \alpha_2^2 + \beta_2^2)^T.$$

Then  $[y_0, y_1, y_2] = [-a, -b/2, a]$  is of the form  $[2a\lambda, b\lambda, -2a\lambda]$ , a contradiction.

This completes our first step. Our second step is to take care of the other exceptional case (exception (1) in Lemma 4.2).

This exceptional case is that there exist two constants  $a', b'$  (not both zero), such that  $a'u_0 + b'u_1 - a'u_2 = 0$  and  $a'u_1 + b'u_2 - a'u_3 = 0$ . We split this exceptional case into two cases depending on whether  $a = 0$  or  $a \neq 0$ , given in the statement of Lemma 4.3.

If  $a = 0$ , then  $b \neq 0$  and  $x_1 = x_2 = 0$ . As  $[x_0, 0, 0, x_3]$  is nondegenerate,  $x_0x_3 \neq 0$ . We can write

$$[x_0, 0, 0, x_3] = x_0 \begin{pmatrix} 1 & 0 \end{pmatrix}^{\otimes 3} + x_3 \begin{pmatrix} 0 & 1 \end{pmatrix}^{\otimes 3},$$

and

$$[u_0, u_1, u_2, u_3] = [x_0, 0, 0, x_3]T^{\otimes 3} = x_0((1 \ 0)T)^{\otimes 3} + x_3((0 \ 1)T)^{\otimes 3}.$$

The existence of  $a', b'$  (not both zero), such that  $a'u_0 + b'u_1 - a'u_2 = 0$  and  $a'u_1 + b'u_2 - a'u_3 = 0$ , implies that the matrix  $\begin{bmatrix} u_2 - u_0 & u_1 \\ u_3 - u_1 & u_2 \end{bmatrix}$  is degenerate. By Proposition 3.5 we have the inner product of  $(1, 0)T = (\alpha_1, \beta_1)$  and  $(0, 1)T = (\alpha_2, \beta_2)$  is zero, i.e.,

$$\alpha_1\alpha_2 + \beta_1\beta_2 = 0.$$

But noticing that  $y_1 = \alpha_1\alpha_2 + \beta_1\beta_2$  and  $a = 0$ , we have  $ay_0 + by_1 - ay_2 = 0$ , a contradiction.

Finally we consider the case  $a \neq 0$ . By a scaling we can assume  $a = 1$ . Then  $x_2 = bx_1 + x_0$ ,  $x_3 = bx_2 + x_1 = b^2x_1 + bx_0 + x_1$ . Assume temporarily that  $b \neq \pm 2i$ , then the recurrence has two distinct eigenvalues, and therefore Proposition 3.5 applies. After some calculation we can obtain an expression for the determinant

$$(4.3) \quad \det \begin{bmatrix} u_0 - u_2 & u_1 \\ u_1 - u_3 & u_2 \end{bmatrix} = \det \begin{bmatrix} \alpha_1 & \beta_1 \\ \alpha_2 & \beta_2 \end{bmatrix}^2 (\alpha_1^2 + \beta_1^2 + b(\alpha_1\alpha_2 + \beta_1\beta_2) - \alpha_2^2 - \beta_2^2)(bx_0x_1 - x_1^2 + x_0^2).$$

However, for a fixed  $T$ , the above polynomial identity is valid for any  $b \neq \pm 2i$ , when we define a recurrence  $x_{k+2} = bx_{k+1} + x_k$ ,  $k = 0, 1$ , and then define  $[u_0, u_1, u_2, u_3] = [x_0, x_1, x_2, x_3]T^{\otimes 3}$ . Hence, it is also valid for  $b = \pm 2i$ , and so we dispense with the temporary assumption that  $b \neq \pm 2i$ .

The condition that there exist two constants  $a', b'$  (not both zero), such that  $a'u_0 + b'u_1 - a'u_2 = 0$  and  $a'u_1 + b'u_2 - a'u_3 = 0$  is equivalent to  $\det \begin{bmatrix} u_0 - u_2 & u_1 \\ u_1 - u_3 & u_2 \end{bmatrix} = 0$ . We know that  $\det \begin{bmatrix} \alpha_1 & \beta_1 \\ \alpha_2 & \beta_2 \end{bmatrix} \neq 0$ ; the second factor  $\alpha_1^2 + \beta_1^2 + b(\alpha_1\alpha_2 + \beta_1\beta_2) - \alpha_2^2 - \beta_2^2$  in (4.3) is exactly  $y_0 + by_1 - y_2$  which is nonzero. So we have  $bx_0x_1 - x_1^2 + x_0^2 = 0$ . Since  $bx_0x_1 + x_0^2 - x_1^2 = x_0x_2 - x_1^2 = \det \begin{bmatrix} x_0 & x_1 \\ x_1 & x_2 \end{bmatrix}$ , and the third column in  $\begin{bmatrix} x_0 & x_1 & x_2 \\ x_1 & x_2 & x_3 \end{bmatrix}$  is linearly dependent on the first two columns, this says that  $[x_0, x_1, x_2, x_3]$  is degenerate, a contradiction.  $\square$



**4.3. A pair of signatures: Case (2).** Lemma 4.4 does the same thing as Lemma 4.3 for the second exceptional case of Lemma 4.2 for a pair of signatures.

LEMMA 4.4. *Let  $[x, y, -x, -y]$  be a symmetric signature with arity three and  $[y_0, y_1, y_2]$  be a symmetric signature with arity two. Suppose they are both nondegenerate. If  $y_0 + y_2 \neq 0$  and  $[y_0, y_1, y_2]$  is not of the form  $[\lambda, 0, \lambda]$ , then  $\text{Holant}^*([x, y, -x, -y], [y_0, y_1, y_2])$  is  $\#P$ -hard. In fact, it is  $\#P$ -hard when restricted to bipartite graphs where  $[x, y, -x, -y]$  is on one side and  $[y_0, y_1, y_2]$  is on another.*

*Proof.* The overall plan of the proof of this Lemma is the same as for Lemma 4.3.

We can choose a basis transformation  $T = \begin{bmatrix} \alpha_1 & \beta_1 \\ \alpha_2 & \beta_2 \end{bmatrix}$  such that  $(y_0, y_1, y_2)^T = T^{\otimes 2}(1, 0, 0, 1)^T$ . Then under the same  $T$ ,  $[x, y, -x, -y]$  is transformed to  $[u_0, u_1, u_2, u_3] = [x, y, -x, -y]T^{\otimes 3}$ . Since  $[x, y, -x, -y]$  as a linear recurrence has the characteristic equation  $\lambda^2 + 1$ , we have

$$[x, y, -x, -y] = c(1, i)^{\otimes 3} + d(1, -i)^{\otimes 3},$$

and hence

$$[u_0, u_1, u_2, u_3] = [x, y, -x, -y]T^{\otimes 3} = c(\alpha_1 + \alpha_2 i, \beta_1 + \beta_2 i)^{\otimes 3} + d(\alpha_1 - \alpha_2 i, \beta_1 - \beta_2 i)^{\otimes 3}.$$

Since we assumed that  $[x, y, -x, -y]$  is nondegenerate, we have  $cd \neq 0$ .

By this holographic reduction, we only need to prove that  $\text{Holant}^*([u_0, u_1, u_2, u_3])$  is  $\#P$ -hard. Since  $[u_0, u_1, u_2, u_3]$  is nondegenerate, by Lemma 4.2, we have to consider two cases.

If  $[u_0, u_1, u_2, u_3]$  is of the form  $[x', y', -x', -y']$ , then by Proposition 3.6

$$0 = (\alpha_1 + \alpha_2 i)^2 + (\beta_1 + \beta_2 i)^2 = (\alpha_1 - \alpha_2 i)^2 + (\beta_1 - \beta_2 i)^2.$$

This gives  $\alpha_1 \alpha_2 + \beta_1 \beta_2 = 0$  and  $\alpha_1^2 + \beta_1^2 = \alpha_2^2 + \beta_2^2$ . Then by  $(y_0, y_1, y_2)^T = T^{\otimes 2}(1, 0, 0, 1)^T$ , we know that  $[y_0, y_1, y_2]$  is of the form  $[\lambda, 0, \lambda]$ , a contradiction.

Now we consider the second exceptional case from Lemma 4.2: There exist two constants  $a, b$  (not both zero) such that  $au_0 + bu_1 - au_2 = 0$  and  $au_1 + bu_2 - au_3 = 0$ . By Proposition 3.5 we have

$$0 = (\alpha_1 + \alpha_2 i)(\alpha_1 - \alpha_2 i) + (\beta_1 + \beta_2 i)(\beta_1 - \beta_2 i) = \alpha_1^2 + \alpha_2^2 + \beta_1^2 + \beta_2^2.$$

This gives  $y_0 + y_2 = \alpha_1^2 + \beta_1^2 + \alpha_2^2 + \beta_2^2 = 0$ , a contradiction.  $\square$

**4.4. A single signature of arity  $n$ .** The following lemma extends the result to a signature with an arbitrary arity.

LEMMA 4.5. *Let  $[x_0, x_1, x_2, \dots, x_n]$  be a nondegenerate symmetric signature with arity  $n > 3$ ; then  $\text{Holant}^*([x_0, x_1, x_2, \dots, x_n])$  is  $\#P$ -hard unless one of the following two statements is true: (1) there exist two constants  $a, b$  (not both zero), such that for all  $k = 0, 1, \dots, n - 2$ , we have  $ax_k + bx_{k+1} - ax_{k+2} = 0$  (the pair  $(a, b)$  is unique up to a scalar factor); (2) for all  $k = 0, 1, \dots, n - 2$ , we have  $x_k + x_{k+2} = 0$ .*

*Proof.* We assume that  $\text{Holant}^*([x_0, x_1, x_2, \dots, x_n])$  is not  $\#P$ -hard. Then we prove that it must satisfy either statement (1) or (2) in Lemma 4.5.

**Case A:** We first consider the case where for every  $k = 0, 1, \dots, n - 2$ , the subsignature  $[x_k, x_{k+1}, x_{k+2}]$  is nondegenerate. Then by Lemma 4.2, there are the following two exceptional cases (1) and (2) to be considered for  $[x_0, x_1, x_2, x_3]$ .

(1) There is a nonzero pair  $(a, b)$  such that  $ax_0 + bx_1 - ax_2 = 0$  and  $ax_1 + bx_2 - ax_3 = 0$ . Such a nonzero pair  $(a, b)$  must be unique up to a scalar factor, since otherwise  $[x_0, x_1, x_2, x_3]$  is all 0. Then  $\text{Holant}^*([x_0, x_1, x_2, x_3], [x_k, x_{k+1}, x_{k+2}]) \leq_P^P$

Holant<sup>\*</sup>  $([x_0, x_1, x_2, \dots, x_n])$  is not #P-hard, and by Lemma 4.3, for every  $k = 0, 1, \dots, n - 2$ ,  $ax_k + bx_{k+1} - ax_{k+2} = 0$  or  $[x_k, x_{k+1}, x_{k+2}]$  is of the form  $[2a\lambda, b\lambda, -2a\lambda]$ .

First we claim  $a \neq 0$ . Suppose otherwise  $a = 0$ , then  $b \neq 0$ , and  $x_1 = x_2 = 0$ , from the linear equations. But then surely  $x_3 \neq 0$ , since  $[x_0, x_1, x_2, x_3]$  is nondegenerate. Consider when  $k = 2$ . If  $ax_2 + bx_3 - ax_4 = 0$ , then we have a contradiction. Hence  $[x_2, x_3, x_4]$  is of the form  $[2a\lambda, b\lambda, -2a\lambda]$ . But then  $x_2 = x_4 = 0$ . It follows that  $[x_1, x_2, x_3, x_4] = [0, 0, x_3, 0]$ . Applying Lemma 4.2 to this signature, it follows that Holant<sup>\*</sup>  $([x_1, x_2, x_3, x_4])$  is #P-hard, contrary to assumption.

So we can assume  $a = 1$ . Now we show that the form  $[2\lambda, b\lambda, -2\lambda]$  cannot appear among all  $[x_k, x_{k+1}, x_{k+2}]$ . This will conclude that statement (1) of Lemma 4.5 holds.

Suppose the form  $[2\lambda, b\lambda, -2\lambda]$  does occur. Such a  $\lambda$  certainly is nonzero; otherwise it is degenerate.

If it occurs as  $[x_k, x_{k+1}, x_{k+2}]$  for some  $k$ , then  $b \neq \pm 2i$ , since  $[2, \pm 2i, -2]$  is degenerate. If it occurs as  $[x_k, x_{k+1}, x_{k+2}]$  for  $k = 0$  or  $1$ , then together with  $ax_0 + bx_1 - ax_2 = 0$  and  $ax_1 + bx_2 - ax_3 = 0$ , we have  $b = \pm 2i$ . A contradiction. Thus, if this form  $\lambda[2, b, -2]$  appears as some  $[x_k, x_{k+1}, x_{k+2}]$ , and consider the minimal  $k$  where it appears, then  $k \geq 2$ , and there is a subsignature  $[x_{k-2}, x_{k-1}, x_k, x_{k+1}, x_{k+2}] = \lambda[x, y, 2, b, -2]$  with arity 4, where  $ax + by - 2a = 0$  and  $ay + 2b - ab = 0$ . So the subsignature is  $\lambda[b^2 + 2, -b, 2, b, -2]$ , of which there is a subsignature  $\lambda[-b, 2, b, -2]$ . Applying Lemma 4.4 to the nondegenerate signature pair  $[-b, 2, b, -2]$  and  $[b^2 + 2, -b, 2, b, -2]$ , we have  $b^2 + 2 = -2$  or  $b^2 + 2 = 2$ . The first case would imply that  $[2, b, -2]$  is degenerate. So we have  $b^2 + 2 = 2$  and thus  $b = 0$ . In this case, the subsignature is  $[2, 0, 2, 0, -2]$ . Using a unary signature  $[1, 1]$ , we can get  $[2, 2, 2, -2]$ . By Lemma 4.2, this is #P-hard, again contrary to assumption.

(2) If  $[x_0, x_1, x_2, x_3]$  is of the form  $[x, y, -x, -y]$ , then by Lemma 4.4, for every  $k = 0, 1, \dots, n - 2$ , we have  $x_k + x_{k+2} = 0$  or  $[x_k, x_{k+1}, x_{k+2}]$  is of the form  $[\lambda, 0, \lambda]$ . We prove that  $[\lambda, 0, \lambda]$  cannot appear. This will conclude that statement (2) of Lemma 4.5 holds.

Suppose the form  $[\lambda, 0, \lambda]$  does appear among  $[x_k, x_{k+1}, x_{k+2}]$ . It is easy to see that if it occurred at  $k = 0$  or  $1$ , then  $[x_0, x_1, x_2, x_3] = [x, y, -x, -y]$  is degenerate. Then look at the first  $k \geq 2$  where it occurred, and to its left we have  $x_\ell + x_{\ell+2} = 0$ , for  $0 \leq \ell < k$ . It follows that there must be a subsignature (after a scaling) of the form  $[1, 0, -1, 0, -1]$ . Now we can use a similar trick with a unary signature  $[1, 1]$  and get  $[1, -1, -1, -1]$ , which is #P-hard by Lemma 4.2.

This completes the proof of Case A.

**Case B:** There exists some  $0 \leq k \leq n - 2$ , such that the subsignature  $[x_k, x_{k+1}, x_{k+2}]$  is degenerate.

If the signature is of the form  $[x_0, 0, 0, \dots, 0, x_n]$ , we can choose  $(a, b) = (0, 1)$ , and statement (1) of Lemma 4.5 holds. In the following we assume the signature is not of this form.

By Proposition 3.3, there exists an  $s$ ,  $0 \leq s \leq n - 3$ , such that a subsignature  $\sigma_3 = [x_s, x_{s+1}, x_{s+2}, x_{s+3}]$  has arity three and is nondegenerate. We want to find a nondegenerate subsignature of arity three that contains a *degenerate* subsignature of arity two. Starting from a degenerate subsignature  $\tau_2 = [x_k, x_{k+1}, x_{k+2}]$  of arity two, if  $s \leq k \leq s + 1$ , then  $\tau_2$  is already a subsignature of  $\sigma_3$ . Otherwise,  $k < s$  or  $k > s + 1$ . Suppose  $k < s$ . Consider the subsignature  $[x_{k+1}, x_{k+2}, x_{k+3}]$ . If it is degenerate, we can substitute it for  $\tau_2$  and continue. If it is nondegenerate, then we can substitute  $\sigma_3$  by  $[x_k, x_{k+1}, x_{k+2}, x_{k+3}]$ , and it will also be nondegenerate. This is similar for  $k > s + 1$ .

Thus we can find a degenerate subsignature of arity two that is a subsignature of a nondegenerate subsignature  $\sigma_3$  of arity three. This  $\sigma_3$  must be of the form  $[s^2, sr, r^2, x]$  or  $[y, s^2, sr, r^2]$ . By symmetry, we only consider  $[s^2, sr, r^2, x]$ . By Lemma 4.2, we have two cases.

For the first case,  $\det \begin{bmatrix} r^2-s^2 & sr \\ x-sr & r^2 \end{bmatrix} = (r^2 - s^2)r^2 - (x - sr)sr = 0$ . This implies that  $sr x = r^4$ . If  $sr \neq 0$ ,  $sr x = r^4$  implies that  $[s^2, sr, r^2, x]$  is degenerate, a contradiction. If  $sr = 0$ , then  $r = 0$  by  $sr x = r^4$ . Since  $[s^2, sr, r^2, x] = [s^2, 0, 0, x]$  is nondegenerate, we must have  $s^2 \neq 0$  and  $x \neq 0$ . Because  $n > 3$ , there must be entries to its left or to its right, say,  $[s^2, 0, 0, x, z]$ . Consider the pair  $[s^2, 0, 0, x]$  and  $[0, x, z]$ . By Lemma 4.3, the (up to a scalar multiple) unique pair for  $[s^2, 0, 0, x]$  is  $(a, b) = (0, 1)$ . If  $z \neq 0$ , then Lemma 4.3 would imply #P-hardness. If  $z = 0$ , then we have a subsignature  $[0, 0, x, 0]$ . This is also #P-hard by Lemma 4.2.

Finally for the other case,  $s^2 + r^2 = 0$  and  $sr + x = 0$ . Then the signature must be  $[s^2, s^2i, -s^2, -s^2i]$  or  $[s^2, -s^2i, -s^2, s^2i]$ . Both are degenerate, a contradiction.  $\square$

**4.5. Proof of Theorem 4.1.** Finally we extend these lemmas to a set of signatures and finish the proof of Theorem 4.1.

Before we give the proof of Theorem 4.1, we make some definitions and observations. We say a signature  $[x_0, x_1, \dots, x_n]$ , where  $n \geq 2$ , is of type I, if there exist two constants  $a$  and  $b$  (not both zero), such that for every  $k = 0, 1, \dots, n - 2$ , we have  $ax_k + bx_{k+1} - ax_{k+2} = 0$ . If we want to specify the parameters  $a$  and  $b$ , we say it is of type  $I(a, b)$ . We say it is of type II, if for every  $k = 0, 1, \dots, n - 2$ , we have  $x_k + x_{k+2} = 0$ . We say a signature  $[x_0, x_1, x_2]$  of arity two is of type  $I^*$ , more precisely type  $I^*(a, b)$ , if it is of the form  $\lambda[2a, b, -2a]$ . Finally we say a signature  $[x_0, x_1, x_2]$  of arity two is of type  $II^*$ , if it is of the form  $\lambda[1, 0, 1]$ .

For any nondegenerate signature of arity at least three, if it is of type  $I(a, b)$ , then this  $(a, b)$  is uniquely determined up to a nonzero constant multiple. We consider such  $(a, b)$ 's are the same; when we say two different pairs  $(a, b)$  and  $(a', b')$ , we mean they are not constant multiples of each other (distinct points on the projective line  $\mathbb{P}\mathbb{C}^1$ ). In particular a nondegenerate signature of arity at least three cannot be of type  $I(a, b)$  and  $I(a', b')$  for two different pairs  $(a, b)$  and  $(a', b')$ . Similarly, any nondegenerate signature of arity at least three cannot be of both type I and type II. These can be directly verified. Also we can verify that any nondegenerate signature of arity two cannot be of both type  $I(a, b)$  and type  $I^*(a, b)$ , cannot be of both type II and type  $II^*$ , and cannot be of both type  $I^*$  and type  $II^*$ .

Now we rephrase the statements of Lemma 4.2 to Lemma 4.5 as follows, where every signature mentioned is nondegenerate. Lemma 4.2 states that, if  $\text{Holant}^*([x_0, x_1, x_2, x_3])$  is not #P-hard, then  $[x_0, x_1, x_2, x_3]$  is either of type I or of type II. Lemma 4.5 proves the same result as Lemma 4.2 for  $\text{Holant}^*([x_0, x_1, x_2, \dots, x_n])$  with arity  $n > 3$ . Lemma 4.3 states that, if  $[x_0, x_1, x_2, x_3]$  is of type  $I(a, b)$  and  $\text{Holant}^*([x_0, x_1, x_2, x_3], [y_0, y_1, y_2])$  is not #P-hard, then  $[y_0, y_1, y_2]$  is either of type  $I(a, b)$  or of type  $I^*(a, b)$ . Lemma 4.4 states that, if  $[x_0, x_1, x_2, x_3]$  is of type II and  $\text{Holant}^*([x_0, x_1, x_2, x_3], [y_0, y_1, y_2])$  is not #P-hard, then  $[y_0, y_1, y_2]$  is either of type II or of type  $II^*$ .

Now we prove Theorem 4.1. The tractability part of the theorem mainly follows from our previous work on Fibonacci gates [13, 12]. However, we will give an alternative proof. The main result here is the hardness proof, which applies Lemma 4.2 to Lemma 4.5.

*Proof.* Since we consider  $\text{Holant}^*$  problems, every degenerate signature can be decomposed into a tensor product of unary signatures, and will be absorbed in the

Holant\* notation. Thus we will assume every explicitly given signature in  $\mathcal{F}$  is non-degenerate and has arity at least 2.

We first consider tractability. Class 1 is when every signature in  $\mathcal{F}$  has arity at most two. Then the graph of the signature grid  $\Omega$  is a disjoint union of paths and cycles (isolated points contribute a constant; we may assume there are no isolated points). By matrix multiplication, we can compute the value for a path. The value for a cycle is the trace of that path. The value  $\text{Holant}^*_\Omega$  is the product over connected components.

Next we consider class 2. We have type  $I(a, b)$  or  $I^*(a, b)$ . If  $a = 0$ , all functions in  $\mathcal{F}$  have the form  $[x_0, 0, \dots, 0, x_n]$  or  $[0, x, 0]$ . Then  $\text{Holant}^*(\mathcal{F})$  is obviously computable in polynomial time, by extending uniquely any initial value at an edge within a connected component. Now suppose  $a \neq 0$ . We first assume the characteristic equation  $a + b\lambda - a\lambda^2 = 0$  has two distinct eigenvalues  $\lambda_1 \neq \lambda_2$ . Note that  $\lambda_1\lambda_2 = -1$ . In this case all functions in  $\mathcal{F}$  have the form  $F = u[\lambda_1]^{1 \otimes n} + v[\lambda_2]^{1 \otimes n}$  or  $G = [2a, b, -2a]$ . We use the nonsingular matrix  $M = \begin{bmatrix} 1 & \lambda_1 \\ 1 & \lambda_2 \end{bmatrix}$  to perform a holographic reduction on  $\#(\{=2\} \cup \mathcal{U}) | (\mathcal{F} \cup \mathcal{U})$ , which is the bipartite form of problem  $\text{Holant}^*(\mathcal{F})$ . Under this basis,  $F$  is turned into  $M^{\otimes n}F = u \begin{bmatrix} 1 + \lambda_1^2 \\ 0 \end{bmatrix}^{\otimes n} + v \begin{bmatrix} 0 \\ 1 + \lambda_2^2 \end{bmatrix}^{\otimes n}$ ,  $G$  is turned into  $M^{\otimes 2}G = [0, (4a^2 + b^2)/a, 0]$ , and  $(=2) = [1, 0, 1]$  is turned into  $[1, 0, 1](M^{-1})^{\otimes 2}$ , which is a nonzero scalar multiple of  $[1 + \lambda_2^2, 0, 1 + \lambda_1^2]$ . In the meanwhile all unary functions are transformed into unary functions. Connecting any unary function to a function of the type  $M^{\otimes n}F$  maintains that type (with arity  $n - 1$ ). Connecting any unary function to any function of arity two, including  $M^{\otimes 2}G$  or  $[1, 0, 1](M^{-1})^{\otimes 2}$  creates another unary function. Hence it follows that this reduction reduces  $\text{Holant}^*(\mathcal{F})$  to an obviously tractable problem.

Continuing for class 2, if the eigenvalues  $\lambda_1 = \lambda_2$ , then  $b = \pm 2ai$ . Then  $a \neq 0$  and we can normalize to  $a = 1, b = \pm 2i, \lambda_1 = \lambda_2 = b/2$ . The signature  $G = [2, b, -2] = [2, \pm 2i, -2]$  is degenerate. Therefore we only need to consider signatures  $F \in \mathcal{F}$  of type  $I(1, b)$ , together with unary signatures. Connecting any unary signature to any  $F$  of type  $I(1, b)$  creates another signature of type  $I(1, b)$  (of lower arity).

Suppose  $b = 2i$  (the case with  $-2i$  is similar). A signature  $X = [x_0, \dots, x_n] \in \mathcal{F}$  of type  $I(1, 2i)$  satisfies  $x_k = ski^{k-1} + ti^k, 0 \leq k \leq n$ , where  $s$  and  $t$  are two constants. Let  $c = t - \frac{sn}{2}$  and  $d = \frac{si}{2}$ . Consider the signature

$$X' = c \begin{bmatrix} 1 \\ 0 \end{bmatrix}^{\otimes n} + d \left[ \begin{bmatrix} 1 \\ 0 \end{bmatrix}^{\otimes(n-1)} \otimes \begin{bmatrix} 0 \\ 1 \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \end{bmatrix}^{\otimes(n-2)} \otimes \begin{bmatrix} 0 \\ 1 \end{bmatrix} \otimes \begin{bmatrix} 1 \\ 0 \end{bmatrix} + \dots + \begin{bmatrix} 0 \\ 1 \end{bmatrix} \otimes \begin{bmatrix} 1 \\ 0 \end{bmatrix}^{\otimes(n-1)} \right].$$

We claim that

$$Z^{\otimes n}X' = X, \quad \text{where} \quad Z = \begin{bmatrix} 1 & 1 \\ i & -i \end{bmatrix}.$$

To wit, the weight  $k$  entry of  $Z^{\otimes n}X'$  is  $ci^k + d[k(-i)i^{k-1} + (n - k)i^k] = i^k[c + d(n - 2k)] = ski^{k-1} + ti^k$ . Under the holographic transformation by  $Z, (=2)$  is transformed to  $Z^T I_2 Z = 2 \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ , which is 2 times the binary DISEQUALITY ( $\neq_2$ ) function in matrix form. The bipartite problem  $\#\{\neq_2\} | \{X'\}$ , where each  $X'$  may have different parameters  $c$  and  $d$ , is tractable by the following argument. In any nonzero evaluation, on account of ( $\neq_2$ ), an assignment must assign exactly half the edges 0 and the other half 1. For any  $X'$  with arity more than two, to be nonzero it must have strictly more incident edges 0. Hence the computation reduces to that of class 1, with arities at most 2.

For class 3, we will prove that the Holant is zero unless the input graph is bipartite, and there is a holographic algorithm for bipartite graphs. If there is a function  $[1, 0, 1]$  applied to two variables  $x$  and  $y$ , we just merge them into one variable and remove this function. Connecting a unary function to a function of type II still gives a function of type II. Hence, we may assume all functions on vertices of an input graph are of type II, i.e., have the form  $[u, v, -u, -v, \dots]$ . If the input graph  $G(V, E)$  is not bipartite, there is a cycle  $v_1, e_1, v_2, \dots, v_k, e_k, v_1$  of odd length. We partition all 0-1 assignments for  $E$  into two parts, with a 1-1 correspondence between them. An assignment  $\sigma$  is mapped to  $\sigma'$ , which assigns the same values on  $E - \{e_1, \dots, e_k\}$ , but opposite values on  $\{e_1, \dots, e_k\}$ . Under  $\sigma$  and  $\sigma'$ , all functions on  $V - \{v_1, \dots, v_k\}$  evaluate to the same value, and if  $\sigma(e_{j-1}) \neq \sigma(e_j)$  (where the index  $j$  is counted mod  $k$ ), the function at  $v_j$  evaluates to the same value under  $\sigma$  and  $\sigma'$ , and if  $\sigma(e_{j-1}) = \sigma(e_j)$ , the function at  $v_j$  evaluates to the opposite values under  $\sigma$  and  $\sigma'$ . Consider  $\sigma(e_1), \sigma(e_2), \dots, \sigma(e_k), \sigma(e_1)$ , there must be an even number of times where the value changes in this sequence. Since  $k$  is odd, there are an odd number of  $v_j$ 's, where the functions at  $v_j$  give opposite values under  $\sigma$  and  $\sigma'$ . Hence, in the summation, the contributions of  $\sigma$  and  $\sigma'$  are canceled.

If the input graph is bipartite, the problem is  $\#(\mathcal{F} \cup \mathcal{U}) | (\mathcal{F} \cup \mathcal{U})$ . Then  $\#(\mathcal{F} \cup \mathcal{U}) | (\mathcal{F} \cup \mathcal{U})$  is turned into  $\#(\mathcal{F}' \cup \mathcal{U}) | (\mathcal{F}' \cup \mathcal{U})$ , where every signature in  $\mathcal{F}'$  has the form  $[x_1, 0, \dots, 0, x_n]$ , by the holographic reduction with matrix  $M = \begin{bmatrix} 1 & \\ & -i \end{bmatrix}$ . This is clearly tractable.

Now we prove that if  $\text{Holant}^*(\mathcal{F})$  is *not*  $\#P$ -hard, then  $\mathcal{F}$  must fall in one of the three classes.

As stated earlier, all signatures in  $\mathcal{F}$  are assumed to be nondegenerate. If no signature has arity  $\geq 3$ , then we are done with class 1. Suppose there are some signatures of arity  $\geq 3$ . Each such signature is of type I or II by Lemma 4.2 and Lemma 4.5. If there is exactly one signature of arity  $\geq 3$ , then if it is of type I, Lemma 4.3 finishes the proof and we are in class 2; if it is of type II, Lemma 4.4 finishes the proof and we are in class 3. So we are finished.

In the following we assume there are more than one signatures in  $\mathcal{F}$  of arity  $\geq 3$ . Of course, each such signature declares a type I or II. As remarked before, any signature in  $\mathcal{F}$ , being nondegenerate of arity  $\geq 3$ , declares a unique type I or II, not both; and in case of type I, it is of type  $I(a, b)$ , for a unique  $(a, b) \in \text{PC}^1$ . We would like to prove that these declarations are all “consistent”; i.e., there is only one type for all nondegenerate signatures of arity at least three.

- Case 1. Suppose there is a nondegenerate type  $I(0, 1)$  signature  $\sigma_3$  of arity  $\geq 3$ . By connecting with the unary  $[1, 1]$  we may assume it is of arity three. Then for any other nondegenerate signature  $\sigma'$  of arity  $\geq 3$  and of a different type, if one exists, by Proposition 3.3, there exists a nondegenerate subsignature  $[x, y, z]$  of arity two. By being *not*  $\#P$ -hard, we apply Lemma 4.3 to the pair  $\sigma_3$  and  $[x, y, z]$  and conclude that there are two cases:  $[x, y, z] = [x, 0, z]$ , or  $[0, y, 0]$ . Being nondegenerate,  $xz \neq 0$ , or  $y \neq 0$ , respectively. Within  $\sigma'$  there is a subsignature containing it. In the first case, it has the form  $[w, x, 0, z]$  (or  $[x, 0, z, w]$ ), then  $[w, x, 0]$  (or  $[0, z, w]$ ) is nondegenerate, and applying Lemma 4.3 once again, we get  $w = 0$ . In the second case, it has the form  $[u, 0, y, 0]$  (or  $[0, y, 0, u]$ ). If  $u = 0$ ,  $\text{Holant}^*([0, 0, y, 0])$  (or  $\text{Holant}^*([0, y, 0, 0])$ ) is  $\#P$ -hard by Lemma 4.2. Hence in all cases we get a signature of the form  $[u, 0, v, 0]$  (or its reversal  $[0, v, 0, u]$ ) where  $uv \neq 0$ . By symmetry consider  $[u, 0, v, 0]$  (the case  $[0, v, 0, u]$  is similar). We connect a unary  $[s, t]$  to it, and get  $[su, tv, sv]$ . As long as  $st \neq 0$  and  $s^2/t^2 \neq v/u$ , this is nondegenerate and Lemma 4.3 on the pair  $\sigma_3$  and  $[su, tv, sv]$  leads to  $\#P$ -hardness.

• Case 2. Now we assume Case 1 does not apply, but there exists a nondegenerate type I( $a, b$ ) signature  $\sigma_3$  of arity  $\geq 3$ , where  $a \neq 0$  and  $b \neq \pm 2ai$ . By Proposition 3.3, we may assume  $\sigma_3$  has arity three. We may normalize it to  $a = 1$ , and  $b \neq \pm 2i$ .

Then the characteristic equation  $\lambda^2 - b\lambda - 1$  has two distinct eigenvalues  $\lambda_1 \neq \lambda_2$ . It follows that  $\lambda_1\lambda_2 = -1$ , and thus  $\lambda_1, \lambda_2 \neq \pm i$ . Define

$$T = \begin{bmatrix} \frac{1}{\sqrt{1+\lambda_1^2}} & \frac{1}{\sqrt{1+\lambda_2^2}} \\ \frac{\lambda_1}{\sqrt{1+\lambda_1^2}} & \frac{\lambda_2}{\sqrt{1+\lambda_2^2}} \end{bmatrix},$$

then  $T$  is orthogonal. It is easy to verify that  $\sigma_3 = T^{\otimes 3}f$  for some  $f = [x_0, 0, 0, x_3]^T$ , a nondegenerate signature of type I(0, 1). Performing a holographic transformation by  $T$  on all signatures in  $\mathcal{F}$ , we have reduced Case 2 to Case 1.

• Case 3. We assume Cases 1 and 2 do not apply, but there exists a nondegenerate type I(1,  $b$ ) signature  $\sigma_3$  of arity  $\geq 3$ , where  $b = \pm 2i$ . By Proposition 3.3, we may assume  $\sigma_3$  has arity three. We still suppose there exists a nondegenerate signature  $\sigma'$  of arity  $\geq 3$  and of a different type, for otherwise we are done. By Proposition 3.3, there exists a nondegenerate subsignature  $[x, y, z]$  of  $\sigma'$ .  $\sigma'$  could be of type I(1,  $b'$ ) or II, where  $b' = \pm 2i$  (but  $b' \neq b$ , and hence  $b' = \bar{b}$ ).

If it is of type I(1,  $b'$ ), note that for  $b = \pm 2i$ , the binary signature  $[2, b, -2]$  is degenerate, and so, by applying Lemma 4.3 to  $\sigma_3$  and  $[x, y, z]$ , we get  $x + by - z = 0$  as well as  $x + b'y - z = 0$ . This implies that  $y = 0$  and  $x = z \neq 0$ . But within  $\sigma'$  there is a subsignature  $[w, x, 0, x]$  or  $[x, 0, x, w]$ . As  $x \neq 0$ ,  $[w, x, 0]$  and  $[0, x, w]$  are nondegenerate. Applying Lemma 4.3 once again we get a contradiction. If  $\sigma'$  is of type II, then we still get  $x + by - z = 0$  as well as  $z = -x$ . Since  $b = \pm 2i$ , this implies that  $[x, y, z]$  is degenerate, also a contradiction.

• Case 4. There are no nondegenerate signatures of type I and arity  $\geq 3$ . Then all of them must be of type II. This concludes the proof.  $\square$

### 5. A dichotomy theorem for $\text{Holant}^c(\mathcal{F})$ .

**THEOREM 5.1.** *Let  $\mathcal{F}$  be a set of real symmetric signatures, and let  $\mathcal{F}_1, \mathcal{F}_2$ , and  $\mathcal{F}_3$  be three families of signatures defined as*

$$\begin{aligned} \mathcal{F}_1 &= \{\lambda([1, 0]^{\otimes k} + i^r[0, 1]^{\otimes k}) \mid \lambda \in \mathbb{C}, k = 1, 2, \dots, \text{ and } r = 0, 1, 2, 3\}; \\ \mathcal{F}_2 &= \{\lambda([1, 1]^{\otimes k} + i^r[1, -1]^{\otimes k}) \mid \lambda \in \mathbb{C}, k = 1, 2, \dots, \text{ and } r = 0, 1, 2, 3\}; \\ \mathcal{F}_3 &= \{\lambda([1, i]^{\otimes k} + i^r[1, -i]^{\otimes k}) \mid \lambda \in \mathbb{C}, k = 1, 2, \dots, \text{ and } r = 0, 1, 2, 3\}. \end{aligned}$$

*Then  $\text{Holant}^c(\mathcal{F})$  is computable in polynomial time if (1)  $\text{Holant}^*(\mathcal{F})$  is computable in polynomial time or (2)  $\mathcal{F} \subseteq \mathcal{F}_1 \cup \mathcal{F}_2 \cup \mathcal{F}_3$ . Otherwise,  $\text{Holant}^c(\mathcal{F})$  is  $\#P$ -hard.*

*Tractability and proof outline for hardness.* By definition, every instance of  $\text{Holant}^c(\mathcal{F})$  is also an instance of  $\text{Holant}^*(\mathcal{F})$ . So it is obvious that if  $\text{Holant}^*(\mathcal{F})$  is computable in polynomial time, then so is  $\text{Holant}^c(\mathcal{F})$ . The polynomial time algorithm for any complex-valued function set  $\mathcal{F} \subseteq \mathcal{F}_1 \cup \mathcal{F}_2 \cup \mathcal{F}_3$  is nontrivial. The tractability crucially depends on algebraic cancellations. The crux of the matter is a polynomial time algorithm to evaluate

$$\sum_{x_1, x_2, \dots, x_n=0,1} i^{L_1(X)+L_2(X)+\dots+L_m(X)},$$

where each  $L_k(X)$  is a 0-1 indicator function of an affine linear function mod 2. Thus the exponent on  $i$  can be viewed as a mod 4 sum of mod 2 sums. It turns out that this tractability leads to a more general polynomial time algorithm that includes asymmetric signatures as well. In [11, 14], we extend these families to a unified algorithm, which eventually leads to a dichotomy theorem for complex-valued Boolean #CSP.

The tractability for Holant<sup>c</sup>( $\mathcal{F}_1 \cup \mathcal{F}_2 \cup \mathcal{F}_3$ ) works for complex-valued signatures. In this paper we give a proof for the case of real-valued signatures, which suffices for Theorem 5.1.

We explicitly list all the signatures in  $\mathcal{F}_1 \cup \mathcal{F}_2 \cup \mathcal{F}_3$ , up to an arbitrary constant multiple from  $\mathbb{C}$ .

1.  $[1, 0, 0, \dots, 0, \pm 1]$ ; ( $\mathcal{F}_1, r = 0, 2$ )
2.  $[1, 0, 0, \dots, 0, \pm i]$ ; ( $\mathcal{F}_1, r = 1, 3$ )
3.  $[1, 0, 1, 0, \dots, 0 \text{ or } 1]$ ; ( $\mathcal{F}_2, r = 0$ )
4.  $[0, 1, 0, 1, \dots, 0 \text{ or } 1]$ ; ( $\mathcal{F}_2, r = 2$ )
5.  $[1, i, 1, i, \dots, i \text{ or } 1]$ ; ( $\mathcal{F}_2, r = 3$ )
6.  $[1, -i, 1, -i, \dots, (-i) \text{ or } 1]$ ; ( $\mathcal{F}_2, r = 1$ )
7.  $[1, 0, -1, 0, 1, 0, -1, 0, \dots, 0 \text{ or } 1 \text{ or } (-1)]$ ; ( $\mathcal{F}_3, r = 0$ )
8.  $[1, 1, -1, -1, 1, 1, -1, -1, \dots, 1 \text{ or } (-1)]$ ; ( $\mathcal{F}_3, r = 1$ )
9.  $[0, 1, 0, -1, 0, 1, 0, -1, \dots, 0 \text{ or } 1 \text{ or } (-1)]$ ; ( $\mathcal{F}_3, r = 2$ )
10.  $[1, -1, -1, 1, 1, -1, -1, 1, \dots, 1 \text{ or } (-1)]$ . ( $\mathcal{F}_3, r = 3$ )

Among these, we single out the real-valued signatures, up to an arbitrary constant multiple from  $\mathbb{R}$ .

1.  $[1, 0, 0, \dots, 0, \pm 1]$ ; ( $\mathcal{F}_1$ )
2.  $[1, 0, 1, 0, \dots, 0 \text{ or } 1]$ ; ( $\mathcal{F}_2$ )
3.  $[0, 1, 0, 1, \dots, 0 \text{ or } 1]$ ; ( $\mathcal{F}_2$ )
4.  $[1, 0, -1, 0, 1, 0, -1, 0, \dots, 0 \text{ or } 1 \text{ or } (-1)]$ ; ( $\mathcal{F}_3$ )
5.  $[1, 1, -1, -1, 1, 1, -1, -1, \dots, 1 \text{ or } (-1)]$ ; ( $\mathcal{F}_3$ )
6.  $[0, 1, 0, -1, 0, 1, 0, -1, \dots, 0 \text{ or } 1 \text{ or } (-1)]$ ; ( $\mathcal{F}_3$ )
7.  $[1, -1, -1, 1, 1, -1, -1, 1, \dots, 1 \text{ or } (-1)]$ . ( $\mathcal{F}_3$ )

For any occurrence of a constraint function of the form  $[1, 0, 0, \dots, 0, 1]$  of arity  $k$ , we can simply identify the  $k$  variables and eliminate this constraint function. Similarly for  $[1, 0, 0, \dots, 0, -1]$  on  $x_1, \dots, x_k$ , we will identify all  $x_1, \dots, x_k$ , and replace it by  $[1, -1] = (-1)^{x_1}$ . For  $[1, 0, 1, 0, \dots, 0 \text{ or } 1]$  on  $x_1, \dots, x_k$ ,  $k > 1$ , we substitute  $x_k = x_1 + \dots + x_{k-1}$  (a mod 2 sum), and for  $[0, 1, 0, 1, \dots, 0 \text{ or } 1]$  we substitute  $x_k = x_1 + \dots + x_{k-1} + 1$ . For  $k = 1$ , we have  $[1, 0]$  and  $[0, 1]$ , respectively, and the variable is simply set to the constant 0 and 1, respectively.

A more interesting case is  $[1, 1, -1, -1, 1, 1, -1, -1, \dots, 1 \text{ or } (-1)]$  on  $x_1, \dots, x_k$ . We claim this is the same function as  $(-1)^{\sum_{1 \leq i < j \leq k} x_i x_j}$ . On any input of Hamming weight  $t$ , the exponent is  $\binom{t}{2}$ , which by Lucas theorem gives the following values:  $0, 0, 1, 1, 0, 0, 1, 1, \dots \text{ mod } 2$ . Similarly  $[1, -1, -1, 1, 1, -1, -1, \dots, 1 \text{ or } (-1)]$  is  $(-1)^{\sum_{1 \leq i \leq k} x_i + \sum_{1 \leq i < j \leq k} x_i x_j}$ .

Finally,  $[1, 0, -1, 0, 1, 0, -1, 0, \dots, 0 \text{ or } 1 \text{ or } (-1)]$  is the product of

$$[1, 0, 1, 0, \dots, 0 \text{ or } 1] \quad \text{and} \quad [1, 1, -1, -1, 1, 1, -1, -1, \dots, 1 \text{ or } (-1)],$$

and  $[0, 1, 0, -1, 0, 1, 0, -1, \dots, 0 \text{ or } 1 \text{ or } (-1)]$  is the product of

$$[0, 1, 0, 1, \dots, 0 \text{ or } 1] \quad \text{and} \quad [1, 1, -1, -1, 1, 1, -1, -1, \dots, 1 \text{ or } (-1)].$$

The above analysis shows that the tractability of  $\text{Holant}^c(\mathcal{F}_1 \cup \mathcal{F}_2 \cup \mathcal{F}_3)$  follows from a polynomial time algorithm to evaluate the sum

$$\sum_{x_1, x_2, \dots, x_n=0,1} (-1)^{Q(x_1, x_2, \dots, x_n)},$$

where  $Q(x_1, x_2, \dots, x_n)$  is a polynomial over  $\mathbb{Z}_2$  of degree at most 2. This algorithm is known [28].

The main result of Theorem 5.1 is hardness. We want to prove that aside from these tractable cases, all remaining problems are  $\#P$ -hard. Here the main technique is polynomial interpolation. We prove the second dichotomy theorem (Theorem 5.1) by a reduction to the first (Theorem 4.1). We will show how to interpolate all unary signatures. The interpolation method used here is briefly described in subsection 5.1. Once we can interpolate all unary signatures by Lemma 5.3, we can make use of the dichotomy theorem for  $\text{Holant}^*(\mathcal{F})$ . The whole proof is organized as a sequence of lemmas (Lemma 5.5 to Lemma 5.9). In each lemma, we prove the theorem for a broader family of  $\mathcal{F}$ , and the remaining unproved ones are the beginning of the next lemma. Finally we prove the theorem for all possible signature sets  $\mathcal{F}$ . In some cases, the attempt to interpolate all unary signatures does not work. In these cases, we employ another starting point of  $\#P$ -hardness, which is the problem of counting PERFECT MATCHINGS on 3-regular graphs [16]. The reduction from PERFECT MATCHING is also by polynomial interpolation, which is done in Lemma 5.4. However, note that counting PERFECT MATCHINGS is computable in polynomial time for planar graphs [25, 26, 29]; therefore our dichotomy theorem for  $\text{Holant}^c$  problems here does not apply to planar graphs as our dichotomy theorem for  $\text{Holant}^*$  problems does.

**5.1. Polynomial interpolation.** In this subsection, we discuss the interpolation method we use for the dichotomy theorem for  $\text{Holant}^c$ . Polynomial interpolation is a powerful tool in the study of counting problems initiated by Valiant [31] and further developed by Vadhan [30], Dyer and Greenhill [21], and many others. The method we use here is essentially the same as Vadhan [30].

Suppose for some set of signatures  $\mathcal{F}$ , we want to show that for all unary signatures  $f = [x, y]$ , we have  $\text{Holant}(\mathcal{F} \cup \{[x, y]\}) \leq_T^p \text{Holant}(\mathcal{F})$ ; i.e.,  $\text{Holant}(\mathcal{F} \cup \{[x, y]\})$  is polynomial time Turing reducible to  $\text{Holant}(\mathcal{F})$ . Let  $\Omega = (G, \mathcal{F} \cup \{[x, y]\}, \pi)$ . We want to compute  $\text{Holant}_\Omega$  in polynomial time using an oracle for  $\text{Holant}(\mathcal{F})$ .

Let  $V_f$  be the subset of vertices in  $G$  assigned  $f$  in  $\Omega$ . Suppose  $|V_f| = n$ . We can classify all 0-1 assignments  $\sigma$  in the Holant sum according to how many vertices in  $V_f$  whose incident edge is assigned a 0 or a 1. Then the Holant value can be expressed as

$$(5.1) \quad \text{Holant}_\Omega = \sum_{0 \leq i \leq n} c_i x^i y^{n-i},$$

where  $c_i$  is the sum over all edge assignments  $\sigma$ , of products of evaluations at all  $v \in V(G) - V_f$ , where  $\sigma$  is such that exactly  $i$  vertices in  $V_f$  have their incident edges assigned 0 (and  $n - i$  have their incident edges assigned 1). If we can evaluate these  $c_i$ , we can evaluate  $\text{Holant}_\Omega$ .

Now suppose  $\{G_s\}$  is a sequence of  $\mathcal{F}$ -gates, and each  $G_s$  has one dangling edge. Denote the signature of  $G_s$  by  $f_s = [x_s, y_s]$ , for  $s = 0, 1, \dots$ . If we replace each occurrence of  $f$  by  $f_s$  in  $\Omega$ , we get a new signature grid  $\Omega_s$ , which is an instance of



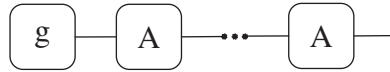


FIG. 5.1. Recursive construction.

Holant( $\mathcal{F}$ ), with

$$(5.2) \quad \text{Holant}_{\Omega_s} = \sum_{0 \leq i \leq n} c_i x_s^i y_s^{n-i}.$$

One can evaluate  $\text{Holant}_{\Omega_s}$  by oracle access to  $\text{Holant}(\mathcal{F})$ . Note that the same set of values  $c_i$  occurs in (5.1) and (5.2). We can treat  $c_i$  in (5.2) as a set of unknowns in a linear system. The idea of interpolation is to find a suitable sequence  $\{f_s\}$  such that the evaluation of  $\text{Holant}_{\Omega_s}$  gives a linear system (5.2) of full rank, from which we can solve all  $c_i$ . In this case, we say that the sequence  $\{G_s\}$  interpolates all unary signatures.

In this paper, the sequence  $\{G_s\}$  will be constructed recursively using suitable gadgetry. There are two gadgets in a recursive construction: one gadget has arity one, giving the initial signature  $g = [x_0, y_0]$ ; the other gadget has arity two, giving the recursive iteration. It is more convenient to use a  $2 \times 2$  matrix  $A$  to denote the signature of this binary gadget. So we can recursively connect them as in Figure 5.1 and get  $\{G_s\}$ .

The signatures of  $\{G_s\}$  have the following relation:

$$(5.3) \quad \begin{bmatrix} x_s \\ y_s \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} x_{s-1} \\ y_{s-1} \end{bmatrix},$$

where  $A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$  and  $g = \begin{bmatrix} x_0 \\ y_0 \end{bmatrix}$ .

We call this gadget pair  $(A, g)$  a recursive construction. Vadhan’s Lemma 6.1 in [30] gives the following lemma.

LEMMA 5.2. *Let  $\alpha, \beta$  be the two eigenvalues of  $A$ . If the following three conditions are satisfied:*

1.  $\det(A) \neq 0$ ;
2.  $g$  is not a column eigenvector of  $A$  (nor the zero vector);
3.  $\alpha/\beta$  is not a root of unity,

*then the recursive construction  $(A, g)$  can interpolate all unary signatures.*

This interpolation method also works in a similar way for signatures with a larger arity but only two dimensions of freedom. For example, it works for all signatures of the form  $[0, x, 0, y]$ . This version is used in the proof of Lemma 5.4.

**5.2. Hardness proof of Theorem 5.1.** In the dichotomy theorem for  $\text{Holant}^*$  ( $\mathcal{F}$ ), we assume the arity of every signature in  $\mathcal{F}$  is greater than one and all signatures in  $\mathcal{F}$  are nondegenerate. In  $\text{Holant}^c(\mathcal{F})$ , not all unary signatures are freely available, so we cannot assume that. In some cases, the presence of some unary signatures or degenerate signatures does change the complexity of the problem. However, we can still do some normalization here to make the proof clearer. Since any degenerate signature  $[x, y]^{\otimes k}$  can be replaced by the corresponding unary signature  $[x, y]$  without changing the complexity of the problem, we always assume that every signature in  $\mathcal{F}$ , whose arity is greater than one, is nondegenerate. Since  $[1, 0]$  and  $[0, 1]$  are freely available, we can construct any subsignature of the original signatures as well as any signature realizable by some  $\mathcal{F}$ -gate.

LEMMA 5.3. *If we can construct a gadget with a (real) signature  $[a, b, c]$ , where  $b^2 \neq ac$ ,  $b \neq 0$ , and  $a + c \neq 0$ , then we can interpolate all unary functions.*

*Proof.* We use the interpolation method as described in subsection 5.1. We consider the recursive construction  $(\begin{bmatrix} a & b \\ b & c \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \end{bmatrix})$  (alternatively  $(\begin{bmatrix} a & b \\ b & c \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix})$ ), show that the construction succeeds in interpolating all unary functions. We use  $A$  to denote  $\begin{bmatrix} a & b \\ b & c \end{bmatrix}$ . Since  $b^2 \neq ac$ ,  $A$  is nondegenerate, the first condition of Lemma 5.2 is satisfied. If  $[1, 0]^T$  is a column eigenvector of  $A$  then  $b = 0$ , a contradiction. So the second condition of Lemma 5.2 is satisfied. Since  $A$  is a real symmetric matrix, both its eigenvalues are real. If the ratio of two real numbers is a root of unity, they must be the same or opposite to each other. If the two eigenvalues are the same, we will have  $b = 0$  and  $a = c$ , a contradiction. If the two eigenvalues are opposite to each other, then the trace  $a + c = 0$ , also a contradiction. Therefore, the third condition of Lemma 5.2 is also satisfied. This completes the proof.  $\square$

If we can construct a gadget with a binary symmetric signature  $[a, b, c]$ , which satisfies all the conditions in Lemma 5.3, then we are done. In most cases, we prove Theorem 5.1 by interpolating all unary signatures. However, in some cases, we are not able to do that. For example, if every signature  $\mathcal{F}$  has the parity condition (i.e., it is zero on all odd weight entries, or zero on all even weight entries), then all unary signatures we can realize from  $\mathcal{F}$  have the form  $[a, 0]$  or  $[0, a]$ . So we cannot interpolate all unary signatures. For these cases, our starting point is the following lemma.

LEMMA 5.4. *If  $a \in \mathbb{R}$  and  $a \neq \pm 1$ , then  $\text{Holant}^c([0, 1, 0, a])$  is  $\#P$ -hard. By symmetry,  $\text{Holant}^c([a, 0, 1, 0])$  is also  $\#P$ -hard.*

*Proof.* We start with  $\text{Holant}([0, 1, 0, 0])$ , which is  $\#P$ -hard. This is the problem of counting perfect matchings in 3-regular graphs [16]. So the problem is  $\#P$ -hard if  $a = 0$ .

Now assume that  $a \notin \{-1, 0, 1\}$ . We will use the signature  $[0, 1, 0, a]$  to interpolate all the signatures of the form  $[0, 1, 0, x]$ ; in particular, we can interpolate  $[0, 1, 0, 0]$  and finish the hardness reduction.

The recursive construction is depicted by Figure 5.2. By a simple parity argument, and an induction on  $i$ , every  $\mathcal{F}$ -gate  $N_i$  has a signature of the form  $[0, x_i, 0, y_i]$ . More precisely, let  $\sigma$  be an assignment to the three dangling edges of  $N_i$ . We show that if the Hamming weight  $\text{wt}(\sigma)$  is even, then the value of  $N_i$  is zero. Each internal edge in  $N_i$ , excluding those internal to  $N_{i-1}$ , appears twice as input to functions at vertices, and hence  $\text{wt}(\sigma)$  is congruent mod 2 to the sum of Hamming weights of assignments at all seven vertices of degree 3 in  $N_i$  (Figure 5.2), where we consider  $N_{i-1}$  as a single vertex of degree 3. To have a nonzero evaluation, each of the seven vertices requires an input of odd Hamming weights. This shows that the signature of  $N_i$  has the form  $[0, x_i, 0, y_i]$ . After some calculation, we can get that they satisfy the following recursive relation:

$$\begin{bmatrix} x_{i+1} \\ y_{i+1} \end{bmatrix} = \begin{bmatrix} 3(a^2 + 1) & (a^3 + a) \\ 3(a^3 + a) & a^6 + 1 \end{bmatrix} \begin{bmatrix} x_i \\ y_i \end{bmatrix}.$$

In this case, the signatures we want to interpolate are of arity three. But since all of them are of the form  $[0, x_i, 0, y_i]$ , we can use the interpolation method in subsection 5.1. Let  $A = \begin{bmatrix} 3(a^2+1) & (a^3+a) \\ 3(a^3+a) & a^6+1 \end{bmatrix}$ , and then  $(A, [1, a]^T)$  forms a recursive construction. Note that here  $[1, a]$  really denotes the initial gadget consisting of a single vertex with signature  $[0, 1, 0, a]$ . Since  $\det(A) = 3(a^4 - 1)^2 \neq 0$ , the first condition holds. Its characteristic equation is  $X^2 - (a^6 + 3a^2 + 4)X + 3(a^4 - 1)^2 = 0$ . For this quadratic equation, its discriminant  $\Delta = (a^6 - 3a^2 - 2)^2 + 12(a + a^3)^2 > 0$ . So  $A$  has two

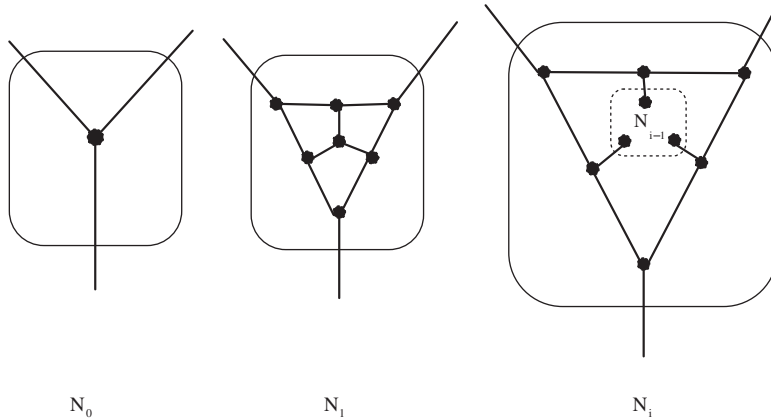


FIG. 5.2. The recursive construction. The signature of every vertex in the gadget is  $[0, 1, 0, a]$ .

distinct real eigenvalues. The sum of the two eigenvalues is  $a^6 + 3a^2 + 4 > 0$ . So they are not opposite to each other. Therefore, the ratio of these two eigenvalues is not a root of unity and the third condition holds. Consider the second condition, if the initial vector  $[1, a]^T$  is a column eigenvector of  $A$ , we have  $A \begin{bmatrix} 1 \\ a \end{bmatrix} = \lambda \begin{bmatrix} 1 \\ a \end{bmatrix}$ , where  $\lambda$  is an eigenvalue of  $A$ . From this, we can conclude that  $a(a^2 - 1)(a^4 - 1) = 0$ , which does not hold given  $a \notin \{-1, 0, 1\}$ . To sum up, this recursive relation satisfies all three conditions of Lemma 5.2 and can be used to interpolate all signatures of the form  $[0, 1, 0, x]$ . This completes the proof.  $\square$

We define some families of real symmetric signatures, which will be used in our proof.

$$\begin{aligned} \mathcal{G}_1 &= \{[a, 0, \dots, 0, b] \mid ab \neq 0, \text{ arity } k \geq 1\}. \\ \mathcal{G}_2 &= \{[x_0, x_1, \dots, x_k] \mid k \geq 1 \quad \forall i \text{ even}, x_i = 0, \text{ or } \forall i \text{ odd}, x_i = 0\}. \\ \mathcal{G}_3 &= \{[x_0, x_1, \dots, x_k] \mid k \geq 1 \quad \forall i, x_i + x_{i+2} = 0\}. \end{aligned}$$

We note that  $\mathcal{G}_1, \mathcal{G}_2$ , and  $\mathcal{G}_3$  are supersets of (the real parts of)  $\mathcal{F}_1, \mathcal{F}_2$ , and  $\mathcal{F}_3$ , respectively (the only exception is that the all zero function is in  $\mathcal{F}_1$ , but not in  $\mathcal{G}_1$ ).

The following several lemmas serve as the main technical argument for the proof of Theorem 5.1. It is a carefully calibrated sequence that is designed to chip away at the difficulty. In an intuitive sense, signature sets within  $\mathcal{G}_1 \cup \mathcal{G}_2 \cup \mathcal{G}_3$  are sufficiently “close” to being tractable for Holant<sup>c</sup>, so that to exactly “carve” out their #P-hardness while keeping  $\mathcal{F}_1 \cup \mathcal{F}_2 \cup \mathcal{F}_3$  in the tractable side (as we must) is a delicate task. All these lemmas have the form “If  $\mathcal{F} \not\subseteq \mathcal{A}$ , then Theorem 5.1 holds.” After proving one lemma, in all subsequent lemmas, we only need to consider the case that  $\mathcal{F} \subseteq \mathcal{A}$ .

LEMMA 5.5. *If  $\mathcal{F} \not\subseteq \mathcal{G}_1 \cup \mathcal{G}_2 \cup \mathcal{G}_3$ , then Theorem 5.1 holds.*

*Proof.* Since  $\mathcal{F} \not\subseteq \mathcal{G}_1 \cup \mathcal{G}_2 \cup \mathcal{G}_3$ , there exists an  $f \in \mathcal{F}$  and  $f \notin \mathcal{G}_1 \cup \mathcal{G}_2 \cup \mathcal{G}_3$ . Since all unary signatures are in  $\mathcal{G}_3$ , the arity of  $f$  is greater than one and  $f$  is nondegenerate. There are two cases according to whether  $f$  has a zero entry or not.

(1)  $f$  has some zero entries. If there exists a subsignature of  $f$  having the form  $[0, a, b]$  or  $[a, b, 0]$ , where  $ab \neq 0$ , then we are done by Lemma 5.3. Otherwise, we can conclude that there are no two successive nonzero entries. So the signature  $f$  takes this form  $[0^{i_0} x_1 0^{i_1} x_2 0^{i_2} \dots x_k 0^{i_k}]$ . Here for all  $1 \leq j \leq k, 0 \neq x_j \in \mathbb{R}; k \geq 2$ , since  $f \notin \mathcal{G}_2; i_0, i_k \geq 0$ , and for all  $1 \leq j \leq k - 1, i_j \geq 1$ . If for all  $1 \leq j \leq k - 1, i_j$  is

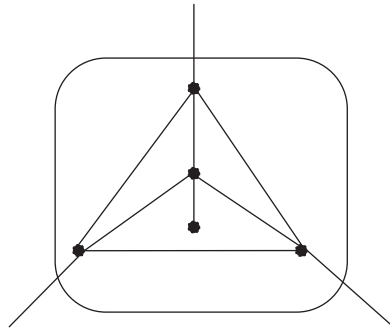


FIG. 5.3. The signature of every degree 4 vertex in the gadget is  $[1, 0, 1, 0, -1]$ . And the signature of the degree 1 vertex in the gadget is  $[1, 0]$ .

odd, then  $f \in \mathcal{G}_2$ , a contradiction. Therefore there exists a subsignature of the form  $[x, 0, \dots, 0, y]$ , where  $xy \neq 0$  and there is an even number of 0's between  $x$  and  $y$ . If this is the entire signature  $f$  itself, then  $f \in \mathcal{G}_1$ , a contradiction. So there is at least one 0 before  $x$  or after  $y$  in  $f$ . By symmetry, we assume there is a 0 before  $x$ , so we have a subsignature  $[0, x, 0, \dots, 0, y]$ , whose arity is even and greater than three. We call its dangling edges  $1, 2, \dots, 2k$ . Then for every  $i = 1, 2, \dots, k - 1$ , we merge dangling edges  $2i + 1$  and  $2i + 2$  together to form a regular edge. After that, we have an  $\mathcal{F}$ -gate with arity two, and its signature is  $[0, x, y]$ . Then we are done by Lemma 5.3.

(2)  $f$  has no zero entries. We only need to prove that we can construct a function  $[a', b', c']$  satisfying the three conditions in Lemma 5.3. Suppose all subsignatures of  $f$  with arity two do not satisfy all three conditions. For each subsignature  $[a', b', c']$ , either  $a' + c' = 0$ , or  $b'^2 = a'c'$ . If all of them satisfy  $a' + c' = 0$ , then  $f \in \mathcal{G}_3$ , and we get a contradiction. If all of them satisfy  $b'^2 = a'c'$ , then  $f$  is degenerate. This is also a contradiction, since the arity of  $f$  is  $> 1$ , and all signatures in  $\mathcal{F}$  of arity  $> 1$  are nondegenerate. Without loss of generality, we can assume there is a subsignature  $[a, b, c, d]$  of  $f$  (or its reversal), such that  $a + c = 0$ ,  $b + d \neq 0$ , and  $c^2 = bd$ . Combining two copies of  $[a, b, c, d]$  by connecting two pairs of dangling edges, we can get a binary function  $[a', b', c'] = [a^2 + 2b^2 + c^2, ab + 2bc + cd, b^2 + 2c^2 + d^2]$ . Then  $a' = 2(b^2 + c^2)$ ,  $b' = c(b + d) \neq 0$ ,  $c' = (b + d)^2$ . And so  $a' + c' > 0$ , and  $a'c' - b'^2 = (b + d)^2(2b^2 + c^2) > 0$ . We are done by Lemma 5.3.  $\square$

LEMMA 5.6. If  $\mathcal{F} \not\subseteq \mathcal{G}_1 \cup \mathcal{F}_2 \cup \mathcal{G}_3$ , then Theorem 5.1 holds.

*Proof.* If  $\mathcal{F} \not\subseteq \mathcal{G}_1 \cup \mathcal{G}_2 \cup \mathcal{G}_3$ , then by Lemma 5.5, we are done. Otherwise, there exists a signature  $f \in \mathcal{F} \subseteq \mathcal{G}_1 \cup \mathcal{G}_2 \cup \mathcal{G}_3$  and  $f \notin \mathcal{G}_1 \cup \mathcal{F}_2 \cup \mathcal{G}_3$ . Then it must be the case that  $f \in \mathcal{G}_2$ . We claim that  $f$  has arity greater than two. Since  $\mathcal{U} \subset \mathcal{G}_3$ , the arity of  $f$  is at least two. Being from  $\mathcal{F}$  and of arity at least two,  $f$  is nondegenerate. If  $f$  has arity two, since  $f \in \mathcal{G}_2$ , it has form  $[*, 0, *]$  or  $[0, *, 0]$ . Being nondegenerate, the former belongs to  $\mathcal{G}_1$ . The latter belongs to  $\mathcal{G}_3$ . Hence  $f$  has arity at least three. Since  $f \notin \mathcal{G}_1$ , there is some nonzero entry in the middle of the signature  $f$ ; after normalization, we can assume there is a subsignature of the form  $[0, 1, 0, x]$  (or  $[x, 0, 1, 0]$ ). If  $x \neq \pm 1$ , then by Lemma 5.4, the problem is  $\#P$ -hard and we are done. Otherwise, for every such pattern, we have  $x = \pm 1$ . Since  $f \notin \mathcal{F}_2$ , then there is some subsignature  $[0, 1, 0, -1]$  and because  $f \notin \mathcal{G}_3$ , there is some subsignature  $\pm[0, 1, 0, 1]$ . Therefore, there is a subsignature  $[1, 0, 1, 0, -1]$  of  $f$ , up to a nonzero multiple. Then we can construct an  $\mathcal{F}$ -gate as in Figure 5.3, whose signature is  $[8, 0, 4, 0]$ . So by Lemma 5.4, we know that the problem is  $\#P$ -hard and we are done. This completes the proof.  $\square$

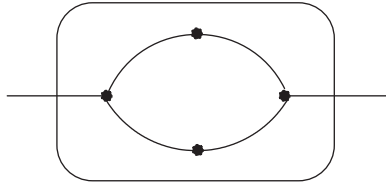


FIG. 5.4. The function on degree 2 nodes is  $[1, a, -1]$ , and the function on degree 3 nodes is  $[1, 0, 0, b']$ .

LEMMA 5.7. If  $\mathcal{F} \not\subseteq \mathcal{G}_1 \cup \mathcal{G}_3$ , then Theorem 5.1 holds.

*Proof.* If  $\mathcal{F} \not\subseteq \mathcal{G}_1 \cup \mathcal{F}_2 \cup \mathcal{G}_3$ , then by Lemma 5.6, we are done. Otherwise, there exists a signature  $f \in \mathcal{F} \subseteq \mathcal{G}_1 \cup \mathcal{F}_2 \cup \mathcal{G}_3$  and  $f \notin \mathcal{G}_1 \cup \mathcal{G}_3$ . Then it must be the case that  $f \in \mathcal{F}_2$ . Note that every signature with arity at most two in  $\mathcal{F}_2$  is also contained in  $\mathcal{G}_1 \cup \mathcal{G}_3$ , so  $f$  is of arity greater than two. Then  $f$  has a subsignature  $[1, 0, 1, 0]$  or  $[0, 1, 0, 1]$ . By symmetry, we assume it is  $[1, 0, 1, 0]$ . If  $\mathcal{F} \subseteq \mathcal{F}_1 \cup \mathcal{F}_2 \cup \mathcal{F}_3$ , then Theorem 5.1 trivially holds and there is nothing to prove. If not, there exists a signature  $g \in \mathcal{F} - \mathcal{F}_1 \cup \mathcal{F}_2 \cup \mathcal{F}_3$ . By  $\mathcal{F} \subseteq \mathcal{G}_1 \cup \mathcal{F}_2 \cup \mathcal{G}_3$ , either  $g \in \mathcal{G}_1 - \mathcal{F}_1 \cup \mathcal{F}_2 \cup \mathcal{F}_3 (\subseteq \mathcal{G}_1 - \mathcal{F}_1)$  or  $g \in \mathcal{G}_3 - \mathcal{F}_1 \cup \mathcal{F}_2 \cup \mathcal{F}_3 (\subseteq \mathcal{G}_3 - \mathcal{F}_3)$ .

For the first case,  $g \in \mathcal{G}_1 - \mathcal{F}_1$ , then after eliminating a scalar factor,  $g$  is of the form  $[1, 0, 0, \dots, b]$ , where  $b \notin \{-1, 0, 1\}$ . If the arity of  $g$  is odd, we can realize  $[1, b]$  by connecting every two dangling edges into one edge and leave one dangling edge. Then connecting this unary signature to one dangling edge of  $[1, 0, 1, 0]$ , we can realize a binary signature  $[1, b, 1]$ . Then by Lemma 5.3, Theorem 5.1 holds. If the arity of  $g$  is even, we can realize  $[1, 0, b]$  (leaving two dangling edges). By connecting one dangling edge of each of 3 copies of  $[1, 0, b]$  to each of 3 dangling edges of  $[1, 0, 1, 0]$ , we can have a new ternary symmetric signature  $[1, 0, b^2, 0]$ , which is equivalent to  $[1/b^2, 0, 1, 0]$ . By Lemma 5.4, we know the problem is #P-hard.

For the second case  $g \in \mathcal{G}_3 - \mathcal{F}_3$ , then  $g$  has a subsignature of the form  $[1, b]$ , where  $b \notin \{-1, 0, 1\}$ . By the same argument as above, Theorem 5.1 holds. This completes the proof.  $\square$

LEMMA 5.8. If  $\mathcal{F} \not\subseteq \mathcal{G}_1 \cup \mathcal{F}_3$ , then Theorem 5.1 holds.

*Proof.* If  $\mathcal{F} \not\subseteq \mathcal{G}_1 \cup \mathcal{G}_3$ , then by Lemma 5.7, we are done. Otherwise, there exists a signature  $f \in \mathcal{F} \subseteq \mathcal{G}_1 \cup \mathcal{G}_3$  and  $f \notin \mathcal{G}_1 \cup \mathcal{F}_3$ . Then it must be the case that  $f \in \mathcal{G}_3$ . Since every unary signature is in  $\mathcal{G}_1 \cup \mathcal{F}_3$ , the arity of  $f$  is at least two. It can be seen that  $f$  has a subsignature of the form  $[1, a, -1]$ , where  $a \notin \{-1, 0, 1\}$ .

If  $\mathcal{F} \subseteq \{[1, 0, 1]\} \cup \mathcal{G}_3$ , then  $\text{Holant}^*(\mathcal{F})$  is polynomial time computable by Theorem 4.1, and as a result Theorem 5.1 holds and we are done.

If not, there exists a signature  $g \in \mathcal{F} \subseteq \mathcal{G}_1 \cup \mathcal{G}_3$  and  $g \notin \{[1, 0, 1]\} \cup \mathcal{G}_3$ . Then it must be the case that  $g \in \mathcal{G}_1$ . The arity of  $g$  is greater than one, since  $g \notin \mathcal{G}_3$ .

If the arity of  $g$  is two, then  $g$  is of the form  $[1, 0, b]$ , where  $b \notin \{-1, 0, 1\}$ . Connecting two copies of  $[1, 0, b]$  to both sides of one binary signature  $[1, a, -1]$ , we can get a new binary signature  $[1, ab, -b^2]$ . It satisfies all the conditions of Lemma 5.3, and we are done. If the arity of  $g$  is greater than two, then we can always realize a signature  $[1, 0, 0, b']$ , where  $b' \neq 0$ . (We can connect the unary signature  $[1, a]$  to all dangling edges of  $g$  except three.) Then we can use an  $\mathcal{F}$ -gate as in Figure 5.4. Its signature is  $[1, a^2b', b'^2]$ , and by Lemma 5.3, we are done. This completes the proof.  $\square$

By the above lemmas, the only case we have to handle is that  $\mathcal{F} \subseteq \mathcal{G}_1 \cup \mathcal{F}_3$ . This is done by the following lemma, which completes the proof of Theorem 5.1.

LEMMA 5.9. *If  $\mathcal{F} \subseteq \mathcal{G}_1 \cup \mathcal{F}_3$ , then Theorem 5.1 holds.*

*Proof.* If  $\mathcal{F} \subseteq \mathcal{F}_1 \cup \mathcal{F}_3$ ,  $\text{Holant}^c(\mathcal{F})$  is computable in polynomial time. Similarly, if  $\mathcal{F} \subseteq \mathcal{U} \cup \mathcal{F}_3 \cup \{[1, 0, 1]\}$ , then by Theorem 4.1,  $\text{Holant}^c(\mathcal{F})$  is computable in polynomial time. Hence in these two cases, Theorem 5.1 holds. Now suppose  $\mathcal{F} \not\subseteq \mathcal{F}_1 \cup \mathcal{F}_3$  and  $\mathcal{F} \not\subseteq \mathcal{U} \cup \mathcal{F}_3 \cup \{[1, 0, 1]\}$ .

There exists  $f \in \mathcal{F} - \mathcal{F}_1 \cup \mathcal{F}_3$ . Since  $\mathcal{F} \subseteq \mathcal{G}_1 \cup \mathcal{F}_3$ , such an  $f$  belongs to  $\mathcal{G}_1$ .

Now there are two cases. The first case is that we have such an  $f \notin \mathcal{U}$ , i.e.,  $f \in \mathcal{F} \cap \mathcal{G}_1 - (\mathcal{F}_1 \cup \mathcal{F}_3 \cup \mathcal{U})$ . The arity of  $f$  is greater than one. By connecting its dangling edges together except two or three depending on the parity of the arity of  $f$ , we can assume  $f$  has the form  $[1, 0, a]$  or  $[1, 0, 0, a]$ , where  $a \notin \{-1, 0, 1\}$ .

The second case is that every  $f \in \mathcal{F} \cap \mathcal{G}_1 - (\mathcal{F}_1 \cup \mathcal{F}_3)$  is also in  $\mathcal{U}$ . By  $\mathcal{F} \not\subseteq \mathcal{U} \cup \mathcal{F}_3 \cup \{[1, 0, 1]\}$ , there exists an  $f_1 \in \mathcal{F} - (\mathcal{U} \cup \mathcal{F}_3 \cup \{[1, 0, 1]\})$ . Since  $\mathcal{F} \subseteq \mathcal{G}_1 \cup \mathcal{F}_3$ , and  $f_1 \notin \mathcal{F}_3$ , we get  $f_1 \in \mathcal{G}_1$ . If  $f_1 \notin \mathcal{F}_1$ , we could use this  $f_1$  as the  $f$  above, namely,  $f_1 \in \mathcal{F} \cap \mathcal{G}_1 - (\mathcal{F}_1 \cup \mathcal{F}_3 \cup \mathcal{U})$ . This contradicts the condition of this case. Thus  $f_1 \in \mathcal{F}_1$ . Also we have some  $f_2 \in \mathcal{F} - (\mathcal{F}_1 \cup \mathcal{F}_3)$ . We have  $f_2 \in \mathcal{G}_1$ , since  $\mathcal{F} \subseteq \mathcal{G}_1 \cup \mathcal{F}_3$ . So  $f_2 \in \mathcal{F} \cap \mathcal{G}_1 - (\mathcal{F}_1 \cup \mathcal{F}_3)$ . Since we are in this second case, certainly  $f_2 \in \mathcal{U}$ .

To recap, we have  $f_1, f_2 \in \mathcal{F} \cap \mathcal{G}_1$  such that  $f_1 \in \mathcal{F}_1$  but  $f_1 \notin \mathcal{U} \cup \mathcal{F}_3 \cup \{[1, 0, 1]\}$ , and  $f_2 \in \mathcal{U}$  but  $f_2 \notin \mathcal{F}_1$ . The arity of  $f_1$  is at least two. We claim it is greater than two. Otherwise,  $f_1$  being from  $\mathcal{F}_1$  and not  $[1, 0, 1]$ , it would be  $f_1 = [1, 0, -1] \in \mathcal{F}_3$ , a contradiction. So  $f_1$  has the form  $[1, 0, 0, \dots, \pm 1]$  of arity at least three.  $f_2$  is of the form  $[1, a']$ , where  $a' \notin \{-1, 0, 1\}$ ; this follows from  $f_2 \in \mathcal{U} \cap \mathcal{G}_1 - \mathcal{F}_1$ . By connecting all the dangling edges of  $f_1$  except two with  $f_2$ , we can construct an  $\mathcal{F}$ -gate with a signature of the form  $[1, 0, a]$ , where  $a \notin \{-1, 0, 1\}$ . This is one of the above two forms after the first case. To sum up, in both cases, we have some  $f$  of the form  $[1, 0, a]$  or  $[1, 0, 0, a]$ , where  $a \notin \{-1, 0, 1\}$ .

If  $\mathcal{F} \subseteq \mathcal{G}_1 \cup \{[0, 1, 0]\} \cup \mathcal{U}$ , then by Theorem 4.1,  $\text{Holant}^c(\mathcal{F})$  is computable in polynomial time and Theorem 5.1 holds. Otherwise, there exists a  $g \in \mathcal{F} \subseteq \mathcal{G}_1 \cup \mathcal{F}_3$ , and  $g \notin \mathcal{G}_1 \cup \{[0, 1, 0]\} \cup \mathcal{U}$ . Being not in  $\mathcal{U}$ ,  $g$  has arity at least two. Being not in  $\mathcal{G}_1$ ,  $g$  must be in  $\mathcal{F}_3$ . Then  $g$  has one of the following subsignatures:  $[1, 1, -1], [1, -1, -1], [1, 0, -1, 0], [0, 1, 0, -1]$ ; this follows from a careful examination of the forms of  $\mathcal{F}_3$ . By symmetry (taking the reversal of both  $f$  and  $g$ ), we only need to consider two cases for  $f = [1, 0, a]$  or  $[1, 0, 0, a]$ , where  $a \notin \{-1, 0, 1\}$ , and two cases for  $g = [1, 1, -1]$  or  $[1, 0, -1, 0]$ .

According to  $f$  and  $g$ , we have four cases. If  $f = [1, 0, a]$  and  $g = [1, 1, -1]$ , then connecting them together into a chain  $fgf$ , we can realize  $[1, a, -a^2]$ . By Lemma 5.3, we are done. If  $f = [1, 0, a]$  and  $g = [1, 0, -1, 0]$ , for each dangling edge of  $g$ , we extend it by one copy of  $f$ . Then we can realize  $[1, 0, -a^2, 0]$ . So by Lemma 5.4, we are done. If  $f = [1, 0, 0, a]$  and  $g = [1, 1, -1]$ , we can connect a unary signature  $[1, 1]$  (a subsignature of  $g$ ) to one dangling edge of  $f$  and realize a binary signature  $f' = [1, 0, a]$ . This reduces it to the first case, which has been proved. If  $f = [1, 0, 0, a]$  and  $g = [1, 0, -1, 0]$ , we can realize a unary signature  $[1, a]$  from  $f$  by connecting two of its dangling edges together, and then connect this unary signature to one dangling edge of  $g$  to realize  $[1, -a, -1]$ . Note that  $[1, -a, -1] \notin \mathcal{G}_1 \cup \mathcal{F}_3$ , and by Lemma 5.8, we are done.  $\square$

**6. Conclusions.** We have introduced an alternative framework, called the Holant problems, in the study of the complexity of counting problems. We have proved some dichotomy theorems for both the Holant\* setting as well as the Holant<sup>c</sup> setting. Some deeper mathematical structure becomes visible when we treat EQUALITY functions as an ordinary function subject to holographic reductions. This is a feature that is

hard to discuss in the traditional #CSP formulation of the complexity of counting problems, where all EQUALITY functions are assumed to be present for free. A major open area for future investigations in #CSP problems is when the constraint functions can take negative values, and thus cancellations occur in the counting problems they define. The Holant framework seems a natural fit when we need to address potential cancellations in the computation of counting problems. Holographic reductions are a main tool in this investigation.

The results on Holant problems in this paper are only a beginning. There are many open problems for future investigations. The results on Holant\* and Holant<sup>c</sup> are restrictions of Holant problems. Naturally one should investigate the unrestricted Holant problems. The local constraint functions in this paper are all symmetric functions. One can investigate asymmetric functions. The domain size in this paper is Boolean. To extend to larger domains would be very challenging; but such a generalization would be a vast extension on all #CSP problems, with cancellations. Currently this is still largely unexplored.

**Acknowledgment.** We thank many colleagues who have expressed an interest and gave us valuable feedbacks on this work. Special thanks to Martin Dyer, Leslie Ann Goldberg, Mark Jerrum, Leslie Valiant, and the anonymous referees for our conference paper at STOC 2009 [11], which contained some of the results reported here. We are also very grateful to the anonymous referees for our journal submission. The journal version contained many more proof details, and we are thankful to the referees for very detailed, insightful, and constructive comments.

## REFERENCES

- [1] L. BLUM, F. CUCKER, M. SHUB, AND S. SMALE, *Complexity and Real Computation*, Springer-Verlag, New York, Secaucus, NJ, 1998.
- [2] A. A. BULATOV, *A dichotomy theorem for constraint satisfaction problems on a 3-element set*, J. ACM, 53 (2006), pp. 66–120.
- [3] A. A. BULATOV, *The complexity of the counting constraint satisfaction problem*, in Proc. ICALP (1), Lecture Notes in Computer Science 5125, L. Aceto, I. Damgård, L. A. Goldberg, M. M. Halldórsson, A. Ingólfssdóttir, and I. Walukiewicz, eds., Springer, New York, 2008, pp. 646–661.
- [4] A. A. BULATOV AND V. DALMAU, *Towards a dichotomy theorem for the counting constraint satisfaction problem*, in FOCS, IEEE Computer Society, New York, 2003, pp. 562–571.
- [5] A. A. BULATOV AND M. GROHE, *The complexity of partition functions*, in Proc. ICALP, 2004, Turku, Finland, Lecture Notes in Comput. Sci. 3142, Springer, 2004, pp. 294–306.
- [6] A. A. BULATOV AND M. GROHE, *The complexity of partition functions*, Theor. Comput. Sci., 348 (2005), pp. 148–186.
- [7] J.-Y. CAI, X. CHEN, AND P. LU, *Graph homomorphisms with complex values: A dichotomy theorem*, in Proc. ICALP (1), 2010, Bordeaux, France, Lecture Notes in Comput. Sci. 6198, Springer, 2010, pp. 275–286.
- [8] J.-Y. CAI AND V. CHOUDHARY, *Valiant’s holant theorem and matchgate tensors*, Theor. Comput. Sci., 384 (2007), pp. 22–32.
- [9] J.-Y. CAI AND P. LU, *Holographic algorithms: From art to science*, J. Comput. Syst. Sci., 77 (2011), pp. 41–61.
- [10] J.-Y. CAI AND P. LU, *On symmetric signatures in holographic algorithms*, Theory Comput. Syst., 46 (2010), pp. 398–415.
- [11] J.-Y. CAI, P. LU, AND M. XIA, *Holant problems and counting CSP*, in Proceedings of STOC 2009: The 41st Annual ACM Symposium on Theory of Computing, M. Mitzenmacher, ed., Bethesda, MD, pp. 715–724, 2009.
- [12] J.-Y. CAI, P. LU, AND M. XIA, *Holographic algorithms by Fibonacci gates*, Linear Algebra Appl. special issue on Tensors and Multilinear Algebra, to appear; available online from <http://www.sciencedirect.com/science/article/pii/S0024379511001625>.

- [13] J.-Y. CAI, P. LU, AND M. XIA, *Holographic algorithms by Fibonacci gates and holographic reductions for hardness*, in FOCS '08: Proceedings of the 49th Annual IEEE Symposium on Foundations of Computer Science, Philadelphia, PA, 2008, IEEE Computer Society, New York, 2008.
- [14] J.-Y. CAI, P. LU, AND M. XIA, *The complexity of complex weighted Boolean #CSP*, Upcoming Journal Submission.
- [15] N. CREIGNOU, S. KHANNA, AND M. SUDAN, *Complexity classifications of Boolean constraint satisfaction problems*, SIAM Monogr. Discrete Math. Appl., SIAM, Philadelphia, 2001.
- [16] P. DAGUM AND M. LUBY, *Approximating the permanent of graphs with large factors*, Theor. Comput. Sci., 102 (1992), pp. 283–305.
- [17] C. T. J. DODSON AND T. POSTON, *Tensor geometry*, Graduate Texts in Mathematics 130, Springer-Verlag, New York, 1991.
- [18] M. E. DYER, L. A. GOLDBERG, AND M. JERRUM, *The complexity of weighted Boolean CSP*, SIAM J. Comput., 38 (2008), pp. 1970–1986.
- [19] M. E. DYER, L. A. GOLDBERG, AND M. PATERSON, *On counting homomorphisms to directed acyclic graphs*, in ICALP (1), Lecture Notes in Computer Science 4051, M. Bugliesi, B. Preneel, V. Sassone, and I. Wegener, eds., Springer, New York, 2006, pp. 38–49.
- [20] M. E. DYER, L. A. GOLDBERG, AND M. PATERSON, *On counting homomorphisms to directed acyclic graphs*, J. ACM, 54 (2007).
- [21] M. E. DYER AND C. S. GREENHILL, *The complexity of counting graph homomorphisms (extended abstract)*, in Proceedings of the 11th Annual ACM-SIAM Symposium on Discrete Algorithms, 2000, San Francisco, CA, ACM/SIAM, pp. 246–255.
- [22] M. E. DYER AND C. S. GREENHILL, *Corrigendum: The complexity of counting graph homomorphisms*, Random Structures Algorithms, 25 (2004), pp. 346–352.
- [23] M. FREEDMAN, L. LOVÁSZ, AND A. SCHRIJVER, *Reflection positivity, rank connectivity, and homomorphism of graphs*, J. AMS, 20 (2007), pp. 37–51.
- [24] L. A. GOLDBERG, M. GROHE, M. JERRUM, AND M. THURLEY, *A complexity dichotomy for partition functions with mixed signs*, CoRR, abs/0804.1932, 2008.
- [25] P. W. KASTELEYN, *The statistics of dimers on a lattice*, Phys., 27 (1961), pp. 1209–1225.
- [26] P. W. KASTELEYN, *Graph theory and crystal physics*, in Graph Theory and Theoretical Physics, F. Harary, ed., Academic Press, London, 1967, pp. 43–110.
- [27] K.-I. KO, *Complexity theory of real functions*, Birkhäuser Boston, Cambridge, MA, 1991.
- [28] R. LIDL AND H. NIEDERREITER, *Finite fields*, Encyclopedia of Mathematics and its Applications, Vol. 20, Cambridge University Press, Cambridge, 1997.
- [29] H. N. V. TEMPERLEY AND M. E. FISHER, *Dimer problem in statistical mechanics-an exact result*, Philos. Mag., 6 (1961), p. 1061C–1063.
- [30] S. P. VADHAN, *The complexity of counting in sparse, regular, and planar graphs*, SIAM J. Comput., 31 (2001), pp. 398–427.
- [31] L. G. VALIANT, *The complexity of enumeration and reliability problems*, SIAM J. Comput., 8 (1979), pp. 410–421.
- [32] L. G. VALIANT, *Quantum circuits that can be simulated classically in polynomial time*, SIAM J. Comput., 31 (2002), pp. 1229–1254.
- [33] L. G. VALIANT, *Accidental algorithms*, in FOCS '06: Proceedings of the 47th Annual IEEE Symposium on Foundations of Computer Science, Washington, DC, 2006, IEEE Computer Society, pp. 509–517.
- [34] L. G. VALIANT, *Holographic algorithms*, SIAM J. Comput., 37 (2008), pp. 1565–1594.
- [35] M. XIA, P. ZHANG, AND W. ZHAO, *Computational complexity of counting problems on 3-regular planar graphs*, Theor. Comput. Sci., 384 (2007), pp. 111–125.