10

Dichotomies for Asymmetric Constraints

Most dichotomy theorems presented so far in Volume I are for counting problems specified by sets of *symmetric* constraint functions. This is true for the first nontrivial dichotomy for Holant* problems in Chapter 2, for spin systems on regular graphs in Chapter 5, for Holant^c problems and for CSP^d problems in Chapter 6, for the Holant dichotomy in Chapter 7, and for the planar Holant problems in Chapter 9. There is an exception for Boolean #CSP in Chapter 3, which is valid for all sets of complex-valued constraint functions that are not necessarily symmetric, over the Boolean domain. The planar version, Theorem 8.34 in Chapter 8, also applies only to sets of symmetric constraint functions.

In this final chapter of Volume I, we discuss the few results currently known regarding classification theorems specified by not necessarily symmetric constraint functions.

The first result we present is the generalization of Theorem 8.34 to the asymmetric setting. We prove that the classification in Theorem 8.34 still holds for sets of not necessarily symmetric constraint functions. Owing to its length, we will not give the full proof of this theorem, but only a proof outline, and illustrate some ideas in the proof of a few lemmas as an example. The full proof of this theorem can be found in [CF16].

The second result we present is a complexity dichotomy for Holant* problems specified by not necessarily symmetric constraint functions. This is a generalization of Theorem 2.12 from Chapter 2. The proof that will be presented also provides a better insight to the Holant* dichotomy, even for the symmetric case.

A common theme of the proofs of these theorems is a reduction to the dichotomies already proved for the symmetric case.

https://doi.org/10.1017/9781107477063.011 Published online by Cambridge University Press

10.1 Planar #CSP with Asymmetric Constraints

In Chapter 8 we proved Theorem 8.34, which shows that the complexity dichotomy in Chapter 3 for #CSP problems on Boolean variables (Theorem 3.7) has an extension to take into account the planar case. Specifically, Theorem 8.34 shows that for #CSP problems over any set of complex-valued *symmetric* constraint functions on Boolean variables, Valiant's holographic transformation followed by Kasteleyn's algorithm (Theorem 4.7) is a *universal* methodology to solve in polynomial time on planar graphs those counting problems that are #P-hard in general.

The first main theorem of Chapter 10 states that this statement is also valid for complex-valued constraint functions on Boolean variables that are *not* necessarily symmetric.

Theorem 10.1. *For any finite set of constraint functions $\mathcal{F}$ over Boolean variables, each taking complex values and not necessarily symmetric, #CSP($\mathcal{F}$) belongs to exactly one of three types according to $\mathcal{F}$:*

1. *It is P-time solvable.*
2. *It is P-time solvable over planar graphs but #P-hard over general graphs.*
3. *It is #P-hard over planar graphs.*

Moreover, type (2) consists of precisely those problems that are holographically reducible to the Kasteleyn's algorithm, also known as the Fisher–Kasteleyn–Temperley (FKT) algorithm.

The theorem has an explicit form. The three fundamental tractable classes $\mathscr{A}$ (affine), $\mathscr{P}$ (product type), and $\mathscr{M}$ (matchgates) are all defined (Definition 3.6, Definition 3.2, and Definition 4.9 respectively) for complex-valued constraint functions that are not necessarily symmetric. Then #CSP($\mathcal{F}$) has type *(1)* if $\mathcal{F} \subseteq \mathscr{A}$ or $\mathcal{F} \subseteq \mathscr{P}$; otherwise it has type *(2)* if $\mathcal{F} \subseteq \widehat{\mathscr{M}}$, where $\widehat{\mathscr{M}}$ is $\mathscr{M}$ under the holographic transformation $H_2 = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$; otherwise it has type *(3)*. This theorem finally settles the full reach of the power of Valiant's holographic algorithms in the #CSP framework over Boolean variables.

The theorem is also valid for infinite sets of constraint functions $\mathcal{F}$ in the following sense. If $\mathcal{F}$ satisfies one of the tractability criteria, then the problem is P-time solvable where the input size of an instance includes the description of the constraints used. On the other hand, if $\mathcal{F}$ does not satisfy the tractability criteria, then there is a finite subset $\mathcal{F}' \subset \mathcal{F}$ such that the problem defined by constraint functions $\mathcal{F}'$ is #P-hard.

https://doi.org/10.1017/9781107477063.011 Published online by Cambridge University Press

From Chapter 9, we know that there are #P-hard Holant problems that become tractable in P-time on planar structures, but they are *not* holographically reducible to the FKT. In view of this, the validity of Theorem 10.1 should not be considered self-evident.

It turns out that the class of Holant problems is more than just a separate framework providing a cautionary reference to the validity of Theorem 10.1. In fact, they form the main arena where we carry out the proof of Theorem 10.1. A basic idea in this proof is a holographic transformation between the #CSP setting and the Holant setting via the Hadamard transformation H_2. As with classical Fourier transform, certain properties are easier to handle in one setting while others are easier after the transformation. We will go back and forth.

In Section 10.1.1 we give a proof outline of Theorem 10.1. In Section 10.1.2 we prove some results illustrating several technical ideas used. Among the new techniques, we introduce a tableau calculus, compressed signatures, integrality gap, and a reduction to symmetric planar #CSP^2 dichotomy.

10.1.1 Outline of the Proof of Theorem 10.1

We now give a proof outline of Theorem 10.1 and also explain some overall vision that guided this proof. An important technique is to view our counting problems in the dual perspectives of planar #CSP *and* planar Holant problems, i.e., we make essential use of the equivalence Pl-#CSP($\mathcal{F}$) $\equiv_{\mathrm{T}}$ Pl-Holant($\widehat{\mathcal{EQ}}, \widehat{\mathcal{F}}$), where $\widehat{\mathcal{F}} = H_2\mathcal{F}$. Some questions are easier to handle in one framework, while others are easier in the other.

We want to show that *either* $\mathcal{F} \subseteq \mathscr{A}$, or $\mathcal{F} \subseteq \mathscr{P}$, or $\mathcal{F} \subseteq \widehat{\mathscr{M}}$, in which case Pl-#CSP($\mathcal{F}$) is computable in P-time, *or else* Pl-#CSP($\mathcal{F}$) is #P-hard. In the Pl-Holant($\widehat{\mathcal{EQ}}, \widehat{\mathcal{F}}$) setting, the tractability condition is expressed as $\widehat{\mathcal{F}} \subseteq \widehat{\mathscr{A}}$, or $\widehat{\mathcal{F}} \subseteq \widehat{\mathscr{P}}$, or $\widehat{\mathcal{F}} \subseteq \mathscr{M}$.

Note that $\mathscr{A}$ is invariant under the transformation by H_2, i.e., $\widehat{\mathscr{A}} = \mathscr{A}$. However, $\widehat{\mathscr{P}}$ is more difficult to reason about than $\mathscr{P}$, while $\mathscr{M}$ is easier than $\widehat{\mathscr{M}}$ to handle. The former suggests that we carry out our proof in the Pl-#CSP setting, while the latter suggests the opposite, that we do so in the Pl-Holant setting instead.

One necessary condition for $\mathscr{M}$ is the parity condition for matchgates, i.e., all nonzero entries of a matchgate signature must have the same parity in Hamming weight. If some signature in $\widehat{\mathcal{F}}$ violates the parity condition, then we have eliminated one possibility $\widehat{\mathcal{F}} \subseteq \mathscr{M}$. In this case if we prove in the Pl-#CSP framework, we can avoid discussing the more difficult class $\widehat{\mathscr{M}}$. On the other hand, if $\widehat{\mathcal{F}}$ satisfies the parity condition, then we have a lucky situation, as follows.

https://doi.org/10.1017/9781107477063.011 Published online by Cambridge University Press

Fact: For any signature $g \in \mathscr{P}$, if the transformation $\hat{g} = H_2^{\otimes \text{arity}(g)} g$ satisfies the parity condition, then $g \in \mathscr{A}$. Consequently $\hat{g} \in \mathscr{A}$ as well since $\widehat{\mathscr{A}} = \mathscr{A}$.

Exercise: Prove this fact for the special case in which g has at most two antipodal points in its support, and then conclude that the statement holds in general.

Hence if every function in $\widehat{\mathcal{F}}$ satisfies the parity condition, then $\widehat{\mathcal{F}} \subseteq \widehat{\mathscr{P}}$ already implies $\widehat{\mathcal{F}} \subseteq \mathscr{A}$, with the consequence that we do not need to specifically discuss the tractability condition $\widehat{\mathcal{F}} \subseteq \widehat{\mathscr{P}}$. Thus in this case we can avoid discussing the irksome class $\widehat{\mathscr{P}}$ altogether.

Therefore there are two main cases:

Case (1): $\widehat{\mathcal{F}}$ does not satisfy the parity condition. Then we want to show that Pl-#CSP($\mathcal{F}$) is #P-hard unless $\mathcal{F} \subseteq \mathscr{A}$ or $\mathcal{F} \subseteq \mathscr{P}$. From any signature in $\widehat{\mathcal{F}}$ violating the parity condition, we can construct a unary signature $[1, w]$ with $w \neq 0$, in the Holant framework Pl-Holant($\widehat{\mathcal{EQ}}$, $\widehat{\mathcal{F}}$). Any signature that violates the parity condition is a witness that $\widehat{\mathcal{F}} \not\subseteq \mathscr{M}$, or equivalently $\mathcal{F} \not\subseteq \widehat{\mathscr{M}}$. This unary signature $[1, w]$ with $w \neq 0$ is such a witness, and it is symmetric. If $\mathcal{F} \subseteq \mathscr{A}$ or $\mathcal{F} \subseteq \mathscr{P}$, then the problem Pl-#CSP($\mathcal{F}$) is tractable. Otherwise, there exist some signatures $f, g \in \mathcal{F}$ such that $f \notin \mathscr{A}$ and $g \notin \mathscr{P}$. We would like to construct some *symmetric* signatures from these that are also nonaffine and nonproduct type, respectively, and then apply the known dichotomy theorem for symmetric signatures. For the nonproduct type we will do so in the Pl-#CSP($\mathcal{F}$) setting, to avoid having to deal with $\widehat{\mathscr{P}}$. For the nonaffine signatures, we can do so in either the Pl-#CSP framework or the Pl-Holant framework as $\mathscr{A}$ is invariant $\widehat{\mathscr{A}} = \mathscr{A}$.

However, the obstacle in this plan is that it is generally difficult to construct *symmetric* signatures from *asymmetric* (i.e., not necessarily symmetric) signatures in a *planar* fashion, for arity > 3. Therefore, a main engine of the proof is *arity reduction*. Starting from a nonproduct type signature of arity $n > 3$, we construct in the Pl-#CSP setting a nonproduct type signature of arity $n - 1$. Then we show how to construct, from any nonproduct type signature of arity 3, a *symmetric* and nonproduct type signature of arity 3. These constructions will need suitable unary signatures which will be constructed, starting with that $[1, w]$ constructed in Pl-Holant($\widehat{\mathcal{EQ}}$, $\widehat{\mathcal{F}}$).

For the construction of nonaffine signatures, we will introduce a *tableau calculus*. Again we will carry out an arity reduction proof, this time all the way down to arity 1. We prove that with the help of three unary signatures $[1, 0]$, $[0, 1]$, $[1, x]$ with $x \neq 0$, we can get a unary nonaffine signature from any nonaffine signature of higher arity in the Pl-Holant setting. This proof heavily depends on the new tableau calculus. Then we construct $[1, 0]$, $[0, 1]$, $[1, x]$ by

https://doi.org/10.1017/9781107477063.011 Published online by Cambridge University Press

shuttling between Pl-#CSP($\mathcal{F}$) and Pl-Holant($\widehat{\mathcal{EQ}}, \widehat{\mathcal{F}}$). There is an exceptional case where all signatures in $\mathcal{F}$ are $\{0, 1\}$-valued in Pl-#CSP($\mathcal{F}$). In this case, we cannot construct $[1, 0]$, $[0, 1]$, $[1, x]$ simultaneously. We resolve this case separately. For $\{0, 1\}$-valued $\mathcal{F}$, we actually also cannot construct all the unary signatures in the arity reduction proof for nonproduct type, if we *only* assume the existence of some $g \in \mathcal{F} \setminus \mathscr{P}$. However, if we have both $g \in \mathcal{F} \setminus \mathscr{P}$ and some $f \in \mathcal{F} \setminus \mathscr{A}$, we can use f to produce the needed unary signatures to help the arity reduction on g. All these use tableau calculus.

Case (2): $\widehat{\mathcal{F}}$ satisfies the parity condition. In this case, if $\widehat{\mathcal{F}} \subseteq \mathscr{A}$, or $\widehat{\mathcal{F}} \subseteq \widehat{\mathscr{P}}$, or $\widehat{\mathcal{F}} \subseteq \mathscr{M}$, then the problem is tractable. However, due to the parity condition, there are only two consequential conditions here, $\widehat{\mathcal{F}} \subseteq \mathscr{A}$ or $\widehat{\mathcal{F}} \subseteq \mathscr{M}$; the containment $\widehat{\mathcal{F}} \subseteq \widehat{\mathscr{P}}$ is subsumed by $\widehat{\mathcal{F}} \subseteq \mathscr{A}$. Therefore we want to prove that if $\widehat{\mathcal{F}} \not\subseteq \mathscr{A}$ and $\widehat{\mathcal{F}} \not\subseteq \mathscr{M}$, then Pl-Holant($\widehat{\mathcal{EQ}}, \widehat{\mathcal{F}}$) is #P-hard.

Again a natural idea is to construct nonaffine and nonmatchgate *symmetric* signatures from any such asymmetric signatures, and then we can apply the known dichotomy theorem for symmetric signatures. The main difficulty of this approach lies in dealing with nonmatchgate signatures. Note that both $\widehat{\mathcal{F}}$ and $\widehat{\mathcal{EQ}}$ (being a subset of $\mathscr{M}$) satisfy the parity condition, and therefore the signature of any construction from $\widehat{\mathcal{EQ}} \cup \widehat{\mathcal{F}}$ must also satisfy the parity condition. On the other hand, it follows from matchgate identities (4.10) that any signature of arity ≤ 3 is a matchgate signature iff it satisfies the parity condition. Hence all constructible nonmatchgate signatures must have arity ≥ 4. But it is difficult to construct a symmetric signature from any asymmetric signature of arity ≥ 4 while preserving planarity.

So we take an alternative approach. For a given nonmatchgate signature, we first prove that we can get a nonmatchgate signature f of arity 4. Then we can construct from f either a special function called the crossover function $\mathfrak{X}$ or the EQUALITY function $(=_4)$ of arity 4. Here the crossover function $\mathfrak{X} = [(x_1 = x_3) \wedge (x_2 = x_4)]$ is a function on four variables x_1, x_2, x_3, x_4 in cyclic order, which ensures that pairs of diagonally crossed variables are equal.

If we have $\mathfrak{X}$, we can finish the proof by the nonplanar #CSP dichotomy theorem: To any nonplanar input signature grid, we just replace any crossing by $\mathfrak{X}$, and obtain an equivalent planar signature grid, and then we can apply Theorem 3.7.

If we have $(=_4)$, then we can get all $\mathcal{EQ}_2 = \{=_2, =_4, \ldots, =_{2k}, \ldots\}$ in Pl-Holant($\widehat{\mathcal{EQ}}, \widehat{\mathcal{F}}$). This implies that

$$\begin{aligned} \text{Pl-Holant}(\widehat{\mathcal{EQ}}, \widehat{\mathcal{F}}) &\equiv_{\mathrm{T}} \text{Pl-Holant}(\mathcal{EQ}_2, \widehat{\mathcal{EQ}}, \widehat{\mathcal{F}}) \\ &\equiv_{\mathrm{T}} \text{Pl-\#CSP}^2(\widehat{\mathcal{EQ}}, \widehat{\mathcal{F}}), \end{aligned} \tag{10.1}$$

https://doi.org/10.1017/9781107477063.011 Published online by Cambridge University Press

where #CSP^2 denotes #CSP in which every variable appears an even number of times.

Now comes a "cognitive dissonance." By (10.1), what used to be the "right-hand side" in the equivalence Pl-#CSP($\mathcal{F}$) $\equiv_{\mathrm{T}}$ Pl-Holant($\widehat{\mathcal{EQ}}, \widehat{\mathcal{F}}$) will be treated as a Pl-#CSP^2 problem:

$$\begin{array}{c} \text{Pl-\#CSP}(\mathcal{F}) \equiv_{\mathrm{T}} \text{Pl-Holant}(\widehat{\mathcal{EQ}}, \widehat{\mathcal{F}}) \equiv_{\mathrm{T}} \text{Pl-Holant}(\mathcal{EQ}_2, \widehat{\mathcal{EQ}}, \widehat{\mathcal{F}}) \\ \qquad\qquad\qquad\

2. If every function in $\widehat{\mathcal{F}}$ takes values in $\{0, 1, -1\}$ up to a constant but not every function in $\widehat{\mathcal{F}}$ takes values in $\{0, 1\}$ up to a constant, then we can construct $[1, 0, -1]$ and complete the proof.
3. If every function in $\widehat{\mathcal{F}}$ takes values in $\{0, 1\}$ then we prove it separately.

In all cases, we use our tableau calculus. The full proof of Theorem 10.1 can be found in [CF16].

10.1.2 Arity Reduction for Nonaffine Asymmetric Signatures

In this subsection we prove a few specific lemmas to illustrate some new techniques. The lemmas are used for arity reduction of nonaffine asymmetric signatures in the dual setting Pl-Holant$(\widehat{\mathcal{EQ}}, \widehat{\mathcal{F}})$.

As described in the proof outline, for any signature $f \notin \mathscr{A}$ (resp. $\mathscr{P}$, $\widehat{\mathscr{M}}$), we want to construct a *symmetric* signature $g \notin \mathscr{A}$ (resp. $\mathscr{P}$, $\widehat{\mathscr{M}}$), and then apply known dichotomy theorems. If f has high arity, it is very difficult to construct a symmetric signature from a nonsymmetric f in a planar fashion. So our first step is to reduce the arity of f. This task can be delicate. One reason is that a minor modification can change a tractable asymmetric signature to a signature that defines a #P-hard problem. For example, let f be the signature with signature matrix

$$M_{x_1x_2,x_4x_3}(f) = \begin{bmatrix} f_{0000} & f_{0010} & f_{0001} & f_{0011} \\ f_{0100} & f_{0110} & f_{0101} & f_{0111} \\ f_{1000} & f_{1010} & f_{1001} & f_{1011} \\ f_{1100} & f_{1110} & f_{1101} & f_{1111} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & -1 & 0 \\ 0 & -1 & -1 & 0 \\ 1 & 0 & 0 & -1 \end{bmatrix},$$

listing the 16 values of $f_{x_1x_2x_3x_4}$. Then $f \in \mathscr{A} \cap \mathscr{M}$. But if we change the sign of any one nonzero entry of f, then it is not in *any* tractable class, and it defines a #P-hard problem.

We will prove two lemmas, Lemmas 10.4 and 10.10. They show, in the case when all signatures in $\widehat{\mathcal{F}}$ satisfy the parity condition, how to reduce the arity of a nonaffine asymmetric signature in Pl-Holant$(\widehat{\mathcal{EQ}}, \widehat{\mathcal{F}})$. In Lemma 10.4 all signatures in $\widehat{\mathcal{F}}$ take values in $\{0, 1\}$. In Lemma 10.10 all signatures in $\widehat{\mathcal{F}}$ take values in $\{0, 1, -1\}$ and we need an additional auxiliary function. These lemmas are used in the proof of Theorem 10.1 in the second main case according to the proof outline; they are used to handle the two special cases (2) and (3) listed toward the end of that proof outline, after we managed to get $(=_4)$ and hence a reduction to #CSP^2.

In order to prove Lemma 10.4 we first present Lemmas 10.2 and 10.3, which help regularize a signature. Lemma 10.2 shows how to use $[0, 1]^{\otimes 2}$ and

https://doi.org/10.1017/9781107477063.011 Published online by Cambridge University Press

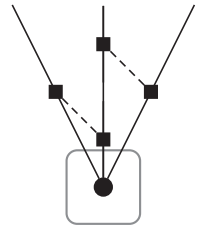

Figure 10.1. Flipping two variables of f that are not adjacent by $[0, 1, 0]^{\otimes 2}$ while preserving planarity. The circle vertex is labeled f and squares are $[0, 1, 0]$. A pair of squares connected by a dashed line forms $[0, 1, 0]^{\otimes 2}$.

$[1, 0, 1, 0] \in \widehat{\mathcal{EQ}}$ to get $[0, 1, 0]^{\otimes 2}$, then to flip any two variables that are not necessarily adjacent, while preserving planarity.

Lemma 10.2. *In* Pl-Holant$(\widehat{\mathcal{EQ}}, [0, 1]^{\otimes 2}, f)$*, if f has arity n, then for any $s \neq t \in [n]$, we can construct g such that $g(x'_1, x'_2, \ldots, x'_n) = f(x_1, x_2, \ldots, x_n)$, where $x'_k = \overline{x_k}$ for $k \in \{s, t\}$ and $x'_k = x_k$ otherwise. Moreover, $f \in \mathscr{A}$ iff $g \in \mathscr{A}$.*

Proof. That $f \in \mathscr{A}$ iff $g \in \mathscr{A}$ follows from the definition of $\mathscr{A}$.

By connecting $[0, 1]^{\otimes 2}$ to two disjoint copies of $[1, 0, 1, 0] \in \widehat{\mathcal{EQ}}$, we get $[0, 1, 0]^{\otimes 2}$. Note that this is a planar gadget where two adjacent pairs of variables are to be flipped. This function is $(x_1 \neq x_2) \wedge (x_3 \neq x_4)$.

Without loss of generality, we assume that $t > s$. If $t = s + 1$, then x_s and x_{s+1} are adjacent variables and we can directly flip both x_s and x_{s+1}. In general (see Figure 10.1 for an illustration), we can flip two adjacent variables sequentially as follows. Let $h^{(0)} = f$ and for $1 \leq j \leq t - s$, define $h^{(j)}$ to be the result of flipping both variables x_{s+j-1} and x_{s+j}, i.e., $h^{(j)}(x_1^{(j)}, x_2^{(j)}, \ldots, x_n^{(j)}) = h^{(j-1)}(x_1^{(j-1)}, x_2^{(j-1)}, \ldots, x_n^{(j-1)})$, where $x_i^{(j)} = \overline{x_i^{(j-1)}}$ for $i \in \{s + j - 1, s + j\}$ and $x_i^{(j)} = x_i^{(j-1)}$ for all others. Then we are done by letting $g = h^{(t-s)}$. In effect, all variables x_i with $s < i < t$ are flipped twice. □

Lemma 10.2 is also valid for $\mathscr{P}$ and $\mathscr{M}$ and is useful in the proof of Theorem 10.1, but we will not use this more generalized form in the exposition here.

Lemma 10.3 implies that in Pl-Holant$(\widehat{\mathcal{EQ}}, f)$, where $f \notin \mathscr{A}$, we can assume that $f_{00\cdots0} = 1$. Moreover, if f satisfies the parity condition, we can assume it satisfies the *even* parity condition.

Lemma 10.3. *If $\widehat{\mathcal{F}}$ contains a signature $f \notin \mathscr{A}$ of arity n, then we can construct a function $f' \notin \mathscr{A}$ of arity n with $f'_{00\cdots0} = 1$ such that*

$$\text{Pl-Holant}(\widehat{\mathcal{EQ}}, f', \widehat{\mathcal{F}}) \leq_{\text{T}} \text{Pl-Holant}(\widehat{\mathcal{EQ}}, \widehat{\mathcal{F}}).$$

https://doi.org/10.1017/9781107477063.011 Published online by Cambridge University Press

Moreover, if f satisfies the parity condition, then f' satisfies the even *parity condition, and if f takes values in $\{0, 1\}$ ($\{0, 1, -1\}$), then f' also takes values in $\{0, 1\}$ ($\{0, 1, -1\}$).*

Proof. If $f_{00\cdots0} \neq 0$, then we simply normalize f by setting $f' = f/f_{00\cdots0}$. So we suppose $f_{00\cdots0} = 0$. By $f \notin \mathscr{A}$, clearly f is not identically 0. Let $\mathrm{wt}(\alpha) = \min_{\beta\in\{0,1\}^n}\{\mathrm{wt}(\beta)|f_\beta \neq 0\}$. Let $S = \{i \mid 1 \leq i \leq n,$ the ith bit of α is $0\}$. Since we have $[1, 0] \in \widehat{\mathcal{EQ}}$ we can get a nonzero multiple of $[0, 1]^{\otimes \mathrm{wt}(\alpha)}$, by connecting a copy of $[1, 0]$ to each variable of f in S. Depending on whether $\mathrm{wt}(\alpha)$ is odd or even, we can get either $[0, 1]$ or $[0, 1]^{\otimes 2}$ respectively, by taking zero or more self loops on $[0, 1]^{\otimes \mathrm{wt}(\alpha)}$. Then we can get either $[0, 1, 0]$ or $[0, 1, 0]^{\otimes 2}$ respectively, by connecting $[0, 1]$ or $[0, 1]^{\otimes 2}$ to one or two copies of $[1, 0, 1, 0] \in \widehat{\mathcal{EQ}}$.

If $\mathrm{wt}(\alpha)$ is odd, and this includes the case when f satisfies the odd parity condition, we have $[0, 1, 0]$ and can flip any variable of f individually. By flipping all variables in $[n] \setminus S$, and normalizing, we obtain f' with the required property. In particular, if f satisfies the odd parity condition, then f' satisfies the even parity condition.

If $\mathrm{wt}(\alpha)$ is even, and this includes the case when f satisfies the even parity condition, we have $[0, 1]^{\otimes 2}$ and $[0, 1, 0]^{\otimes 2}$. By Lemma 10.2 we can flip any two variables of f. By applying the construction in Lemma 10.2 simultaneously on $\mathrm{wt}(\alpha)/2$ pairs of variables of f, we can transform f to f' by a planar construction so that $f'_{00\cdots0} = f_\alpha \neq 0$. By normalizing, we obtain the required f' with $f'_{00\cdots0} = 1$. In particular, if f satisfies the even parity condition, then f' also satisfies the even parity condition. We get f' from f by flipping some variables in all cases. Thus if f takes values in $\{0, 1\}$ ($\{0, 1, -1\}$ respectively), then f' also takes values in $\{0, 1\}$ ($\{0, 1, -1\}$ respectively). □

Lemma 10.4. *Suppose all signatures in $\widehat{\mathcal{F}}$ take values in $\{0, 1\}$ and satisfy the parity condition. If $\widehat{\mathcal{F}}$ contains a signature $f \notin \mathscr{A}$ of arity $n \geq 3$, then there exists a signature $g \notin \mathscr{A}$ of arity $< n$, such that*

$$\text{Pl-Holant}(\widehat{\mathcal{EQ}}, g, \widehat{\mathcal{F}}) \leq_{\mathrm{T}} \text{Pl-Holant}(\widehat{\mathcal{EQ}}, \widehat{\mathcal{F}}).$$

Furthermore, g satisfies the even *parity condition.*

Proof. The proof is by induction on n, and uses the tableau calculus.

If there exists $i \in [n]$ such that $f^{x_i=0} \notin \mathscr{A}$, then we are done since we have $[1, 0] \in \widehat{\mathcal{EQ}}$, and connecting $[1, 0]$ to x_i of f gives $f^{x_i=0}$. In the following, we assume that $f^{x_i=0} \in \mathscr{A}$ for $1 \leq i \leq n$. By Lemma 10.3, we can assume that $f_{00\cdots0} = 1$ and f satisfies the *even* parity condition.

https://doi.org/10.1017/9781107477063.011 Published online by Cambridge University Press

As a $\{0,1\}$-valued signature, $f \in \mathscr{A}$ iff $\mathrm{supp}(f)$ is an affine subspace. So $\mathrm{supp}(f)$ is not an affine subspace, in particular not a linear subspace. There exist $\mathbf{a} = a_1 a_2 \cdots a_n$, $\mathbf{b} = b_1 b_2 \cdots b_n$, such that $\mathbf{a}, \mathbf{b} \in \mathrm{supp}(f)$, $\mathbf{c} = \mathbf{a} \oplus \mathbf{b} = c_1 c_2 \cdots c_n \notin \mathrm{supp}(f)$. Since $00\ldots0 \in \mathrm{supp}(f)$, $\mathbf{a} \neq \mathbf{b}$, so there exists $i \in [n]$ such that $a_i \neq b_i$. Without loss of generality, we assume that $a_1 = 0, b_1 = 1$. It follows that $c_1 = 1$. Let $\mathbf{a}' = a_3 \cdots a_n$, $\mathbf{b}' = b_3 \cdots b_n$, $\mathbf{c}' = c_3 \cdots c_n$.

By connecting the variables x_2, x_1 of $[1,0,1,0]$ to the variables x_1, x_2 of f respectively, the planar gadget gives the signature

$$g(x', x_3, x_4, \ldots, x_n) = \sum_{x_1, x_2 \in \{0,1\}} [1,0,1,0](x_2, x_1, x') f(x_1, x_2, \ldots, x_n).$$

If $x' = 0$ then the parity function $[1,0,1,0](x_2, x_1, x')$ forces $x_1 x_2 = 00$ or 11, and if $x' = 1$ then the parity function forces $x_1 x_2 = 01$ or 10. So we have

$$g(x_1 \oplus x_2, x_3, \ldots, x_n) = f(x_1, x_2, x_3, \ldots, x_n) + f(\bar{x}_1, \bar{x}_2, x_3, \ldots, x_n).$$

We show that $g \notin \mathscr{A}$.

Now comes the tableau calculus. We will examine all four cases of membership of $\bar{a}_1 \bar{a}_2 \mathbf{a}'$ and $\bar{b}_1 \bar{b}_2 \mathbf{b}'$ in $\mathrm{supp}(f)$.

- Suppose $\bar{a}_1 \bar{a}_2 \mathbf{a}' \in \mathrm{supp}(f)$ and $\bar{b}_1 \bar{b}_2 \mathbf{b}' \in \mathrm{supp}(f)$. By assumption $\mathbf{a} = 0 a_2 \mathbf{a}' \in \mathrm{supp}(f)$ and $0 \bar{b}_2 \mathbf{b}' = \bar{b}_1 \bar{b}_2 \mathbf{b}' \in \mathrm{supp}(f)$. Since $f^{x_1=0} \in \mathscr{A}$, $\mathrm{supp}(f^{x_1=0})$ is an affine subspace. As $0\ldots0 \in \mathrm{supp}(f^{x_1=0})$, $\mathrm{supp}(f^{x_1=0})$ is a linear subspace. We have

$$\begin{array}{rll} & a_2 \mathbf{a}' & \in \mathrm{supp}(f^{x_1=0}) \\ \oplus & \bar{b}_2 \mathbf{b}' & \in \mathrm{supp}(f^{x_1=0}) \\ \hline & \bar{c}_2 \mathbf{c}' & \end{array}$$

  Hence $\bar{c}_2 \mathbf{c}' \in \mathrm{supp}(f^{x_1=0})$. This means that $f_{\bar{c}_1 \bar{c}_2 \mathbf{c}'} \neq 0$ since $\bar{c}_1 = 0$. By definition

$$g_{(c_1 \oplus c_2)\mathbf{c}'} = f_{c_1 c_2 \mathbf{c}'} + f_{\bar{c}_1 \bar{c}_2 \mathbf{c}'}.$$

  Thus $g_{(c_1 \oplus c_2)\mathbf{c}'} = 1$ since $f_{c_1 c_2 \mathbf{c}'} = 0$ and $f_{\bar{c}_1 \bar{c}_2 \mathbf{c}'} = 1$ (as f is $\{0,1\}$-valued). Moreover, by

$$g_{(a_1 \oplus a_2)\mathbf{a}'} = f_{a_1 a_2 \mathbf{a}'} + f_{\bar{a}_1 \bar{a}_2 \mathbf{a}'},$$

  we have $g_{(a_1 \oplus a_2)\mathbf{a}'} = 2$, since $f_{a_1 a_2 \mathbf{a}'} = f_{\bar{a}_1 \bar{a}_2 \mathbf{a}'} = 1$. This implies that g is not affine, as g has nonzero values of different norm, a necessary condition to be in $\mathscr{A}$.

https://doi.org/10.1017/9781107477063.011 Published online by Cambridge University Press

- Suppose $\bar{a}_1\bar{a}_2\mathbf{a}' \notin \mathrm{supp}(f)$ and $\bar{b}_1\bar{b}_2\mathbf{b}' \notin \mathrm{supp}(f)$. By

$$g_{00\cdots0} = f_{000\cdots0} + f_{110\cdots0},$$

we have $g_{00\cdots0} \neq 0$ since $f_{000\cdots0} = 1$ and $f_{110\cdots0} \in \{0, 1\}$. Suppose g is affine, for otherwise we are done. Then $\mathrm{supp}(g)$ is a linear subspace, since $00\cdots0 \in \mathrm{supp}(g)$. By

$$g_{(a_1\oplus a_2)\mathbf{a}'} = f_{a_1a_2\mathbf{a}'} + f_{\bar{a}_1\bar{a}_2\mathbf{a}'},$$
$$g_{(b_1\oplus b_2)\mathbf{b}'} = f_{b_1b_2\mathbf{b}'} + f_{\bar{b}_1\bar{b}_2\mathbf{b}'},$$

we have $g_{(a_1\oplus a_2)\mathbf{a}'} = 1$, $g_{(b_1\oplus b_2)\mathbf{b}'} = 1$ since $f_{a_1a_2\mathbf{a}'} = f_{b_1b_2\mathbf{b}'} = 1$ and $f_{\bar{a}_1\bar{a}_2\mathbf{a}'} = f_{\bar{b}_1\bar{b}_2\mathbf{b}'} = 0$. This implies that $(a_1 \oplus a_2)\mathbf{a}', (b_1 \oplus b_2)\mathbf{b}' \in \mathrm{supp}(g)$. Then by

$$\begin{array}{rll} & (a_1 \oplus a_2)\mathbf{a}' & \in \mathrm{supp}(g) \\ \oplus & (b_1 \oplus b_2)\mathbf{b}' & \in \mathrm{supp}(g) \\ \hline & (c_1 \oplus c_2)\mathbf{c}' & \end{array}$$

we have $(c_1 \oplus c_2)\mathbf{c}' \in \mathrm{supp}(g)$. This implies that

$$g_{(c_1\oplus c_2)\mathbf{c}'} = f_{c_1c_2\mathbf{c}'} + f_{\bar{c}_1\bar{c}_2\mathbf{c}'} \neq 0.$$

Thus $f_{\bar{c}_1\bar{c}_2\mathbf{c}'} \neq 0$ since $f_{c_1c_2\mathbf{c}'} = 0$. So $\bar{c}_1\bar{c}_2\mathbf{c}' \in \mathrm{supp}(f)$. Hence, $\bar{c}_2\mathbf{c}' \in \mathrm{supp}(f^{x_1=0})$ since $\bar{c}_1 = 0$. By

$$\begin{array}{rll} & a_2\mathbf{a}' & \in \mathrm{supp}(f^{x_1=0}) \\ \oplus & \bar{c}_2\mathbf{c}' & \in \mathrm{supp}(f^{x_1=0}) \\ \hline & \bar{b}_2\mathbf{b}' & \end{array}$$

we have $\bar{b}_2\mathbf{b}' \in \mathrm{supp}(f^{x_1=0})$. Thus $\bar{b}_1\bar{b}_2\mathbf{b}' \in \mathrm{supp}(f)$ as $\bar{b}_1 = 0$. This is a contradiction.

- Suppose $\bar{a}_1\bar{a}_2\mathbf{a}' \in \mathrm{supp}(f)$ and $\bar{b}_1\bar{b}_2\mathbf{b}' \notin \mathrm{supp}(f)$. Then we consider

$$\begin{aligned} g_{(a_1\oplus a_2)\mathbf{a}'} &= f_{a_1a_2\mathbf{a}'} + f_{\bar{a}_1\bar{a}_2\mathbf{a}'}, \\ g_{(b_1\oplus b_2)\mathbf{b}'} &= f_{b_1b_2\mathbf{b}'} + f_{\bar{b}_1\bar{b}_2\mathbf{b}'}. \end{aligned} \tag{10.2}$$

This implies that $g_{(a_1\oplus a_2)\mathbf{a}'} = 2$ and $g_{(b_1\oplus b_2)\mathbf{b}'} = 1$. Thus $g \notin \mathscr{A}$.

- Suppose $\bar{a}_1\bar{a}_2\mathbf{a}' \notin \mathrm{supp}(f)$ and $\bar{b}_1\bar{b}_2\mathbf{b}' \in \mathrm{supp}(f)$. Then (10.2) implies that $g_{(a_1\oplus a_2)\mathbf{a}'} = 1$ and $g_{(b_1\oplus b_2)\mathbf{b}'} = 2$. Thus $g \notin \mathscr{A}$. □

Looking at the proof of Lemma 10.4 closely, it becomes clear that to be able to move a support point to $00\cdots0$ using Lemma 10.3 simplifies greatly the argument in the tableau calculus. For affine spaces V, the closure property is that whenever $\mathbf{a}, \mathbf{b}, \mathbf{c} \in V$, we have $\mathbf{a} \oplus \mathbf{b} \oplus \mathbf{c} \in V$. But for linear spaces V, the closure property is simply whenever $\mathbf{a}, \mathbf{b} \in V$, we have $\mathbf{a} \oplus \mathbf{b} \in V$. Instead of

https://doi.org/10.1017/9781107477063.011 Published online by Cambridge University Press

having to deal with $\mathrm{supp}(f)$ as an affine space, by moving the support point to $00\cdots0$ we can deal with linear space, and as such we only need to have two known points in $\mathrm{supp}(f)$ to obtain another point in $\mathrm{supp}(f)$.

In order to prove Lemma 10.10, it is helpful to have some facts about multilinear representation of signatures. Lemma 10.5 shows that for a $\{1,-1\}$-valued signature of arity k, there exists a unique multilinear polynomial $P(x_1,\ldots,x_k)\in\mathbb{Z}_2[x_1,\ldots,x_k]$ such that

$$f(x_1,\ldots x_k)=(-1)^{P(x_1,\ldots,x_k)}. \tag{10.3}$$

Thus if (10.3) holds where P has total degree >2, then $f\notin\mathscr{A}$.

Lemma 10.5. *If f is a signature of arity k taking values in $\{\pm1\}$, then there exists a unique multilinear polynomial $P(x_1,\ldots,x_k)\in\mathbb{Z}_2[x_1,\ldots,x_k]$ such that $f(x_1,\ldots x_k)=(-1)^{P(x_1,\ldots,x_k)}$.*

Proof. For any input $\alpha=a_1\ldots a_k\in\{0,1\}^k$, there exists $r_\alpha\in\{0,1\}$ such that $f_\alpha=(-1)^{r_\alpha}$ since f takes values in $\{1,-1\}$. Let $P(x_1,\ldots,x_k)=\sum_{\alpha\in\{0,1\}^k} r_\alpha\prod_{i=1}^k\widetilde{x_i}\in\mathbb{Z}_2[x_1,\ldots,x_k]$, where $\alpha=a_1\ldots a_k$, $\widetilde{x_i}=x_i$ if $a_i=1$ and $1-x_i$ if $a_i=0$. Then $f(x_1,\ldots x_k)=(-1)^{P(x_1,\ldots,x_k)}$.

To prove uniqueness, it is equivalent to prove that if f is the constant 1 function then $P(x_1,\ldots,x_k)=0$ in $\mathbb{Z}_2[x_1,\ldots,x_k]$. For a contradiction suppose $\prod_{i\in S}x_i$ is a nonzero term in $P(x_1,\ldots,x_k)$ with minimum cardinality $|S|$. Set $x_i=1$ for all $i\in S$, and all other $x_i=0$. Then P evaluates to 1 in $\mathbb{Z}_2$, and f evaluates to $-1\neq1$. This is a contradiction. □

Corollary 10.6. *Let $f\in\mathscr{A}$ be a $\{1,-1\}$-valued signature of arity n with support of dimension $k\geq0$. Let $Y=\{x_{i_1},\ldots,x_{i_k}\}$ be a set of free variables. Then there exists a unique $Q(Y)\in\mathbb{Z}_2[Y]$ such that*

$$f(x_1,x_2,\ldots,x_n)=(-1)^{Q(Y)}$$

for $(x_1,x_2,\ldots,x_n)\in\mathrm{supp}(f)$, where $Q(Y)$ is a multilinear polynomial of degree ≤2.

Corollary 10.7. *Let f be a binary signature and $M_{x_1,x_2}(f)=\begin{bmatrix}f_{00}&f_{01}\\f_{10}&f_{11}\end{bmatrix}=\begin{bmatrix}1&b\\c&d\end{bmatrix}$. If $b,c,d\in\{1,-1\}$ then $f\in\mathscr{A}$.*

Definition 10.8. If f has affine support of dimension k, and $Y=\{x_{i_1},\ldots,x_{i_k}\}$ is a set of free variables, then the *compressed signature* of f (with respect to Y)

https://doi.org/10.1017/9781107477063.011 Published online by Cambridge University Press

is the unique function $\underline{f}$ defined on Y such that

$$\underline{f}(x_{i_1}, \ldots, x_{i_k}) = f(x_1, x_2, \ldots, x_n),$$

where $(x_1, x_2, \ldots, x_n) \in \mathrm{supp}(f)$.

Corollary 10.9. *Let f be a signature of arity n having affine support of dimension k. Suppose f takes values in $\{0, \pm 1\}$, and $Y = \{x_{i_1}, \ldots, x_{i_k}\}$ is a set of free variables. Then $f \in \mathscr{A}$ iff $\underline{f} \in \mathscr{A}$.*

Proof. If $\underline{f}(x_{i_1}, \ldots, x_{i_k}) = (-1)^{Q(x_{i_1}, \ldots, x_{i_k})}$, then

$$f(x_1, \ldots, x_n) = \chi_{\mathrm{supp}(f)}(x_1, \ldots, x_n) \cdot (-1)^{Q(x_{i_1}, \ldots, x_{i_k})}.$$

So $f \in \mathscr{A}$ iff $\underline{f} \in \mathscr{A}$. □

Now we are ready to prove Lemma 10.10. We will use the tableau calculus again, which is used numerous times in the proof of Theorem 10.1. Another technique is an integrality gap (0 mod 4 versus 2 mod 4) that guarantees that a function is not in $\mathscr{A}$. Here the arity reduction is more difficult, and we need the help of another auxiliary function $[1, 0, -1]$. When we use it in the proof of Theorem 10.1, this auxiliary function will have to be produced by other means.

Lemma 10.10. *Suppose all signatures in $\widehat{\mathcal{F}}$ take values in $\{0, 1, -1\}$ and satisfy the parity condition. If $\widehat{\mathcal{F}}$ contains the binary signature $[1, 0, -1]$ and a signature $f \notin \mathscr{A}$ of arity $n \geq 3$, then there exists a signature $h \notin \mathscr{A}$ of arity $< n$, such that*

$$\text{Pl-Holant}(\widehat{\mathcal{EQ}}, h, \widehat{\mathcal{F}}) \leq_{\mathrm{T}} \text{Pl-Holant}(\widehat{\mathcal{EQ}}, \widehat{\mathcal{F}}).$$

Furthermore, h satisfies the even *parity condition.*

Proof. Again the proof is by induction on n. As in the proof of Lemma 10.4 we may assume that $f^{x_i=0} \in \mathscr{A}$ for $1 \leq i \leq n$, $f_{00\cdots0} = 1$ and f satisfies the *even* parity condition.

♣ Step 1: Suppose $\mathrm{supp}(f)$ is not an affine subspace.

Being not an affine subspace, $\mathrm{supp}(f)$ is certainly not a linear subspace. Thus there exist $\mathbf{a} = a_1 a_2 \cdots a_n$, $\mathbf{b} = b_1 b_2 \cdots b_n$, such that $\mathbf{a}, \mathbf{b} \in \mathrm{supp}(f)$, $\mathbf{c} = \mathbf{a} \oplus \mathbf{b} = c_1 c_2 \cdots c_n \notin \mathrm{supp}(f)$. Since $f_{00\cdots0} = 1$, i.e., $00\cdots0 \in \mathrm{supp}(f)$, we have $\mathbf{a} \neq \mathbf{b}$. Without loss of generality, we assume that $a_1 = 0, b_1 = 1$. It follows that $c_1 = 1$. Let $\mathbf{a}' = a_3 \cdots a_n$, $\mathbf{b}' = b_3 \cdots b_n$, $\mathbf{c}' = c_3 \cdots c_n$.

https://doi.org/10.1017/9781107477063.011 Published online by Cambridge University Press

Connecting the first variable of $[1, 0, -1]$ to the first variable of f, this gives the signature

$$\widetilde{f}(x_1, x_2, \ldots, x_n) = \sum_{x_1' \in \{0,1\}} [1, 0, -1](x_1', x_1) f(x_1', x_2, \ldots, x_n).$$

Since $[1, 0, -1]$ is a weighted binary equality that takes value 1 on 00 and -1 on 11,

$$\widetilde{f}(x_1, x_2, \ldots, x_n) = (-1)^{x_1} f(x_1, x_2, \ldots, x_n).$$

Moreover, by connecting the variables x_2, x_1 of $[1, 0, 1, 0] \in \widehat{\mathcal{EQ}}$ to the variables x_1, x_2 of f respectively, the planar construction gives the signature

$$h(x', x_3, x_4, \ldots, x_n) = \sum_{x_1, x_2 \in \{0,1\}} [1, 0, 1, 0](x_2, x_1, x') f(x_1, x_2, \ldots, x_n).$$

As in the proof of Lemma 10.4, we have

$$h(x_1 \oplus x_2, x_3, \ldots, x_n) = f(x_1, x_2, x_3, \ldots, x_n) + f(\bar{x}_1, \bar{x}_2, x_3, \ldots, x_n).$$

Similarly, by connecting the variables x_2, x_1 of $[1, 0, 1, 0]$ to the variables x_1, x_2 of $\widetilde{f}$ respectively, the planar gadget gives the signature $\widetilde{h}(x', x_3, x_4, \ldots, x_n)$ with

$$\begin{aligned}\widetilde{h}(x_1 \oplus x_2, x_3, \ldots, x_n) = {} & (-1)^{x_1} f(x_1, x_2, x_3, \ldots, x_n) \\ & + (-1)^{\bar{x}_1} f(\bar{x}_1, \bar{x}_2, x_3, \ldots, x_n).\end{aligned}$$

Since $f^{x_i=0} \in \mathscr{A}$ and affine functions have affine support, and furthermore $f^{x_i=0}_{0\cdots0} = 1$, we conclude that $\mathrm{supp}(f^{x_i=0})$ is a linear subspace of $\mathbb{Z}_2^{n-1}$. Recall that all nonzero values of an affine signature have the same norm.

Next comes the tableau calculus. We will examine all four cases of membership of $\bar{a}_1\bar{a}_2\mathbf{a}'$ and $\bar{b}_1\bar{b}_2\mathbf{b}'$ in $\mathrm{supp}(f)$. In each case we prove that either h or $\widetilde{h}$ is not affine with arity $n-1$.

- Suppose $\bar{a}_1\bar{a}_2\mathbf{a}' \in \mathrm{supp}(f)$ and $\bar{b}_1\bar{b}_2\mathbf{b}' \in \mathrm{supp}(f)$. By assumption $\mathbf{a} = 0a_2\mathbf{a}' \in \mathrm{supp}(f)$ and $0\bar{b}_2\mathbf{b}' = \bar{b}_1\bar{b}_2\mathbf{b}' \in \mathrm{supp}(f)$, and since $\mathrm{supp}(f^{x_1=0})$ is a linear subspace, we have

$$\begin{array}{rll} & a_2\mathbf{a}' & \in \mathrm{supp}(f^{x_1=0}) \\ \oplus & \bar{b}_2\mathbf{b}' & \in \mathrm{supp}(f^{x_1=0}) \\ \hline & \bar{c}_2\mathbf{c}' & \end{array}.$$

https://doi.org/10.1017/9781107477063.011 Published online by Cambridge University Press

Hence $\bar{c}_2\mathbf{c}' \in \mathrm{supp}(f^{x_1=0})$. This means that $f_{\bar{c}_1\bar{c}_2\mathbf{c}'} \neq 0$ since $\bar{c}_1 = 0$. By definition of h and $\widetilde{h}$,

$$\begin{aligned} h_{(c_1\oplus c_2)\mathbf{c}'} &= \quad f_{c_1c_2\mathbf{c}'} + f_{\bar{c}_1\bar{c}_2\mathbf{c}'}, \\ \widetilde{h}_{(c_1\oplus c_2)\mathbf{c}'} &= -f_{c_1c_2\mathbf{c}'} + f_{\bar{c}_1\bar{c}_2\mathbf{c}'}. \end{aligned}$$

Thus $|h_{(c_1\oplus c_2)\mathbf{c}'}| = |\widetilde{h}_{(c_1\oplus c_2)\mathbf{c}'}| = 1$ since $f_{c_1c_2\mathbf{c}'} = 0$ and $f_{\bar{c}_1\bar{c}_2\mathbf{c}'} = \pm 1$. Moreover, by

$$\begin{aligned} h_{(a_1\oplus a_2)\mathbf{a}'} &= f_{a_1a_2\mathbf{a}'} + f_{\bar{a}_1\bar{a}_2\mathbf{a}'}, \\ \widetilde{h}_{(a_1\oplus a_2)\mathbf{a}'} &= f_{a_1a_2\mathbf{a}'} - f_{\bar{a}_1\bar{a}_2\mathbf{a}'}, \end{aligned} \tag{10.4}$$

if both $h_{(a_1\oplus a_2)\mathbf{a}'}$ and $\widetilde{h}_{(a_1\oplus a_2)\mathbf{a}'}$ are 0, then $f_{a_1a_2\mathbf{a}'} = 0$, by treating (10.4) as a linear equation system. This contradicts that $\mathbf{a} \in \mathrm{supp}(f)$. Thus we have $h_{(a_1\oplus a_2)\mathbf{a}'} \neq 0$ or $\widetilde{h}_{(a_1\oplus a_2)\mathbf{a}'} \neq 0$. Both $f_{a_1a_2\mathbf{a}'} \neq 0$ and $f_{\bar{a}_1\bar{a}_2\mathbf{a}'} \neq 0$ by assumption, and since f takes values in $\{0, 1, -1\}$, we have $f_{a_1a_2\mathbf{a}'}, f_{\bar{a}_1\bar{a}_2\mathbf{a}'} \in \{1, -1\}$. This implies that one of $|h_{(a_1\oplus a_2)\mathbf{a}'}|$ or $|\widetilde{h}_{(a_1\oplus a_2)\mathbf{a}'}|$ is 2. So h or $\widetilde{h}$ is not affine, because at least one of them has nonzero values of unequal norm.

- Suppose $\bar{a}_1\bar{a}_2\mathbf{a}' \notin \mathrm{supp}(f)$ and $\bar{b}_1\bar{b}_2\mathbf{b}' \notin \mathrm{supp}(f)$. By treating the following as a linear system

$$\begin{aligned} h_{00\cdots0} &= f_{000\cdots0} + f_{110\cdots0}, \\ \widetilde{h}_{00\cdots0} &= f_{000\cdots0} - f_{110\cdots0}, \end{aligned} \tag{10.5}$$

we have $h_{00\cdots0} \neq 0$ or $\widetilde{h}_{00\cdots0} \neq 0$ since $f_{000\cdots0} \neq 0$. Without loss of generality, we assume that $h_{00\cdots0} \neq 0$. The same argument can be applied to $\widetilde{h}$ if $\widetilde{h}_{00\cdots0} \neq 0$. Suppose h is affine, for otherwise we are done. Since $00\cdots0 \in \mathrm{supp}(h)$, $\mathrm{supp}(h)$ is a linear subspace. By

$$\begin{aligned} h_{(a_1\oplus a_2)\mathbf{a}'} &= f_{a_1a_2\mathbf{a}'} + f_{\bar{a}_1\bar{a}_2\mathbf{a}'}, \\ h_{(b_1\oplus b_2)\mathbf{b}'} &= f_{b_1b_2\mathbf{b}'} + f_{\bar{b}_1\bar{b}_2\mathbf{b}'}, \end{aligned}$$

we have $|h_{(a_1\oplus a_2)\mathbf{a}'}| = 1$, $|h_{(b_1\oplus b_2)\mathbf{b}'}| = 1$ since $f_{\bar{a}_1\bar{a}_2\mathbf{a}'} = f_{\bar{b}_1\bar{b}_2\mathbf{b}'} = 0$ and $f_{a_1a_2\mathbf{a}'}, f_{b_1b_2\mathbf{b}'} \in \{1, -1\}$. This implies that $(a_1 \oplus a_2)\mathbf{a}', (b_1 \oplus b_2)\mathbf{b}' \in \mathrm{supp}(h)$. Then by

$$\begin{array}{rl} & (a_1 \oplus a_2)\mathbf{a}' \quad \in \mathrm{supp}(h) \\ \oplus & (b_1 \oplus b_2)\mathbf{b}' \quad \in \mathrm{supp}(h) \\ \hline & (c_1 \oplus c_2)\mathbf{c}' \end{array}$$

we have $(c_1 \oplus c_2)\mathbf{c}' \in \mathrm{supp}(h)$. This implies that

$$h_{(c_1\oplus c_2)\mathbf{c}'} = f_{c_1c_2\mathbf{c}'} + f_{\bar{c}_1\bar{c}_2\mathbf{c}'} \neq 0.$$

https://doi.org/10.1017/9781107477063.011 Published online by Cambridge University Press

Thus $f_{\bar{c}_1\bar{c}_2\mathbf{c}'} \neq 0$ since $f_{c_1c_2\mathbf{c}'} = 0$. So $\bar{c}_1\bar{c}_2\mathbf{c}' \in \mathrm{supp}(f)$. Therefore, $\bar{c}_2\mathbf{c}' \in \mathrm{supp}(f^{x_1=0})$ since $\bar{c}_1 = 0$. Recall that by assumption $\mathbf{a} = 0a_2\mathbf{a}' \in \mathrm{supp}(f)$. By

$$\begin{array}{rll} & a_2\mathbf{a}' & \in \mathrm{supp}(f^{x_1=0}) \\ \oplus & \bar{c}_2\mathbf{c}' & \in \mathrm{supp}(f^{x_1=0}) \\ \hline & \overline{b_2}\mathbf{b}' & \end{array}$$

we have $\bar{b}_2\mathbf{b}' \in \mathrm{supp}(f^{x_1=0})$. Thus $\bar{b}_1\bar{b}_2\mathbf{b}' \in \mathrm{supp}(f)$ as $\bar{b}_1 = 0$. This is a contradiction.

- Suppose $\bar{a}_1\bar{a}_2\mathbf{a}' \in \mathrm{supp}(f)$ and $\bar{b}_1\bar{b}_2\mathbf{b}' \notin \mathrm{supp}(f)$. By treating the following as a linear system,

$$\begin{aligned} h_{(b_1\oplus b_2)\mathbf{b}'} &= \quad f_{b_1b_2\mathbf{b}'} + f_{\bar{b}_1\bar{b}_2\mathbf{b}'}, \\ \widetilde{h}_{(b_1\oplus b_2)\mathbf{b}'} &= -f_{b_1b_2\mathbf{b}'} + f_{\bar{b}_1\bar{b}_2\mathbf{b}'}, \end{aligned}$$

we have $|h_{(b_1\oplus b_2)\mathbf{b}'}| = |\widetilde{h}_{(b_1\oplus b_2)\mathbf{b}'}| = 1$ since $f_{b_1b_2\mathbf{b}'} = \pm 1$ and $f_{\bar{b}_1\bar{b}_2\mathbf{b}'} = 0$. Then by

$$\begin{aligned} h_{(a_1\oplus a_2)\mathbf{a}'} &= f_{a_1a_2\mathbf{a}'} + f_{\bar{a}_1\bar{a}_2\mathbf{a}'}, \\ \widetilde{h}_{(a_1\oplus a_2)\mathbf{a}'} &= f_{a_1a_2\mathbf{a}'} - f_{\bar{a}_1\bar{a}_2\mathbf{a}'}, \end{aligned}$$

and $f_{a_1a_2\mathbf{a}'}, f_{\bar{a}_1\bar{a}_2\mathbf{a}'} \in \{1, -1\}$, we have $|h_{(a_1\oplus a_2)\mathbf{a}'}| = 2$ or $|\widetilde{h}_{(a_1\oplus a_2)\mathbf{a}'}| = 2$. This implies that h or $\widetilde{h}$ is not affine, since one of them has nonzero values of unequal norm.
- The proof for the case $\bar{a}_1\bar{a}_2\mathbf{a}' \notin \mathrm{supp}(f)$ and $\bar{b}_1\bar{b}_2\mathbf{b}' \in \mathrm{supp}(f)$ is symmetric.

♣ Step 2: Now we may assume that $\mathrm{supp}(f)$ is an affine subspace with dimension $k \geq 0$.

Let Y be a set of k free variables, $Y = \{y_1, y_2, \ldots, y_k\} = \{x_{i_1}, x_{i_2} \ldots, x_{i_k}\}$. Let $\underline{f}(Y)$ be the compressed signature of f. By Corollary 10.9, $\underline{f} \notin \mathscr{A}$. However, for $\{1, -1\}$-valued $\underline{f}$, if $k = 2$, then $\underline{f} \in \mathscr{A}$ by Corollary 10.7, a contradiction. Also trivially $\underline{f} \in \mathscr{A}$ if $k < 2$. Hence $k \geq 3$.

By Lemma 10.5, and since $\underline{f}$ takes values in $\{1, -1\}$, there exists a unique multilinear polynomial $Q(y_1, y_2, \ldots, y_k) \in \mathbb{Z}_2[X]$ such that $\underline{f}(y_1, y_2, \ldots, y_k) = (-1)^{Q(y_1,y_2,\ldots,y_k)}$. $\underline{f}$ is affine iff $Q(y_1, y_2, \ldots, y_k)$ is a multilinear polynomial of total degree ≤ 2.

If $k \geq 4$ and there exists a term $y_{i_1}y_{i_2}\cdots y_{i_s}$ with nonzero coefficient, where $3 \leq s < k$, then there exists some $y_j \notin \{y_{i_1}, y_{i_2}, \ldots, y_{i_s}\}$, such that

$$\underline{f}^{y_j=0}(y_1, \ldots, \widehat{y_j}, \ldots, y_k) = (-1)^{Q'(y_1,\ldots,\widehat{y_j},\ldots,y_k)},$$

https://doi.org/10.1017/9781107477063.011 Published online by Cambridge University Press

where Q' is a polynomial on $k-1$ variables ($\widehat{y_j}$ denotes the absence of y_j), where the term $y_{i_1}y_{i_2}\cdots y_{i_s}$ of degree $s>2$ still appears. This implies that $f^{y_j=0}$ is not affine. This is a contradiction. Thus we may assume that $Q(y_1, y_2, \ldots, y_k) = P(y_1, y_2, \ldots, y_k) + ay_1y_2\cdots y_k$, where $P(y_1, y_2, \ldots, y_k) \in \mathbb{Z}_2[X]$ is a multilinear polynomial of total degree ≤ 2, and $a \in \mathbb{Z}_2$. Note that this statement is also vacuously true if $k = 3$. If $a = 0$, then f is affine. This is a contradiction. Otherwise, $Q(y_1, y_2, \ldots, y_k) = P(y_1, y_2, \ldots, y_k) + y_1y_2\cdots y_k$. Moreover, by connecting the first variable of $[1, 0, -1]$ to y_i of f, the gadget gives the signature f' such that

$$f'(x_1, x_2, \ldots, x_n) = (-1)^{y_i} f(x_1, x_2, \ldots, x_n).$$

This implies that f' has the same support of f and

$$\underline{f'}(y_1, y_2, \ldots, y_k) = (-1)^{y_i + P(y_1, y_2, \ldots, y_k) + y_1y_2\cdots y_k},$$

where $\underline{f'}$ is the compressed signature of f' for Y. Thus $f' \notin \mathscr{A}$. This implies that we can add a linear term to $P(y_1, y_2, \ldots, y_k)$ in $\mathbb{Z}_2[X]$ freely.

We connect all variables of f except for y_1 to $n-1$ variables of the even parity function $\frac{1}{2}\{[1, 1]^{\otimes n} + [1, -1]^{\otimes n}\} = [1, 0, 1, \ldots, 0 \text{ (or } 1)] \in \widehat{\mathcal{EQ}}$. This planar construction gives a binary signature f^* that also has even parity, since both f and $[1, 0, 1, \ldots, 0 \text{ (or } 1)]$ have even parity. Hence $f^*_{01} = f^*_{10} = 0$, and the binary signature f^* has the form $[f^*_{00}, 0, f^*_{11}]$. In particular, f^* is symmetric. We have

$$\begin{aligned} f^*_{00} &= \sum_{y_2, y_3, \ldots, y_k \in \{0,1\}} \underline{f}^{y_1=0}(y_2, y_3, \ldots, y_k), \\ f^*_{11} &= \sum_{y_2, y_3, \ldots, y_k \in \{0,1\}} \underline{f}^{y_1=1}(y_2, y_3, \ldots, y_k). \end{aligned} \tag{10.6}$$

These values are as stated because for any $y_1 = 0, 1$ as the first input bit to f^*, there is exactly one term $(x_1, x_2, \ldots x_n) \in \mathrm{supp}(f)$ in the sum defining f^*, for every $(y_1, y_2, \ldots, y_k)$. Because f has even parity, $\sum_{i=1}^n x_i \equiv 0 \bmod 2$, thus the sum of the internal $n-1$ edges has the same parity as y_1. On the other hand, the even parity function $[1, 0, 1, \ldots, 0 \text{ (or } 1)]$ ensures that this sum of internal edges has the same parity as the second input bit of f^*.

Now we take a closer look at f^*. First, we consider the special case that the coefficient of y_iy_j in $P(y_1, y_2, \ldots, y_k)$ is nonzero for all $1 \le i < j \le k$.

- If $k = 3$, we may assume that $P(y_1, y_2, y_3) = y_1 + y_2 + y_3 + y_1y_2 + y_1y_3 + y_2y_3$ since we can add linear terms to $P(y_1, y_2, \ldots, y_k)$ at will. Then we have $\underline{f}(y_1, y_2, y_3) = (-1)^{P(y_1, y_2, y_3) + y_1y_2y_3}$. The polynomial $P(y_1, y_2, y_3) + y_1y_2y_3 = 1 + (1 + y_1)(1 + y_2)(1 + y_3) \in \mathbb{Z}_2[y_1, y_2, y_3]$ corresponds to the

https://doi.org/10.1017/9781107477063.011 Published online by Cambridge University Press

Or function on three bits, $y_1 \vee y_2 \vee y_3$. Thus the signature matrix

$$M_{y_1,y_2y_3}(\underline{f}) = \begin{bmatrix} 1 & -1 & -1 & -1 \\ -1 & -1 & -1 & -1 \end{bmatrix}.$$

Thus $f^* = [-2, 0, -4]$, which has nonzero terms of unequal norm, thus not in $\mathscr{A}$ and we are done.

- For $k \geq 4$, we may assume that $P(y_1, y_2, \ldots, y_k)$ has no linear terms since we can add linear terms to $P(y_1, y_2, \ldots, y_k)$ freely. Since P has all terms $y_i y_j$, both $\underline{f}^{y_1=0}$ and $\underline{f}^{y_1=1}$ are symmetric signatures. For $\underline{f}^{y_1=0}$, the entry of Hamming weight ℓ is $(\underline{f}^{y_1=0})_\ell = (-1)^{\frac{\ell(\ell-1)}{2}}$ for $0 \leq \ell \leq k-1$. For $\underline{f}^{y_1=1}$, we have $(\underline{f}^{y_1=1})_\ell = (-1)^{\frac{\ell(\ell+1)}{2}}$ for $0 \leq \ell \leq k-2$ and $(\underline{f}^{y_1=1})_{k-1} = (-1)^{\frac{k(k-1)}{2}+1}$. This implies that

$$\begin{aligned}
\underline{f}^{y_1=0} &= [1, 1, -1, -1, \ldots, (-1)^{\frac{(k-1)(k-2)}{2}}] \\
&= \frac{1}{1+\mathrm{i}} \left\{ [1, \mathrm{i}]^{\otimes k-1} + \mathrm{i}[1, -\mathrm{i}]^{\otimes k-1} \right\}, \\
\underline{f}^{y_1=1} &= [1, -1, -1, 1, \ldots, (-1)^{\frac{k(k-1)}{2}}] - 2(-1)^{\frac{k(k-1)}{2}}[0, 1]^{\otimes k-1} \\
&= \frac{1}{1-\mathrm{i}} \left\{ [1, \mathrm{i}]^{\otimes k-1} - \mathrm{i}[1, -\mathrm{i}]^{\otimes k-1} \right\} - 2(-1)^{\frac{k(k-1)}{2}}[0, 1]^{\otimes k-1}.
\end{aligned}$$

Thus

$$\begin{aligned}
f^*_{00} = \sum_{\beta \in \{0,1\}^{k-1}} (\underline{f}^{y_1=0})_\beta &= \frac{1}{1+\mathrm{i}} \sum_{w=0}^{k-1} \binom{k-1}{w} [\mathrm{i}^w + \mathrm{i}(-\mathrm{i})^w] \\
&= \frac{1}{1+\mathrm{i}} [(1+\mathrm{i})^{k-1} + \mathrm{i}(1-\mathrm{i})^{k-1}] \\
&= (1+\mathrm{i})^{k-2} + (1-\mathrm{i})^{k-2} \\
&= 2^{\frac{k}{2}} \cos((k-2)\pi/4), \\
f^*_{11} = \sum_{\beta \in \{0,1\}^{k-1}} (\underline{f}^{y_1=1})_\beta &= \frac{1}{1-\mathrm{i}} \sum_{w=0}^{k-1} \binom{k-1}{w} [\mathrm{i}^w - \mathrm{i}(-\mathrm{i})^w] - 2(-1)^{\frac{k(k-1)}{2}} \\
&= \frac{1}{1-\mathrm{i}} [(1+\mathrm{i})^{k-1} - \mathrm{i}(1-\mathrm{i})^{k-1}] - 2(-1)^{\frac{k(k-1)}{2}} \\
&= -2^{\frac{k}{2}} \sin((k-2)\pi/4) - 2(-1)^{\frac{k(k-1)}{2}}.
\end{aligned}$$

For $k \equiv 1 \mod 2$, $|f^*_{00}| = 2^{\frac{k-1}{2}}$, and $|f^*_{11}| = 2^{\frac{k-1}{2}} \pm 2$ (since $k \geq 4$). Therefore $f^*_{11} f^*_{00} \neq 0$ and $|f^*_{11}| \neq |f^*_{00}|$ (again using $k \geq 4$). Thus $f^* \notin \mathscr{A}$ and we are done.

https://doi.org/10.1017/9781107477063.011 Published online by Cambridge University Press

For $k \equiv 0 \mod 4$, $f^*_{00} = 0, |f^*_{11}| = 2^{\frac{k}{2}} \pm 2 \neq 0$. This implies that we have $f^* = f^*_{11}[0,1]^{\otimes 2}$. Using $[1,0,-1], [0,1]^{\otimes 2}$ and $f \notin \mathscr{A}$, by a separate lemma we can get a binary signature of the form $[x,0,y] \notin \mathscr{A}$. (The proof of this separate lemma is nontrivial and is omitted here; see [CF16] for details.)

For $k \equiv 2 \mod 4$, $|f^*_{00}| = 2^{\frac{k}{2}} \geq 4$ since $k \geq 4$, and $|f^*_{11}| = 2$, so $f^* \notin \mathscr{A}$ and we are done.

Now we assume that there exist $i \neq j \in [k]$ such that the coefficient of $y_i y_j$ is 0 in $P(y_1, y_2, \ldots, y_k)$. For notational simplicity, without loss of generality we assume that $i = k-1$, $j = k$. Then we can assume that (with the linear term y_{k-1} and y_k removed if needed)

$$P(y_1, y_2, \ldots, y_k) = y_1(L_1 + \epsilon_1) + y_2(L_2 + \epsilon_2) + \cdots + y_{k-2}(L_{k-2} + \varepsilon_{k-2}),$$

where

$$L_1 = \sum_{i=2}^{k} a_{1i} y_1, \quad L_2 = \sum_{i=3}^{k} a_{2i} y_i, \quad \ldots, \quad L_{k-2} = \sum_{i=k-1}^{k} a_{(k-2)i} y_i$$

and $a_{ji} \in \mathbb{Z}_2$ are fixed, but we can choose $\epsilon_i \in \mathbb{Z}_2$ freely since we can add linear terms freely. Let $F_{(0)} = \underline{f}$, $F_{(i)} = \underline{f}^{y_1=0, y_2=0, \ldots, y_i=0}$ for $i \in [k-2]$. We claim that there exist $\epsilon_1, \epsilon_2, \ldots, \epsilon_{k-2} \in \mathbb{Z}_2$ such that

$$\sum_{y_{i+1}, y_{i+2}, \ldots, y_k \in \{0,1\}} F_{(i)}(y_{i+1}, y_{i+2}, \ldots, y_k) \geq 4,$$

for all $1 \leq i \leq k-2$. We prove this claim by induction. The base case is for $F_{(k-2)}$. Note that $P(0, \ldots, 0, y_{k-1}, y_k)$ is identically 0. Thus

$$\sum_{y_{k-1}, y_k \in \{0,1\}} F_{(k-2)}(y_{k-1}, y_k) = 4.$$

By induction, we may assume that

$$\sum_{y_{i+1}, y_{i+2}, \ldots, y_k \in \{0,1\}} F_{(i)}(y_{i+1}, y_{i+2}, \ldots, y_k) \geq 4$$

and prove that

$$\sum_{y_i, y_{i+1}, \ldots, y_k \in \{0,1\}} F_{(i-1)}(y_i, y_{i+1}, \ldots, y_k) \geq 4$$

for $i \leq k-2$. Note that

$$F_{(i-1)}(1, y_{i+1}, y_{i+2}, \ldots, y_k) = (-1)^{(L_i + \epsilon_i) + y_{i+1}(L_{i+1} + \epsilon_{i+1}) + \cdots + y_{k-2}(L_{k-2} + \epsilon_{k-2})},$$
$$F_{(i-1)}(0, y_{i+1}, y_{i+2}, \ldots, y_k) = (-1)^{y_{i+1}(L_{i+1} + \epsilon_{i+1}) + \cdots + y_{k-2}(L_{k-2} + \epsilon_{k-2})}. \tag{10.7}$$

https://doi.org/10.1017/9781107477063.011 Published online by Cambridge University Press

By inductive hypothesis,

$$\sum_{y_{i+1},\ldots,y_k\in\{0,1\}} F_{(i-1)}(0,y_{i+1},\ldots,y_k) = \sum_{y_{i+1},\ldots,y_k\in\{0,1\}} F_{(i)}(y_{i+1},\ldots,y_k) \geq 4.$$

If L_i is identically 0, then we set $\epsilon_i = 0$. It follows that

$$\sum_{y_{i+1},\ldots,y_k\in\{0,1\}} F_{(i-1)}(1,y_{i+1},\ldots,y_k) = \sum_{y_{i+1},\ldots,y_k\in\{0,1\}} F_{(i-1)}(0,y_{i+1},\ldots,y_k).$$

Thus we have

$$\sum_{y_i,y_{i+1},\ldots,y_k\in\{0,1\}} F_{(i-1)}(y_i,y_{i+1},\ldots,y_k) = 2 \sum_{y_{i+1},\ldots,y_k\in\{0,1\}} F_{(i-1)}(0,y_{i+1},\ldots,y_k) \geq 8$$

and we are done.

Otherwise, $L_i = 0$ defines a subspace V of $\mathbb{Z}_2^{k-i}$ that has dimension $k-i-1$. Let $V' = \mathbb{Z}_2^{k-i} \setminus V$; then V' is an affine space defined by $L_i = 1$ which has dimension $k-i-1$.

Let

$$a = \sum_{y_{i+1}y_{i+2}\cdots y_k\in V} F_{(i-1)}(0,y_{i+1},y_{i+2},\ldots,y_k),$$

$$b = \sum_{y_{i+1}y_{i+2}\cdots y_k\in V'} F_{(i-1)}(0,y_{i+1},y_{i+2},\ldots,y_k),$$

$$a' = \sum_{y_{i+1}y_{i+2}\cdots y_k\in V} F_{(i-1)}(1,y_{i+1},y_{i+2},\ldots,y_k),$$

and

$$b' = \sum_{y_{i+1}y_{i+2}\cdots y_k\in V'} F_{(i-1)}(1,y_{i+1},y_{i+2},\ldots,y_k).$$

Then

$$\sum_{y_i,y_{i+1},\ldots,y_k\in\{0,1\}} F_{(i-1)}(y_i,y_{i+1},\ldots,y_k) = a+b+a'+b'.$$

By induction, we have $a+b \geq 4$. Thus $a \geq 2$ or $b \geq 2$. If $a \geq 2$, we choose $\epsilon_i = 0$, then $a = a'$, $b = -b'$ by (10.7). Thus $a+a'+b+b' = 2a \geq 4$. If $b \geq 2$, we choose $\epsilon_i = 1$, then $a = -a'$, $b = b'$ by (10.7). Thus $a+a'+b+b' = 2b \geq 4$. This finishes the proof of the claim.

The claim shows that

$$f_{00}^* = \sum_{y_2,y_3,\ldots,y_k\in\{0,1\}} \underline{f}^{y_1=0}(y_2,y_3,\ldots,y_k) = \sum_{y_2,y_3,\ldots,y_k\in\{0,1\}} F_{(1)}(y_2,y_3,\ldots,y_k) \geq 4.$$

https://doi.org/10.1017/9781107477063.011 Published online by Cambridge University Press

Let g be the n-ary signature with the same support as f (thus satisfies the even parity condition) and on its support

$$g(x_1, x_2, \ldots, x_n) = (-1)^{P(y_1, y_2, \ldots, y_k)},$$

then $\underline{f}_\beta = \underline{g}_\beta$ for any $\beta \in \{0, 1\}^k$ other than $\beta = 11 \cdots 1$. For $\beta = 11 \cdots 1$,

$$\underline{f}_\beta = (-1)^{P(1,1,\ldots,1)+1\cdot 1\cdots 1} = -(-1)^{P(1,1,\ldots,1)} = -\underline{g}_\beta.$$

This implies that

$$\underline{f} = \underline{g} \pm 2[0, 1]^{\otimes k}. \tag{10.8}$$

(Note that we do not really construct g in Pl-Holant$(\widehat{\mathcal{EQ}}, \widehat{\mathcal{F}})$. We just use g to argue that $f^* \notin \mathscr{A}$.) Since both $[1, 0, 1, \ldots, 0 \text{ (or } 1)]$ and g are affine signatures, the following construction would define an affine signature: Connect all variables of g other than y_1 to $n - 1$ variables of $[1, 0, 1, \ldots, 0 \text{ (or } 1)]$. This construction gives a binary signature $g^* = [g^*_{00}, 0, g^*_{11}]$. (By the even parity condition, the weight 1 entry must be 0.) Note that

$$g^*_{00} = \sum_{y_2, y_3, \ldots, y_k \in \{0,1\}} \underline{g}^{y_1=0}(y_2, y_3, \ldots, y_k),$$

$$g^*_{11} = \sum_{y_2, y_3, \ldots, y_k \in \{0,1\}} \underline{g}^{y_1=1}(y_2, y_3, \ldots, y_k).$$

Thus by (10.6) and (10.8), we have

$$f^*_{00} = g^*_{00}, \quad f^*_{11} = g^*_{11} \pm 2. \tag{10.9}$$

Since g^* is a binary affine signature, we must have

$$\text{either} \quad g^*_{00} = 0 \quad \text{or} \quad g^*_{11} = 0 \quad \text{or} \quad (g^*_{00})^4 = (g^*_{11})^4.$$

Since we have $g^*_{00} = f^*_{00} \geq 4$ and both g^*_{00} and g^*_{11} are real numbers, we must have

$$\text{either} \quad g^*_{11} = 0 \quad \text{or} \quad g^*_{11} = \pm g^*_{00}.$$

If $g^*_{11} = 0$ then by (10.9) $f^*_{11} = \pm 2$ has a different nonzero norm than f^*_{00}. If $g^*_{11} = \pm g^*_{00}$, then by (10.9) $g^*_{11} = \pm f^*_{00}$ has norm ≥ 4, and thus $f^*_{11} = g^*_{11} \pm 2$ has norm $|g^*_{11}| \pm 2 = |f^*_{00}| \pm 2$. And so in this case f^*_{11} also has a different nonzero norm than f^*_{00}. In each case, $|f^*_{00}| \neq |f^*_{11}|$ and $f^*_{00} f^*_{11} \neq 0$. This implies that $f^* \notin \mathscr{A}$ and we are done. □

https://doi.org/10.1017/9781107477063.011 Published online by Cambridge University Press

10.2 Holant* Dichotomy

In this section we prove a complexity dichotomy theorem for Holant*$(\mathcal{F})$, where $\mathcal{F}$ is a set of complex-valued, *not necessarily symmetric*, constraint functions on Boolean variables. In Theorem 10.14 we identify four classes of problems that are polynomial time computable; all other problems are proved to be #P-hard. The main proof technique and indeed the formulation of the theorem use holographic transformations.

Theorem 2.12 from Chapter 2 proved a complexity dichotomy theorem for Holant*$(\mathcal{F})$, where $\mathcal{F}$ is a set of complex-valued *symmetric* constraint functions on Boolean variables. Theorem 3.5 from Chapter 3 is an equivalent statement of Theorem 2.12 stated in terms of holographic transformations (except for case 3, which has to do with vanishing signatures $\mathscr{V}$ introduced in Chapter 7). Theorem 10.14 is a generalization of Theorem 2.12 to asymmetric signatures; Theorem 2.12 can be viewed as a special case of Theorem 10.14 restricted to symmetric signatures. Furthermore, Theorem 10.14 also helps us gain a deeper and clearer understanding of the tractable cases for the symmetric ones.

10.2.1 Dichotomy Statement

For any (constraint) function F, or synonymously a signature, of arity n, we also use the same symbol F to denote the column vector indexed by $\{0,1\}^n$ as an expression of F, listing all its values in lexicographic order. Sometimes we also partition the variable set into two parts $\{x_1, x_2, \ldots, x_n\} = I \cup J$, and write F as a matrix with rows indexed by $\{0,1\}^{|I|}$ and columns indexed by $\{0,1\}^{|J|}$. For example, for a ternary function F, we can write it as a 4×2 matrix $F = F_{x_1x_2,x_3}$. As another example, for a binary function $F(x, y)$ we can write it both as a vector in $\mathbb{C}^4$, or a 2×2 matrix $F = F_{x,y}$, with row index x and column index y ranging over $\{0,1\}$. As constraint functions, F and cF, for any nonzero $c \in \mathbb{C}$, are equivalent in terms of the complexity of Holant problems they define, and we denote it by $F \cong cF$. $F \cong 0$ means that F is (identically) 0.

We use $F^{x_i=x_j}$ to denote the function of arity $n-2$ realized by a function F of arity $n \geq 2$ after taking a self-loop by connecting two edges corresponding to variables x_i and x_j. Another operation is called juxtaposition, or tensor product. Suppose F is a function of arity n and $\mathcal{I} = \{I_1, \ldots, I_k\}$ is a partition of $[n]$. If $F(X) = \prod_{j=1}^{k} F_j(X|_{I_j})$ for some functions $F_1, \ldots, F_k$, where $X = \{x_1, \ldots, x_n\}$ and $X|_{I_j} = \{x_s | s \in I_j\}$ (we also denote it by X_j), then we say F can be decomposed into type $\mathcal{I}$, or simply F has type $\mathcal{I}$. We say $\mathcal{I}$ is a *type specification*. We denote such an F by $F = \bigotimes_{\mathcal{I}}(F_1, \ldots, F_k)$. If each F_j is the function of some gadget, then $\bigotimes_{\mathcal{I}}(F_1, \ldots, F_k)$ is the function of the gadget which is the disjoint

https://doi.org/10.1017/9781107477063.011 Published online by Cambridge University Press

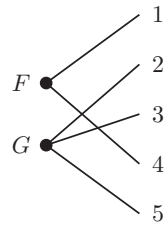

Figure 10.2. An example of juxtaposition $\bigotimes_{\mathcal{I}}(F, G)$, where $\mathcal{I} = \{\{1, 4\}, \{2, 3, 5\}\}$.

union of these gadgets for F_j, with variables arranged according to $\mathcal{I}$. When the indexing is clear, we also use notation $F_1 \otimes \cdots \otimes F_k$. Note that this tensor product notation $\otimes$ is consistent with tensor product of matrices and vectors (see Figure 10.2.) This definition of $F = \bigotimes_{\mathcal{I}}(F_1, \ldots, F_k)$ can be easily generalized to the case where $\mathcal{I} = \{I_1, \ldots, I_k\}$ is a partition of an arbitrary finite set of indices.

We use $F^{x_{j_1}=U_1,\ldots,x_{j_k}=U_k}$ to denote the function realized by connecting the unary function U_s to the input variable x_{j_s} of F, (for $s = 1, \ldots, k$). $F^{x_j=0}$ and $F^{x_j=1}$ denote respectively $F^{x_j=[1,0]}$ and $F^{x_j=[0,1]}$, and when x_j is clear we also use FU to denote $F^{x_j=U}$. The notation FU is also consistent with matrix-vector product. Matrix multiplication will be used often. For example, suppose $A = A_{x_1,x_2}$, $B = B_{x_3,x_4}$ and $C = C_{x_5,x_6}$ are three binary functions. Then the function $(A \otimes B \otimes C)^{x_2=x_3,x_4=x_5}$ is formed by linking the three functions A, B, C in a path, and its matrix form is just the matrix product ABC indexed by x_1 (for row) and x_6 (for column). Note that $A_{\emptyset,x_1}B_{x_2,\emptyset}$ or $A^{\mathrm{T}}B$ is the dot product of unary functions A and B. Similarly, $A_{x_1,\emptyset}B_{\emptyset,x_2}$ or AB^{T} is the matrix form of the tensor product function $\bigotimes_{\{\{1\},\{2\}\}}(A, B)$ (or just $A \otimes B$) of unary functions A and B.

We say a function set $\mathcal{F}$ is closed under tensor product if for any functions A and B on two disjoint sets of variables indexed by I and J respectively, $A, B \in \mathcal{F}$ implies that $\bigotimes_{\mathcal{I}}(A, B) \in \mathcal{F}$, where $\mathcal{I} = \{I, J\}$. Given $\mathcal{F}$, its tensor closure $\langle\mathcal{F}\rangle$ is the minimum set containing $\mathcal{F}$ and is closed under tensor product. This closure exists, being the set of all functions obtained by taking a finite sequence of tensor products from $\mathcal{F}$. Because tensor products are defined for functions on disjoint sets of variables, for a holographic transformation defined by any matrix M, $\langle M\mathcal{F}\rangle = M\langle\mathcal{F}\rangle$.

The following three classes of functions are important for our dichotomy.

Definition 10.11. A function $F : \{0, 1\}^n \to \mathbb{C}$ is in the class $\mathcal{E}$ iff F is zero except (possibly) on two inputs $(a_1, \ldots, a_n)$ and $(\bar{a}_1, \ldots, \bar{a}_n) = (1 - a_1, \ldots, 1 - a_n)$.

https://doi.org/10.1017/9781107477063.011 Published online by Cambridge University Press

In other words, $F \in \mathcal{E}$ iff its support is contained in a pair of complementary points. We think of $\mathcal{E}$ as a generalization to the asymmetric setting of GENERALIZED EQUALITY $[a, 0, \ldots, 0, b]$. The tensor closure of $\mathcal{E}$ is a class we are already familiar with.

Exercise: Prove that the tensor closure $\langle \mathcal{E} \rangle$ is the class of signatures of product type $\mathscr{P}$.

Definition 10.12. A function $F : \{0, 1\}^n \to \mathbb{C}$ is in the class $\mathcal{M}$ iff F is 0 except (possibly) on $n + 1$ inputs whose Hamming weight is ≤ 1.

The name $\mathcal{M}$ is given for *matching*. But this is not to be confused with the matchgate signatures $\mathscr{M}$, which is defined by gadgets called matchgates using weighted *perfect* matching signatures.

Definition 10.13. $\mathcal{T}$ is the set of all functions of arity ≤ 2.

Note that $\mathcal{U}$ is a subset of all three classes, $\mathcal{E}$, $\mathcal{M}$, and $\mathcal{T}$. $\langle \mathcal{U} \rangle$ is the class of degenerate signatures. A binary function belongs to $\langle \mathcal{U} \rangle$ iff its matrix has determinant 0. A ternary function $F(x_1, x_2, x_3)$ belongs to $\langle \mathcal{T} \rangle$ iff $F^{x_j=U} \cong 0$ for some $1 \leq j \leq 3$ and some unary $U \ncong 0$. If furthermore the ternary function $F(x_1, x_2, x_3)$ is symmetric, then the following statements are all equivalent: (1) $F \in \langle \mathcal{T} \rangle$; (2) $F \in \langle \mathcal{U} \rangle$; (3) $F = [a, b]^{\otimes 3}$ for some unary $[a, b]$; and (4) $FU \cong 0$ for some unary $U \ncong 0$ (take $U = [b, -a]$ if $[a, b] \ncong 0$, or *any* unary $U \ncong 0$ if $[a, b] \cong 0$).

Denote

$$Z_1 = \begin{bmatrix} 1 & 1 \\ \mathrm{i} & -\mathrm{i} \end{bmatrix} \quad \text{and} \quad Z_2 = \begin{bmatrix} 1 & 1 \\ -\mathrm{i} & \mathrm{i} \end{bmatrix}.$$

Note that *symmetric* signatures f in $Z_1\mathcal{M}$ are the same as those in $\mathscr{R}_2^+$, and equivalently $\mathrm{rd}^+(f) \leq 1$. Similarly Z_2 corresponds to $\mathscr{R}_2^-$, and rd^-.

Theorem 10.14. *Let $\mathcal{F}$ be any set of complex-valued functions in Boolean variables. The problem Holant$^*(\mathcal{F})$ is polynomial time computable, if*

1. *$\mathcal{F} \subseteq \langle \mathcal{T} \rangle$, or*
2. *there exists an orthogonal matrix H such that $\mathcal{F} \subseteq \langle H\mathcal{E} \rangle$, or*
3. *for $Z = Z_1$ or Z_2, $\mathcal{F} \subseteq \langle Z\mathcal{E} \rangle$, or*
4. *for $Z = Z_1$ or Z_2, $\mathcal{F} \subseteq \langle Z\mathcal{M} \rangle$.*

In all other cases, Holant$^*(\mathcal{F})$ is #P-hard. The dichotomy is still valid even if the inputs are restricted to planar graphs.

https://doi.org/10.1017/9781107477063.011 Published online by Cambridge University Press

Note that since $\langle M\mathcal{F}\rangle = M\langle\mathcal{F}\rangle$, and $\langle\mathcal{E}\rangle = \mathscr{P}$, the tractable classes can also be stated as $\mathcal{F} \subseteq \langle\mathcal{T}\rangle$, $\mathcal{F} \subseteq H\mathscr{P}$, $\mathcal{F} \subseteq Z\mathscr{P}$, or $\mathcal{F} \subseteq Z\langle\mathcal{M}\rangle$.

Compared to Theorems 2.12 and 3.5, the case $\mathcal{F} \subseteq \langle\mathcal{T}\rangle$ clearly corresponds to sets of functions of arity ≤ 2 (case 1 of Theorem 3.5, also Class $\mathfrak{A}$ of Theorem 2.12). The cases $\mathcal{F} \subseteq \langle H\mathcal{E}\rangle$ and $\mathcal{F} \subseteq \langle Z\mathcal{E}\rangle$ correspond to case 2 of Theorem 3.5, also to all of Class $\mathfrak{B}$ (except the case I(1, $\pm$2i)) as well as Class $\mathfrak{C}$ in Theorem 2.12. This is the same as $\mathscr{P}$-transformable for asymmetric signatures. The case $\mathcal{F} \subseteq \langle Z\mathcal{M}\rangle$ corresponds to case 3 of Theorem 3.5, also to the case I(1, $\pm$2i) of Class $\mathfrak{B}$ in Theorem 2.12.

Exercise: Prove that, for symmetric signatures, Theorem 10.14 is equivalent to Theorems 3.5 and 2.12. Moreover, the tractable classes correspond to each other as described above.

Exercise: Explain how the tractable cases in Theorem 7.19 related to the vanishing signatures $\mathscr{V}$ present themselves in Theorem 10.14. Explain why the tractable classes $\mathscr{A}$-transformable signatures and $\mathscr{M}$-transformable signatures do not appear in Theorem 10.14.

10.2.2 Tractability

We first prove the four classes in Theorem 10.14 are polynomial time computable.

Theorem 10.15. *Holant*$^*(\langle\mathcal{T}\rangle)$, Holant$^*(\langle H\mathcal{E}\rangle)$, Holant$^*(\langle Z\mathcal{E}\rangle)$ and Holant$^*(\langle Z\mathcal{M}\rangle)$ are polynomial time computable.*

Proof. By "decoupling" a vertex v into several vertices according to its tensor product factors of the function at v, Holant$^*(\langle\mathcal{F}\rangle)$ is trivially equivalent to Holant$^*(\mathcal{F})$, for any $\mathcal{F}$.

Firstly, to show the tractability of Holant$^*(\mathcal{T})$, we consider any input graph G. Each connected component of G is either a path or a cycle, since the maximum degree of G is 2. So we only need to compute some m steps of matrix multiplications and trace operations, where m is the number of edges in G. This is clearly polynomial time computable.

Next we prove the tractability of Holant$^*(H\mathcal{E})$ for any orthogonal matrix H, and of Holant$^*(Z\mathcal{E})$ for both $Z = Z_1$ and Z_2. Let $M = H$ or Z. We first reformulate it as a bipartite Holant problem Holant$(=_2 |M\mathcal{E})$ (since $\mathcal{U} = M\mathcal{U} \subset M\mathcal{E}$, we can drop the $*$ notation in Holant). Here the edges are replaced by the binary Equality function $(=_2) = [1, 0, 1]$. Now we perform a holographic reduction by M, where the right-hand side (RHS) is transformed to $\mathcal{E}$ and the left-hand side (LHS) is transformed to $[1, 0, 1]H^{\otimes 2} = [1, 0, 1] \in \mathcal{E}$ or

https://doi.org/10.1017/9781107477063.011 Published online by Cambridge University Press

$[1, 0, 1]Z^{\otimes 2} \cong [0, 1, 0] \in \mathcal{E}$. Hence after a holographic reduction, our problem becomes a case of Holant($\mathcal{E}$). This is clearly polynomial time computable: Any 0-1 assignment to a particular edge can be uniquely propagated to all edges in each connected component, unless we find a contradiction (in effect we discover a cycle, implied by the connections of the signatures from $\mathcal{E}$, that requires an odd number of switches of 0, 1 values. In that case we return 0.) So, Holant$^*(H\mathcal{E})$ is polynomial time computable.

Finally we prove the tractability of Holant$^*(Z\mathcal{M})$. After a holographic reduction by Z, it becomes Holant$(\{\neq_2\}|\mathcal{M})$. (Again we can drop the $*$ notation in Holant, because $\mathcal{U} = Z\mathcal{U} \subset Z\mathcal{M}$.) If two unary functions from $\mathcal{M}$ are connected via $(\neq)$, it produces a constant factor. If a unary function U in $\mathcal{M}$ is connected via $(\neq)$ to any $F \in \mathcal{M}$ of arity $n > 1$, it produces FU' of arity $n - 1$ for another unary U'. Clearly $FU' \in \mathcal{M}$, since the function value of FU' on any input with Hamming weight ≥ 2 is certainly 0. Recursively apply these replacement steps until there are no more unary functions left. Hence, we only need to show that Holant$(\{\neq_2\}|\mathcal{M} - \mathcal{U})$ is tractable. The input graph is a bipartite graph. Because all functions on the LHS vertex set are $\neq_2$, in order to have a nonzero evaluation, any assignment must have exactly half of all edges assigned 0 and the other half assigned 1. All functions on the RHS are from $\mathcal{M} - \mathcal{U}$, thus have arity ≥ 2. If there is a vertex of degree ≥ 3 in the RHS, then the RHS requires that strictly less than half of all edges are 1, so the value of this problem is 0. Thus we only need to calculate on graphs where all vertices have degree 2 (each connected component is a cycle), which is tractable by matrix multiplication and taking trace. □

10.2.3 Outline of the Hardness Proof

We outline the hardness proof of Theorem 10.14. The main idea is to reduce the general Holant* problem with asymmetric signatures to the symmetric case, and then apply Theorem 2.12. However, it is not easy to do that when functions have large arities. In Subsection 10.2.5, we first establish an arity reduction theorem. We show that for any of the four tractable families $\mathcal{F}'$, starting from any $F \in \mathcal{F} - \mathcal{F}'$, unless F already has some minimum arity, we can construct a function Q such that

1. Holant$^*(\mathcal{F} \cup \{Q\}) \equiv_T$ Holant$^*(\mathcal{F})$,
2. $Q \notin \mathcal{F}'$, and
3. Q has a reduced arity.

So assuming that the given set of functions is not a subset of any of the four tractable families, we can continue this arity reduction until reaching some

https://doi.org/10.1017/9781107477063.011 Published online by Cambridge University Press

minimum arity. More specifically this will end in a finite number of steps producing the following: In the case of $\langle \mathcal{T} \rangle$, we will end up with an arity 3 signature that is not in $\langle \mathcal{T} \rangle$. For any family of the type $\langle H\mathcal{E} \rangle$, $\langle Z\mathcal{E} \rangle$, $\langle Z\mathcal{M} \rangle$, we can get a signature of arity 2 that is not in the family.

Having these functions with small arities (2 or 3) in hand, we can construct some simple gadgets to get *symmetric* functions, which we do in Section 10.2.6. The hope is that these symmetric functions are also out of various tractable families. However, we come cross some difficulties by doing this. For example, using a single function of arity 3 alone that is not in $\langle \mathcal{T} \rangle$, we do not know how to construct a symmetric arity 3 function that is not in $\langle \mathcal{T} \rangle$ in all cases. In our proof, we get help from other signatures. Namely, we not only use a signature of arity 3 that is not in $\langle \mathcal{T} \rangle$, but also some binary signatures that are not in $\langle H\mathcal{E} \rangle$, $\langle Z\mathcal{E} \rangle$ or $\langle Z\mathcal{M} \rangle$, respectively, to construct a *symmetric* signature Q of arity 3 that is not in $\langle \mathcal{T} \rangle$. This is proved in Theorem 10.24. Similarly, in Theorem 10.25, we prove that we can also construct binary *symmetric* signatures Q' that are not in $\langle H\mathcal{E} \rangle$, $\langle Z\mathcal{E} \rangle$ or $\langle Z\mathcal{M} \rangle$, respectively, starting from a given binary signature not in the respective tractable class *together* with the ternary signature $Q \notin \langle \mathcal{T} \rangle$, *provided* Q happens to belong to the respective tractable class $\langle H\mathcal{E} \rangle$, $\langle Z\mathcal{E} \rangle$ or $\langle Z\mathcal{M} \rangle$.

Then we apply the dichotomy for symmetric signatures, Theorem 2.12, as follows. Either Holant*(Q) is #P-hard, or Q belongs to one of the tractable families, which are the restrictions of the listed four cases in Theorem 10.14 to symmetric signatures. If Holant*(Q) is already #P-hard then we are done. Otherwise, since $Q \notin \langle \mathcal{T} \rangle$, it must belong to $\langle H\mathcal{E} \rangle$, $\langle Z\mathcal{E} \rangle$ or $\langle Z\mathcal{M} \rangle$. So the additional condition on Q is fulfilled and by Theorem 10.25, we can construct binary symmetric signatures that are not in the same tractable family, and therefore we also get #P-hardness by Theorem 2.12.

We note that, all our starting problems for #P-hardness are already hard for planar graphs and all the gadgets we use in the reduction are planar. As a result, our final dichotomy also holds for planar graphs. In the proofs later, we will not explicitly state this.

One technical lemma is used extensively in both Sections 10.2.5 and 10.2.6, which substantially simplifies the proof. We call it the separation lemma, which is stated and proved in Section 10.2.4.

10.2.4 Separation Lemma

In this subsection, we introduce a simple lemma that is used frequently in the proof, and its main purpose is proof simplification. This lemma is applied in the following situation. We have identified a finite set of requirements, where

https://doi.org/10.1017/9781107477063.011 Published online by Cambridge University Press

the violation of each requirement can be expressed as a system of polynomial equations in the entries of some signatures. These entries can be ultimately expressed as polynomials in the free variables constituting the entries of arbitrary unary signatures we use. Then to show all these requirements can be simultaneously satisfied, we only need to prove each requirement can be individually satisfied, without regard to the consistency of the satisfying variable assignments.

We will state several versions of the separation lemma. The following lemma is well known. For completeness we give a proof.

Lemma 10.16. *Let $\mathbb{F}$ be an infinite field. Suppose $\{P_1, P_2, \ldots, P_m\}$ is a finite set of nonzero polynomials in $\mathbb{F}[x_1, x_2, \ldots, x_n]$. Then there exist values $a_1, a_2, \ldots, a_n \in \mathbb{F}$ such that $P_i(a_1, a_2, \ldots, a_n) \neq 0$ for all $1 \leq i \leq m$.*

Proof. For $n = 1$, the conclusion holds obviously.

Suppose the conclusion holds for $n - 1$. Let $P_i = \sum_{j=0}^{m_i} p_{i,j}(x_1, \ldots, x_{n-1}) x_n^j$. Because P_i is not the zero polynomial, we may assume p_{i,m_i} is a nonzero polynomial in $\mathbb{F}[x_1, x_2, \ldots, x_{n-1}]$. By induction, there exist values $a_1, a_2, \ldots, a_{n-1} \in \mathbb{F}$ such that $p_{i,m_i}(a_1, a_2, \ldots, a_{n-1}) \neq 0$, and $P_i(a_1, a_2, \ldots, a_{n-1}, x_n) \in \mathbb{F}[x_n]$ is a nonzero polynomial in x_n, for all $1 \leq i \leq m$. It follows that there exists $a_n \in \mathbb{F}$ such that $P_i(a_1, a_2, \ldots, a_n) \neq 0$ for all $1 \leq i \leq m$. □

We will give various gadget constructions which use some unary functions $U_k = [x_k, y_k]$, $k = 1, 2, \ldots, m$. Technically the gadget is defined only when specific values for x_k, y_k have been chosen. A signature is expressed as an ordered set of values; this is true for the given constraint functions as well as the signature of the constructed gadget. The entries of the signature of the constructed gadget can be expressed as polynomials in x_k, y_k (the coefficients depend on the given constraint functions). Frequently we have a finite set of conditions, where the negation of each condition is expressible as polynomial equations on x_k, y_k. A construction succeeds if we satisfy all these conditions. The following lemma lets us deal with these conditions separately.

Lemma 10.17. *Let F be the signature of a gadget construction using unary functions $U_k = [x_k, y_k]$, $k = 1, 2, \ldots, m$.*

Suppose $S_1, S_2, \ldots, S_N$ are sets of functions, where a function $K \in S_i$ iff the signature entries of K satisfy a finite system of polynomial equations $\{P_{i,1}, P_{i,1}, \ldots, P_{i,m_i}\}$.

If for every $1 \leq i \leq N$, there exists an assignment σ of x_k and y_k ($k = 1, 2, \ldots, m$), such that $F \notin S_i$ under σ; then there exists an assignment σ of x_k and y_k ($k = 1, 2, \ldots, m$), such that for all $1 \leq i \leq N$, $F \notin S_i$ under σ.

https://doi.org/10.1017/9781107477063.011 Published online by Cambridge University Press

Equivalently, suppose for every assignment σ of x_k and y_k ($k = 1, 2, \ldots, m$), there exists $1 \le i \le N$, such that $F \in S_i$ under σ, then there exists $1 \le i \le N$, such that for all assignments σ of x_k and y_k ($k = 1, 2, \ldots, m$), $F \in S_i$ under σ.

Proof. Every signature entry of F is expressible as a polynomial in x_k and y_k, $k = 1, 2, \ldots, m$. Substituting these expressions in $\{P_{i,1}, P_{i,1}, \ldots, P_{i,m_i}\}$ we can express the condition $F \in S_i$ by a finite set of polynomial equations $\{P'_{i,1}, P'_{i,1}, \ldots, P'_{i,m_i}\}$ on x_k and y_k. If for every $1 \le i \le N$, there exists an assignment σ of x_k and y_k ($k = 1, 2, \ldots, m$), such that $F \notin S_i$ under σ; then for every $1 \le i \le N$, there exists some $1 \le j_i \le m_i$ such that P'_{i,j_i} is not identically 0 as a polynomial on x_k and y_k ($k = 1, 2, \ldots, m$). Therefore there is an assignment σ on x_k and y_k ($k = 1, 2, \ldots, m$), such that for all $1 \le i \le N$, $P'_{i,j_i}|_\sigma \neq 0$, hence $F \notin S_i$ under σ. □

The following lemma is another version of the same statement.

Lemma 10.18. *Suppose a gadget construction using unary functions $U_k = [x_k, y_k]$ ($k = 1, 2, \ldots, m$) succeeds if it satisfies a finite set of properties R_i, $i = 1, 2, \ldots, N$. Suppose violation of each property R_i is specified by a finite set of polynomial equations. If for each i we can find unary functions $U_k = [x_k, y_k]$ to satisfy property R_i, then we can find unary functions $U_k = [x_k, y_k]$ so that the construction succeeds.*

10.2.5 Arity Reduction

In the next two subsections we prove the hardness part of Theorem 10.14, that is, if $\mathcal{F} \not\subseteq \langle \mathcal{T} \rangle$, $\mathcal{F} \not\subseteq \langle H\mathcal{E} \rangle$, $\mathcal{F} \not\subseteq \langle Z\mathcal{E} \rangle$, $\mathcal{F} \not\subseteq \langle Z\mathcal{M} \rangle$, then Holant*($\mathcal{F}$) is #P-hard.

In this subsection, we show how to reduce the arity of a nontractable signature to a minimum.

Theorem 10.19. *Let $\mathcal{F}'$ be any one of $\langle \mathcal{T} \rangle$, or $\langle H\mathcal{E} \rangle$ where H is orthogonal, or $\langle Z\mathcal{E} \rangle$, or $\langle Z\mathcal{M} \rangle$, where $Z = Z_1$ or Z_2. Let $r = 3$ if $\mathcal{F}' = \langle \mathcal{T} \rangle$, and $r = 2$ in the other three cases.*

Suppose there is a function $F \in \mathcal{F} - \mathcal{F}'$. If arity$(F) > r$, *then we can realize a function Q by connecting F with some unary functions, such that:*

(1) Holant$(\mathcal{F} \cup \{Q\}) \equiv_{\mathrm{T}}$ Holant*$(\mathcal{F})$;*
(2) $Q \notin \mathcal{F}'$ and
(3) $r \le$ arity$(Q) <$ arity(F).

In order to prove Theorem 10.19, we first prove two lemmas. First, we show that any *type specification* in a tensor product decomposition can be described by a finite system of polynomial equations. Note that the type a signature has

https://doi.org/10.1017/9781107477063.011 Published online by Cambridge University Press

is not unique; e.g., if $\mathcal{J}$ is a refinement of $\mathcal{I}$, then a signature having type $\mathcal{J}$ also has type $\mathcal{I}$.

Lemma 10.20. *For any type specification $\mathcal{I}$, there is a finite set of polynomial equations $E_{\mathcal{I}}$ in the entries of a signature F, such that F has type $\mathcal{I}$ iff F satisfies $E_{\mathcal{I}}$.*

Proof. If $\mathcal{I} = \{[n]\}$, there is no requirement on F to have type $\mathcal{I}$. We can use a trivial equation such as $0 = 0$.

Consider the case $\mathcal{I} = \{I_1, I_2\}$. Suppose F has type $\mathcal{I}$. Then obviously, for any two values a_1, b_1 of X_1 and any two values a_2, b_2 of X_2,

$$\det \begin{bmatrix} F(a_1, a_2) & F(a_1, b_2) \\ F(b_1, a_2) & F(b_1, b_2) \end{bmatrix} = 0. \tag{10.10}$$

Without loss of generality we list the indices in I_1 before those of I_2. Hence the collection of all these equations $E_{\{I_1, I_2\}}$ is a necessary condition that F has type $\mathcal{I}$. It is also a sufficient condition. Arrange the values of F into a matrix $F_{X_1, X_2} = F(X_1, X_2)$, where the row indices (respectively column indices) are all possible values of X_1 (respectively X_2). The conditions (10.10) for all a_1, b_1 and all a_2, b_2 imply that any 2×2 submatrix of F is singular, and so $\operatorname{rank}(F_{X_1, X_2}) \leq 1$. Hence, F_{X_1, X_2} is the product of a column vector and a row vector. It follows that F has type $\mathcal{I}$.

Now consider a general partition $\mathcal{I} = \{I_1, \ldots, I_k\}$, and again suppose F has type $\mathcal{I}$. It follows that for any $1 < i \leq k$, and for any fixed values $a_{i+1}, \ldots, a_k$ for $X_{i+1}, \ldots, X_k$ (this is vacuous if $i = k$), we get a function F' from F by setting variables in $X_{i+1}, \ldots, X_k$ to $a_{i+1}, \ldots, a_k$ respectively. F' has type $\{\bigcup_{j=1}^{i-1} I_j, I_i\}$. We define the following set of equations: For all $1 < i \leq k$, and for all assignment values $a_{i+1}, \ldots, a_k$ for $X_{i+1}, \ldots, X_k$, include all equations in $E_{\{\bigcup_{j=1}^{i-1} I_j, I_i\}}$ for the function F'. This is a finite set of polynomial equations. Obviously, this is a necessary condition for F to have type $\mathcal{I}$.

We prove that it is also a sufficient condition. If F is the zero function, then F has type $\mathcal{I}$ trivially. Assume F is not the zero function. Let $i = k$, by what has been proved when $k = 2$, $F = \bigotimes_{\{\bigcup_{j=1}^{k-1} I_j, I_k\}}(P_{k-1}, F_k)$ for some functions P_{k-1} and F_k. Because F is not the zero function, there exists a value a_k for X_k such that $F_k(a_k) \neq 0$. The remaining conditions, for $1 < i \leq k - 1$, yield a finite set of homogeneous equations for $F^{X_k = a_k}(X_1, \ldots, X_{k-1}) = P_{k-1}(X_1, \ldots, X_{k-1}) F_k(a_k)$. After canceling the nonzero factor $F_k(a_k)$, by induction, we obtain a sufficient condition that P_{k-1} has type $\{I_1, \ldots, I_{k-1}\}$. Hence F has type $\mathcal{I}$. □

https://doi.org/10.1017/9781107477063.011 Published online by Cambridge University Press

Next, we prove a property of this decomposition. This property is used throughout in the proof of Theorem 10.19.

Lemma 10.21. *Suppose there exists some type $\mathcal{I} = \{I_1, \ldots, I_k\}$ over $[n-1]$, such that for all unary functions $U = [x, y]$, $F^{x_n=U} = FU = xF^{x_n=0} + yF^{x_n=1}$ has the same type $\mathcal{I}$. Furthermore, suppose $F^{x_n=0} = \bigotimes_{\mathcal{I}}(F_1, F_2, \ldots, F_k)$ and $F^{x_n=1} = \bigotimes_{\mathcal{I}}(K_1, K_2, \ldots, K_k)$ are linearly independent as two vectors. Then there exists* exactly one *index $i \in [k]$ such that F_i and K_i are linearly independent.*

Proof. If F_i and K_i are linearly dependent for all $i \in [k]$, then obviously $F^{x_n=0}$ and $F^{x_n=1}$ would be linearly dependent. Now for a contradiction, suppose there are two distinct indices $i \in [k]$, such that F_i and K_i are linearly independent. Without loss of generality, let $i = 1, 2$ respectively.

Because $F^{x_n=0}, F^{x_n=1}$ are linearly independent, F_j and K_j are not the zero function for any $j \in [k]$. For any $j \in [k] - \{1, 2\}$, by Lemma 10.16, there exist $|I_j|$ many unary functions such that both F_j and K_j become nonzero constants when combined with them. This is because we can write a finite system of polynomial equations Φ_j in a_k, b_k for the $|I_j|$ many unary functions $U_k = [a_k, b_k]$, that expresses the condition that F_j becomes 0 when connecting variables x_k to U_k, for all $k \in I_j$. Since F_j is not the zero function, some 0-1 assignment violates some equation in Φ_j. Hence at least one polynomial P_1 in Φ_j is not identically 0. Similarly there is a finite system of polynomial equations Ψ_j for K_j and some polynomial P_2 is not identically 0. Then by Lemma 10.16 there is a common assignment to a_k, b_k ($k \in I_j$) such that both F_j and K_j are nonzero. After combining $F^{x_n=0}$ and $F^{x_n=1}$ with these unary functions, for all $j \in [k] - \{1, 2\}$, we obtain respectively the functions $c_0 F_1 \otimes F_2$ and $c_1 K_1 \otimes K_2$ over the variables in $I_1 \cup I_2$, where $c_0, c_1 \neq 0$.

Suppose $U = [x, y]$ and $xy \neq 0$. If we combine $FU = xF^{x_n=0} + yF^{x_n=1}$ with the same set of $\left|\bigcup_{j=3}^{k} I_j\right|$ many unary functions, the resulting function is $c_0 xF_1 \otimes F_2 + c_1 yK_1 \otimes K_2$. By the assumption, FU has type $\mathcal{I}$, then this function has type $\{I_1, I_2\}$. However, we show that, for any $xy \neq 0$, this function does not have type $\{I_1, I_2\}$. The matrix form (row index is $X|_{I_1}$, column index is $X|_{I_2}$) of this function is the $2^{|I_1|} \times 2^{|I_2|}$ matrix

$$\begin{bmatrix} F_1 & K_1 \end{bmatrix} \begin{bmatrix} c_0 x & 0 \\ 0 & c_1 y \end{bmatrix} \begin{bmatrix} F_2^{\mathrm{T}} \\ K_2^{\mathrm{T}} \end{bmatrix},$$

where F_1, K_1, F_2, K_2 are column vectors. Since F_1 and K_1 are linearly independent, and F_2 and K_2 are linearly independent, this matrix has rank 2. If this function has type $\{I_1, I_2\}$, its matrix form would have rank ≤ 1. This contradiction proves the lemma. □

https://doi.org/10.1017/9781107477063.011 Published online by Cambridge University Press

Proof of Theorem 10.19: $\mathcal{F}' = \langle \mathcal{T} \rangle$

Suppose $F \in \mathcal{F} - \langle \mathcal{T} \rangle$, with arity$(F) > 3$. Being out of $\langle \mathcal{T} \rangle$, F is not the zero function. If for some unary function $U = [x, y]$, $FU \notin \langle \mathcal{T} \rangle$, then we are done by setting $Q = FU$. Hence we assume for any unary function $U = [x, y]$, $FU = xF^{x_n=0} + yF^{x_n=1}$ has some type $\mathcal{J}$, where each set $J_j \in \mathcal{J}$ has cardinality $|J_j| \leq 2$. For the fixed arity(F), there are only finitely many such types, which are specifiable by a finite set of polynomial equations in x, y. It follows from Lemma 10.17 that, there exists some type $\mathcal{I} = \{I_1, \ldots, I_k\}$, where each $|I_j| \leq 2$, such that for all x, y, FU has the same type $\mathcal{I}$. In particular, both $F^{x_n=0}$ and $F^{x_n=1}$ have type $\mathcal{I}$.

If $F^{x_n=0}$ and $F^{x_n=1}$ are linearly dependent, then $F \in \langle \mathcal{T} \rangle$, having type $\mathcal{I} \cup \{\{n\}\}$.

So we assume $F^{x_n=0} = \bigotimes_{\mathcal{I}}(F_1, F_2, \ldots, F_k)$ and $F^{x_n=1} = \bigotimes_{\mathcal{I}}(K_1, K_2, \ldots, K_k)$ are linearly independent. Being linearly independent, none of the tensor factors of $F^{x_n=0}$ and $F^{x_n=1}$ can be the zero function. By Lemma 10.21, there is exactly one pair of linearly independent tensor factors, without loss of generality, say F_1 and K_1. Expressing K_i in terms of F_i, for $i \geq 2$, there exists a nonzero constant c, such that $F^{x_n=1} = \bigotimes_{\mathcal{I}}(cK_1, F_2, \ldots, F_k)$. If $|I_1| = 1$, that is, F_1 and K_1 are unary functions, then $F \in \langle \mathcal{T} \rangle$, of type $\{I_1 \cup \{n\}, I_2, \ldots, I_k\}$. Thus, $|I_1| = 2$.

Without loss of generality, let $I_1 = \{1, 2\}$. We can fix the variables of F in $I_2, \ldots, I_k$ to some values, such that $F_2, \ldots, F_k$ each contributes a nonzero factor. By this we get a ternary function Q in variables x_1, x_2, x_n, and $F = Q \otimes F_2 \otimes \cdots \otimes F_k$. If $Q \in \langle \mathcal{T} \rangle$, then $F \in \langle \mathcal{T} \rangle$, contrary to assumption. Hence $Q \notin \langle \mathcal{T} \rangle$.

Proof of Theorem 10.19: $\mathcal{F}' = \langle H\mathcal{E} \rangle$ **or** $\langle Z\mathcal{E} \rangle$

For any function F of arity n and invertible matrix M, $F \in \langle M\mathcal{E} \rangle$ iff $(M^{-1})^{\otimes n} F \in \langle \mathcal{E} \rangle$. Since we will realize our function Q of a lower arity than F by connecting F with some unary functions, and since unary functions are transformed to other unary functions under any invertible holographic transformation, we only need to prove for $\langle \mathcal{E} \rangle$. Suppose $F \notin \langle \mathcal{E} \rangle$, and arity$(F) = n > 2$. F is not the zero function. If for some unary function $U = [x, y]$, $FU \notin \langle \mathcal{E} \rangle$, we are done with $Q = FU$. Hence we assume for any unary function $U = [x, y]$, $FU = xF^{x_n=0} + yF^{x_n=1} \in \langle \mathcal{E} \rangle$.

For any partition $\mathcal{I} = \{I_1, \ldots, I_k\}$ of $[n]$, and any $\mathcal{A} = \{A_1, \ldots, A_k\}$, such that $A_j \in \{0, 1\}^{|I_j|}$, we define a set of functions $(\mathcal{I}, \mathcal{A})$. Each A_j is a 0-1 string of length $|I_j|$. In the definition for $(\mathcal{I}, \mathcal{A})$ below we use the set $\{A_j, \bar{A}_j\}$ where A_j and its complement $\bar{A}_j$ play symmetric roles, and so we may normalize the first bit of A_j to be 0. We say a function P belongs to the set $(\mathcal{I}, \mathcal{A})$ iff P has type

https://doi.org/10.1017/9781107477063.011 Published online by Cambridge University Press

$\mathcal{I}$, and $P = \bigotimes_{\mathcal{I}}(P_1, P_2, \ldots, P_k)$ for some functions $P_1, P_2, \ldots, P_k$ such that for any $j \in [k]$, $P_j(X|_{I_j})$ is zero if $X|_{I_j} \notin \{A_j, \bar{A}_j\}$. Thus, $P_j \in \mathcal{E}$ for each $j \in [k]$, and any function in $(\mathcal{I}, \mathcal{A})$ belongs to $\langle \mathcal{E} \rangle$.

Indeed functions in $\langle \mathcal{E} \rangle$ of arity n is the union of these finitely many function sets $(\mathcal{I}, \mathcal{A})$. Obviously, functions in $(\mathcal{I}, \mathcal{A})$ can be described by a finite system of polynomial equations. Since $FU \in \langle \mathcal{E} \rangle$ for all $U = [x, y]$, by Lemma 10.17, there must exist one $(\mathcal{I}, \mathcal{A})$, such that for any x, y, FU belongs to the same set $(\mathcal{I}, \mathcal{A})$. In particular, both $F^{x_n=0}$ and $F^{x_n=1}$ belong to $(\mathcal{I}, \mathcal{A})$.

If $F^{x_n=0}$ and $F^{x_n=1}$ are linearly dependent, then $F \in \langle \mathcal{E} \rangle$, because $F^{x_n=0}$, $F^{x_n=1} \in \langle \mathcal{E} \rangle$, and $F = F^{x_n=b} \otimes [\lambda, \mu]$ for some $b \in \{0, 1\}$, and $\lambda, \mu \in \mathbb{C}$, where $[\lambda, \mu] \in \mathcal{U} \subset \mathcal{E}$ is a unary function on x_n.

Let $F^{x_n=0} = \bigotimes_{\mathcal{I}}(F_1, F_2, \ldots, F_k)$ and $F^{x_n=1} = \bigotimes_{\mathcal{I}}(K_1, K_2, \ldots, K_k)$ be linearly independent. Being linearly independent, none of the tensor factors of $F^{x_n=0}$ and $F^{x_n=1}$ can be the zero function. By Lemma 10.21, there is exactly one pair of linearly independent tensor factors, say, F_1 and K_1. Expressing K_i in terms of F_i, for $i \geq 2$, there exists a nonzero constant c, such that $F^{x_n=1} = \bigotimes_{\mathcal{I}}(cK_1, F_2, \ldots, F_k)$.

We can fix the variables of F in $I_2, \ldots, I_k$ to some values, such that $F_2, \ldots, F_k$ each contributes a nonzero factor. We obtain a function K. After dividing out a nonzero constant, we have $K^{x_n=0} = F_1$ and $K^{x_n=1} = cK_1$, where $c \neq 0$. K evaluates to zero, except on possibly four inputs $\{A_1 0, \bar{A}_1 0, A_1 1, \bar{A}_1 1\}$. Combine the $|I_1| - 1$ variables of K other than x_n and the first variable in I_1 with the function $[1, 1]$, we get a binary function in matrix form

$$Q = \begin{bmatrix} K(A_1 0) & K(A_1 1) \\ K(\bar{A}_1 0) & K(\bar{A}_1 1) \end{bmatrix},$$

where we index the row by the first variable in I_1 and the column by x_n. Note that we have used the definition of $(\mathcal{I}, \mathcal{A})$. For example, in the sum defining $Q(0, 0)$, only one (possibly) nonzero term $K(A_1 0)$ is involved because the first and last bits are both set to 0. The first column in the 2×2 matrix form of Q is F_1 at entries $A_1, \bar{A}_1$, and similarly the second column is from cK_1. Because F_1 and K_1 are linearly independent, and these are the only possible nonzero entries of K, we have $\det Q \neq 0$. We claim that $Q \notin \langle \mathcal{E} \rangle$. For otherwise, being nondegenerate, $Q \in \mathcal{E}$, then the support of K is contained in either $\{A_1 0, \bar{A}_1 1\}$ or $\{A_1 1, \bar{A}_1 0\}$ This implies that $K \in \mathcal{E}$, and hence $F = K \otimes F_2 \otimes \cdots \otimes F_k \in \langle \mathcal{E} \rangle$.

Proof of Theorem 10.19: $\mathcal{F}' = \langle Z\mathcal{M} \rangle$

Again we only need to prove for $\langle \mathcal{M} \rangle$. Suppose $F \notin \langle \mathcal{M} \rangle$, and arity$(F) = n > 2$. Again we may assume for any unary function $U = [x, y]$, $FU = xF^{x_n=0} + yF^{x_n=1} \in \langle \mathcal{M} \rangle$; otherwise, we are done.

https://doi.org/10.1017/9781107477063.011 Published online by Cambridge University Press

For any $U = [x, y]$, $FU \in \langle \mathcal{M} \rangle$ has some type $\mathcal{I}$, and each tensor factor belongs to $\mathcal{M}$, that is, each tensor factor is 0 except on inputs of Hamming weight ≤ 1. Each type $\mathcal{I}$ can be specified by a finite system of polynomial equations $E_{\mathcal{I}}$ by Lemma 10.20. Now we put $E_{\mathcal{I}}$ together with the requirement that for each tensor factor, all entries of Hamming weight > 1 are 0. This requirement can also be stated as a finite system of polynomial equations. (For example, if we require that $(a_{i,j})$ is a tensor product $(b_i) \otimes (c_j)$, where $1 \leq i \leq n$, $1 \leq j \leq m$, and for some subsets $B \subseteq [n]$, $C \subseteq [m]$, $[\forall i \in B, \forall j \in C, b_i = c_j = 0]$, then we include the equations for the type specification from Lemma 10.20 together with $[a_{i,j} = 0, \forall i \in B, \textit{ or } \forall j \in C]$. The equations from Lemma 10.20 imply that a tensor factorization $(a_{i,j}) = (b_i) \otimes (c_j)$ exists. If for all $i \in [n]$, $b_i = 0$, or if for all $j \in [m]$, $c_j = 0$, then $a_{i,j}$ is identically 0. On the other hand, if for some $i_0 \in [n]$, and $j_0 \in [m]$, $b_{i_0} \neq 0$ and $c_{j_0} \neq 0$, Then $b_i = a_{i,j_0}/c_{j_0} = 0$ for all $i \in B$. Similarly $c_j = 0$ for all $j \in C$.)

Applying Lemma 10.17, we conclude that there is one type $\mathcal{I}$ such that for any $U = [x, y]$, FU has a decomposition having the same type specified by $\mathcal{I}$ with tensor factors from $\mathcal{M}$.

If $F^{x_n=0}$ and $F^{x_n=1}$ are linearly dependent, obviously, $F \in \langle \mathcal{M} \rangle$, as $U \in \mathcal{M}$, contrary to assumption.

Let $F^{x_n=0} = \bigotimes_{\mathcal{I}}(F_1, F_2, \ldots, F_k)$ and $F^{x_n=1} = \bigotimes_{\mathcal{I}}(K_1, K_2, \ldots, K_k)$ be linearly independent. As before, none of the tensor factors of $F^{x_n=0}$ and $F^{x_n=1}$ can be the zero function, and exactly one pair among F_i and K_i are linearly independent, say, F_1 and K_1. We can fix the variables of F in $I_2, \ldots, I_k$ to some values, such that $F_2, \ldots, F_k$ contribute a nonzero factor. We get a function in matrix form $K = \begin{bmatrix} F_1^{\mathrm{T}} \\ cK_1^{\mathrm{T}} \end{bmatrix}$, where the row index is $x_n = 0, 1$, columns are indexed by $\{0, 1\}^{|I_1|}$, and $c \neq 0$. Here the first row is $K^{x_n=0} = F_1^{\mathrm{T}}$. The second row is $K^{x_n=1} = cK_1^{\mathrm{T}}$. Columns are indexed by $A \in \{0, 1\}^{|I_1|}$. If the weight of A is greater than 1, then the column indexed by A is 0 because $F_1, K_1 \in \mathcal{M}$. Let S_0 denote the column indexed by $0 \cdots 0 \in \{0, 1\}^m$, and S_i denote the column indexed by the bit sequence $A \in \{0, 1\}^m$, where only the ith bit of A is 1.

For simplicity of notations, assume $I_1 = \{1, 2, \ldots, m\}$. There exists a 0-1 string $A \in \{0, 1\}^m$ of Hamming weight 1, such that $K_1(A) \neq 0$; otherwise, the only nonzero entry for K all have Hamming weight ≤ 1, and so $K \in \mathcal{M}$. This would imply $F \in \langle \mathcal{M} \rangle$, contrary to assumption. Hence there is a column S_i, $1 \leq i \leq m$, whose second entry is not zero. Without loss of generality we assume this S_i is S_m. Because F_1 and K_1 are linearly independent, There exists a column S_j linearly independent with the nonzero column S_m. If S_0 is such a column, then $Q = K^{x_1=0,\ldots,x_{m-1}=0}$ is a binary function on $\{x_n, x_m\}$, and has the matrix form $Q = [S_0, S_m]$. Note that the index for the column S_m is

https://doi.org/10.1017/9781107477063.011 Published online by Cambridge University Press

$A = 0 \cdots 01 \in \{0,1\}^m$. If S_0 is linearly dependent with S_m. Then for some $1 \le j \le m-1$, S_j is linearly independent with S_m. Let $Q = K^{x_1=0,\ldots,x_{j-1}=0,x_j=[x,y],x_{j-1}=0,\ldots,x_{m-1}=0} = [xS_0 + yS_j, xS_m]$, where $x \neq 0$ and $y \neq 0$. (Here in Q, the row index is by x_n and column index is by x_m.) We have obtained our Q such that Q is not degenerate and $Q(1,1) \neq 0$, i.e., $Q \notin \langle \mathcal{M} \rangle$. □

10.2.6 From Asymmetric to Symmetric

In this subsection, we show how to get a symmetric function from some asymmetric functions, keeping the property of *not* belonging to any one of the four tractable classes, $\langle \mathcal{T} \rangle$, $\langle H\mathcal{E} \rangle$, $\langle Z\mathcal{E} \rangle$ or $\langle Z\mathcal{M} \rangle$.

Recall that $F \cong cF$ for $c \neq 0$. In the following lemma, when we count the number of solutions, we count in terms of equivalence classes under $\cong$.

Lemma 10.22. *Suppose F is ternary function. Then $F^{x_3=U} \cong 0$ for some $U \ncong 0$ iff F has type $\{\{1,2\},\{3\}\}$. Suppose $F \notin \langle \mathcal{T} \rangle$. Then $F^{x_3=U} \ncong 0$ for any nonzero unary function U, and there exist exactly one or two nonzero $U = [x,y]$ such that $F^{x_3=U}$ is degenerate.*

Proof. If F has type $\{\{1,2\},\{3\}\}$, then $F = T \otimes [a,b]$. If $a = b = 0$ then F is identically 0, and $F^{x_3=U} \cong 0$ for any unary U. If $[a,b] \ncong 0$, then $F^{x_3=U} \cong 0$ for $U = [b,-a] \ncong 0$. Conversely, if $F^{x_3=U} \cong 0$ for some $U \ncong 0$, then $F^{x_3=0}$ and $F^{x_3=1}$ are linearly dependent, and hence F has type $\{\{1,2\},\{3\}\}$. It follows that if $F \notin \langle \mathcal{T} \rangle$, then $F^{x_3=U} \ncong 0$ for any nonzero unary function U.

Let $F' = F^{x_3=U}$. Then F' is degenerate iff $\det \begin{bmatrix} F'(0,0) & F'(0,1) \\ F'(1,0) & F'(1,1) \end{bmatrix} = 0$. Let $U = [x,y]$, then the entries of F' are linear homogeneous polynomials of x and y, so the determinant equation is a quadratic homogeneous equation. It has either one or two solutions $U \ncong 0$, or it is identically zero. We only need to prove the latter case contradicts $F \notin \langle \mathcal{T} \rangle$.

Suppose F' is degenerate for all U, then we have in particular $F^{x_3=0} = F_1 \otimes F_2$ and $F^{x_3=1} = K_1 \otimes K_2$. If $F^{x_3=0}$ and $F^{x_3=1}$ are linearly dependent, then $F \in \langle \mathcal{T} \rangle$, and so $F^{x_3=0}$ and $F^{x_3=1}$ are linearly independent. Then by Lemma 10.21, exactly one of the two pairs of functions $\{F_1, K_1\}$ and $\{F_2, K_2\}$ is linearly independent, say, $\{F_1, K_1\}$ is linearly dependent. Then F is the tensor product of F_1 and one binary function on the remaining two variables $\{x_2, x_3\}$, and so $F \in \langle \mathcal{T} \rangle$. □

For any ternary function $F(x_1, x_2, x_3) \notin \langle \mathcal{T} \rangle$, the conclusion of Lemma 10.22 certainly applies to all three variables. There is a simple

https://doi.org/10.1017/9781107477063.011 Published online by Cambridge University Press

relationship, among $1 \le i \le 3$, between the nonzero unary functions U_i such that $F^{x_i=U_i}$ is degenerate. Suppose $F^{x_1=U_1}$ is degenerate, where $U_1 \ncong 0$, then $F^{x_1=U_1} = L \otimes R$, where L and R are unary functions on x_2 and x_3 respectively. Since $F^{x_1=U_1} \ncong 0$, both $L \ncong 0$ and $R \ncong 0$. The matrix form of $F^{x_1=U_1}$ has rank exactly 1. It follows that the decomposition $L \otimes R$ is unique under $\cong$. If we define $[x, y]^\perp = [y, -x]$, and let $U_2 = L^\perp$, then $F^{x_1=U_1,x_2=U_2}$ is identically 0. Given U_1 such that $F^{x_1=U_1}$ is degenerate but not identically 0, the property that $F^{x_1=U_1,x_2=U_2} \cong 0$ also uniquely determines a nonzero U_2 as $U_2 \cong L^\perp$, where L is uniquely determined by U_1. Because $(F^{x_2=U_2})^{x_1=U_1} \cong 0$, we have $F^{x_2=U_2}$ is degenerate. This mapping from $U_1 \mapsto U_2 = L^\perp$ is well defined under $\cong$. It is also 1-1: Suppose $F^{x_1=U_1,x_2=U_2}$ and $F^{x_1=U_1',x_2=U_2}$ are both identically 0 for $U_1 \ncong 0$ and $U_1' \ncong 0$. Then $F^{x_2=U_2}$ is degenerate, and expressible as $A(x_1) \otimes B(x_3)$, where A and B are nonzero unary functions. It follows that both $U_1 \cong A^\perp$ and $U_1' \cong A^\perp$. Thus $U_1 \cong U_1'$. By symmetry the inverse map is also well defined.

The same statement holds for x_3.

We summarize this in the following lemma. Suppose $F \notin \langle \mathcal{T} \rangle$ is a ternary function. Let

$$\mathcal{U}_i = \{U \ncong 0 \mid F^{x_i=U} \text{ is degenerate}\}, \qquad 1 \le i \le 3.$$

Lemma 10.23. *There is a one-to-one correspondence between $\mathcal{U}_1, \mathcal{U}_2$, and $\mathcal{U}_3$, as follows. For $\{i, j, k\} = \{1, 2, 3\}$, each $U_i \in \mathcal{U}_i$ gives a unique factorization $F^{x_i=U_i} = V_j(x_j) \otimes V_k(x_k)$, where $V_j^\perp \in \mathcal{U}_j$ and $V_k^\perp \in \mathcal{U}_k$. In particular, $|\mathcal{U}_1| = |\mathcal{U}_2| = |\mathcal{U}_3| = 1$ or 2.*

Now we will prove a crucial theorem for the hardness part of Theorem 10.14.

Theorem 10.24. *Suppose in Holant$^*(\mathcal{F})$, we can realize the following functions*

1. *$F \notin \langle \mathcal{T} \rangle$ of arity 3.*
2. *For any orthogonal matrix H, some $P_H \notin \langle H\mathcal{E} \rangle$ of arity 2.*
3. *For both $Z = Z_1$ or Z_2, some $P_Z \notin \langle Z\mathcal{E} \rangle$ of arity 2.*
4. *For both $Z = Z_1$ or Z_2, some $S_Z \notin \langle Z\mathcal{M} \rangle$ of arity 2.*

Then we can realize a symmetric *ternary function $Q \notin \langle \mathcal{T} \rangle$ in Holant$^*(\mathcal{F})$.*

Proof. We use the gadget shown in Figure 10.3 to realize a symmetric ternary function Q. (In some cases we will need to modify it to define Q; this will be discussed later.) This gadget consists of nine copies of the function F, three copies of a unary function U_1, and three copies of a unary function U_2. The unary functions are to be determined later. Each shaded triangle labeled with

https://doi.org/10.1017/9781107477063.011 Published online by Cambridge University Press

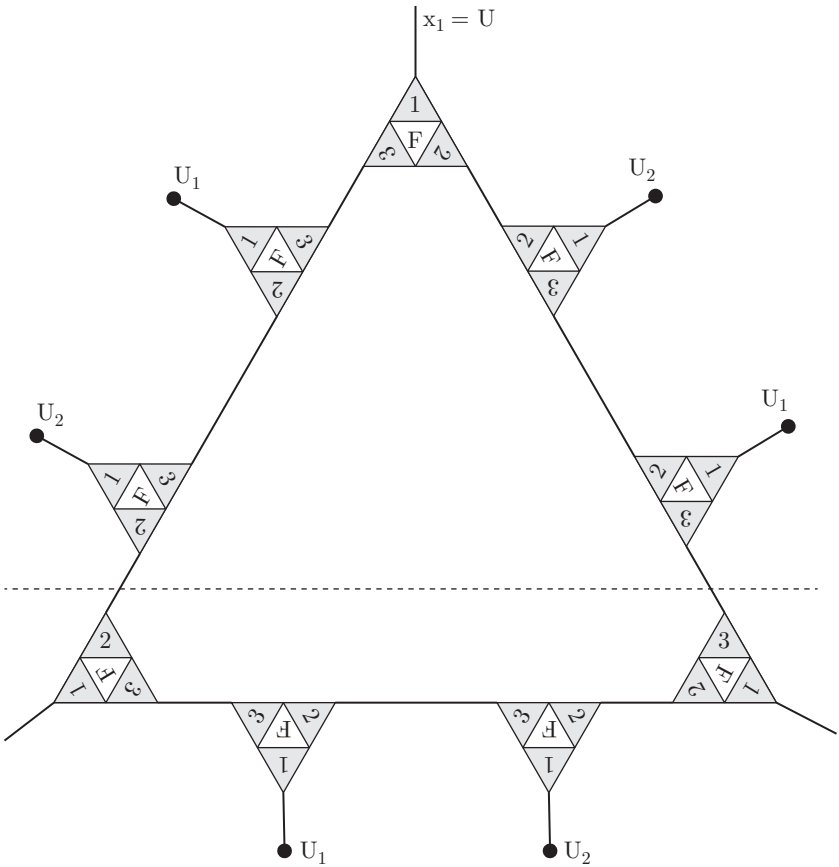

Figure 10.3. Gadget to realize a symmetric ternary function.

F in a central inner triangle represents the function $F(x_1, x_2, x_3) \notin \langle \mathcal{T} \rangle$. The labels $1, 2, 3$ inside the shaded triangle indicate which edge corresponds to variables x_1, x_2, x_3. This gadget remains unchanged if we rotate it $\frac{2\pi}{3}$. Hence, $Q(x_1, x_2, x_3) = Q(x_2, x_3, x_1) = Q(x_3, x_1, x_2)$. It follows that Q is symmetric (notice that this conclusion uses the fact that each variable x_i is a Boolean variable).

Our goal is to prove that there exist nonzero unary functions U_1 and U_2, such that $Q \notin \langle \mathcal{T} \rangle$. Since Q is symmetric, this is equivalent to: There exists no nonzero unary function U satisfying $Q^{x_1=U} \cong 0$, by Lemma 10.22.

To prove this, we divide the gadget into two parts, as shown by the dashed line in Figure 10.3. We establish two properties, one property for each part respectively. The upper part is a ternary function, denoted by S. The first

https://doi.org/10.1017/9781107477063.011 Published online by Cambridge University Press

property is that if $U \not\equiv 0$, then $S^{x_1=U} \not\equiv 0$. The matrix form of $S^{x_1=U}$ is the matrix product $F^{x_1=U_2}_{x_2,x_3} F^{x_1=U_1}_{x_2,x_3} F^{x_1=U}_{x_3,x_2} F^{x_1=U_2}_{x_2,x_3} F^{x_1=U_1}_{x_2,x_3}$, where $F^{x_1=U'}_{x_2,x_3}$ denotes the 2×2 matrix form of $F^{x_1=U'}$ with row index x_2 and column index x_3. Because $F \notin \langle \mathcal{T} \rangle$, if $U \not\equiv 0$, then $F^{x_1=U} \not\equiv 0$, by Lemma 10.22. To satisfy this property on S, we only need some U_1 and U_2 such that $F^{x_1=U_1}$ and $F^{x_1=U_2}$ are nondegenerate. By Lemma 10.22, there exist such U_1 and U_2.

The lower part is a function of arity 4, denoted by P. Two inputs of P are the original inputs x_2, x_3 of Q, corresponding to the lower left and lower right corners of the gadget respectively. The other two inputs correspond to edges connecting P with S, denoted by y_2, y_3 respectively. The second property is that the 4×4 matrix $P_{x_2x_3,y_2y_3}$ is nonsingular.

If there exist U_1 and U_2 such that both properties hold, then for any nonzero unary function U, the vector form of $Q^{x_1=U}$ is the matrix-vector product $P_{x_2x_3,y_2y_3} S^{x_1=U}$, where $S^{x_1=U}$ takes its vector form as a vector of dimension 4. Hence $Q^{x_1=U}$ is not the zero function, because $S^{x_1=U}$ is a nonzero column vector (the first property) and $P_{x_2x_3,y_2y_3}$ is a nonsingular matrix (the second property). This proves $Q \notin \langle \mathcal{T} \rangle$.

To establish the two properties, we can apply the separation lemma 10.18, and prove the two properties individually. We have proved the first one. Now we prove the second one. (The separation lemma allows us to choose unary functions U_1 and U_2 separately for the two parts in order to satisfy the two properties, even though in the actual gadget construction the three occurrences of U_1 must be the same, and similarly for U_2, in order to produce a symmetric Q.)

The idea for the proof of the second property on P will be counterintuitive. Our goal is to choose unary functions U_1 and U_2 such that the function P has a full-rank matrix. We will do this by a nonzero unary function U_1 such that $F^{x_1=U_1}$ has a *singular* matrix. (This should be surprising as we seem to go the opposite direction.) However, once $F^{x_1=U_1}$ is degenerate, this effectively severs the bottom path in this gadget P. (This entanglement on the path connecting the two copies of F on the lower two corners makes it difficult to analyze P.) Consequently the matrix $P_{x_2x_3,y_2y_3}$ become a tensor product of two matrices $A_{x_2,y_2} \otimes B_{x_3,y_3}$. We then aim to guarantee that both A_{x_2,y_2} and B_{x_3,y_3} are nonsingular 2×2 matrices.

Since $F \notin \langle \mathcal{T} \rangle$, by Lemma 10.22 there exists $U_1 \not\equiv 0$ such that $F^{x_1=U_1}$ is degenerate, and $F^{x_1=U_1} = L_L \otimes R_L$, or in more detail, $F^{x_1=U_1}(z_3, z_2) = L_L(z_3)R_L(z_2)$. L_L and R_L are not the zero function. We also want the matrix form A_{x_2,y_2} of $F^{x_3=L_L}$ to be nonsingular. In the notation of Lemma 10.23, by the 1-1 correspondence from $\mathcal{U}_1$ to $\mathcal{U}_3$, $U_1 \in \mathcal{U}_1$ gives L_L and then gives a corresponding $L_L^{\perp} \in \mathcal{U}_3$. Each one of (at most two) $U_1 \in \mathcal{U}_1$ gives a unique $L_L^{\perp} \in \mathcal{U}_3$. We want to choose a $U_1 \in \mathcal{U}_1$ such that its corresponding $L_L \notin \mathcal{U}_3$. By the 1-1

https://doi.org/10.1017/9781107477063.011 Published online by Cambridge University Press

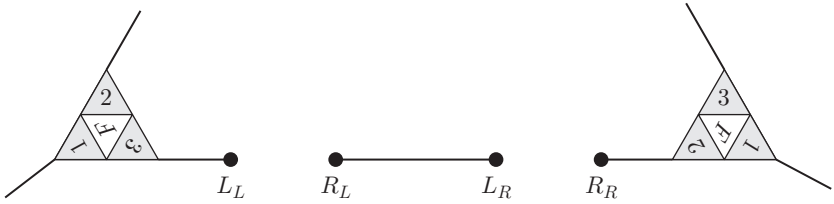

Figure 10.4. Replace $F^{x_1=U_1}$ by $L_L \otimes R_L$, and $F^{x_1=U_2}$ by $L_R \otimes R_R$.

correspondence, this is equivalent to choosing some unary $U \in \mathcal{U}_3$, such that $U^\perp \notin \mathcal{U}_3$. Such a $U \in \mathcal{U}_3$, by the inverse map of the 1-1 correspondence, gives us the desired $U_1 \in \mathcal{U}_1$.

We have a similar requirement for U_2 and $B = F^{x_2=R_R}$, on the right half of the gadget P: $U_2 \in \mathcal{U}_1$, $F^{x_1=U_2} = L_R \otimes R_R$, and $R_R \notin \mathcal{U}_2$. Suppose $F^{x_1=U_1} = L_L \otimes R_L$ and $F^{x_1=U_2} = L_R \otimes R_R$, then we can replace them by unary functions, and combine $R_L^{\mathrm{T}} L_R$ to get a scalar factor c. Then $P_{x_2x_3,y_2y_3}$ is $cA_{x_2,y_2} \otimes B_{x_3,y_3}$, where A_{x_2,y_2} and B_{x_3,y_3} are the matrix forms for $F^{x_3=L_L}$ and $F^{x_2=R_R}$ respectively (Figure 10.4). So we also want $c = R_L^{\mathrm{T}} L_R \neq 0$, in addition to A_{x_2,y_2} and B_{x_3,y_3} being nonsingular.

Note that if we write in matrix form for P from left to right (see Figures 10.3 and 10.4), we have another 4×4 matrix form for P as $P_{y_2x_2,y_3x_3} = F_{x_2x_1,x_3}(L_L R_L^{\mathrm{T}})(L_R R_R^{\mathrm{T}}) F_{x_2,x_3x_1}$. Taking out the dot product value $c = R_L^{\mathrm{T}} L_R$ (a scalar), the remainder of the function is $(F_{x_2x_1,x_3} L_L)(R_R^{\mathrm{T}} F_{x_2,x_3x_1})$, a product of a 4×1 matrix with a 1×4 matrix. This matrix has rank 1. However, this matrix form for P is not the same as $P_{x_2x_3,y_2y_3}$, which is a rotated version. It is $P_{x_2x_3,y_2y_3}$ that we want to ensure that it has rank 4.

To summarize for P, for the second property, we identify three conditions whose conjunction is sufficient.

Condition (1): $F^{x_1=U_1} = L_L \otimes R_L$ is degenerate and $F^{x_3=L_L}$ is nondegenerate.
Condition (2): $F^{x_1=U_2} = L_R \otimes R_R$ is degenerate and $F^{x_2=R_R}$ is nondegenerate.
Condition (3): $R_L^{\mathrm{T}} L_R \neq 0$.

There are three cases, depending on $\mathcal{U}_3$, where one cannot pick U_1 to satisfy condition (1). (We will deal with conditions (2) and (3) separately.)

a. $|\mathcal{U}_3| = 1$ and for the unique $U \in \mathcal{U}_3$, it also holds that $U^\perp \in \mathcal{U}_3$.
b. $|\mathcal{U}_3| = 2$ and for both $U \in \mathcal{U}_3$, it also holds that $U^\perp \cong U \in \mathcal{U}_3$.
c. $|\mathcal{U}_3| = 2$ and $\mathcal{U}_3 = \{U, U^\perp\}$.

Now we will use the given binary signatures $P_H \notin \langle H\mathcal{E} \rangle$, $P_Z \notin \langle Z\mathcal{E} \rangle$, and $S_Z \notin \langle Z\mathcal{M} \rangle$.

In case (a.): $U^\perp \cong U$, and thus $U \cong [1, \mathrm{i}]$ or $[1, -\mathrm{i}]$.

https://doi.org/10.1017/9781107477063.011 Published online by Cambridge University Press

If $U \cong [1, \mathrm{i}]$ (resp. $[1, -\mathrm{i}]$), we show that $S_{Z_1}U$ (resp. $S_{Z_2}U$) does not belong to $\mathcal{U}_3$. Because the binary $S_{Z_1} \notin \langle Z_1\mathcal{M}\rangle$, in matrix form $S_{Z_1} = Z_1TZ_1^{\mathrm{T}}$ for some $T \notin \langle\mathcal{M}\rangle$. Note that $\mathcal{U} \subset \mathcal{M}$, so that if the binary signature T is degenerate, then $T \in \langle\mathcal{M}\rangle$, a contradiction. Hence not only $T \notin \langle\mathcal{M}\rangle$, but also $T \notin \mathcal{M}$, that is $T(1, 1) \neq 0$. Note that the matrix-vector product $Z_1^{\mathrm{T}}U = \begin{bmatrix} 1 & \mathrm{i} \\ 1 & -\mathrm{i} \end{bmatrix} \begin{bmatrix} 1 \\ \mathrm{i} \end{bmatrix} \cong \begin{bmatrix} 0 \\ 1 \end{bmatrix}$. Then we calculate $U^{\mathrm{T}}S_{Z_1}U = U^{\mathrm{T}}Z_1TZ_1^{\mathrm{T}}U \cong \begin{bmatrix} 0 & 1 \end{bmatrix} T \begin{bmatrix} 0 \\ 1 \end{bmatrix} = T(1, 1) \neq 0$. Hence, $S_{Z_1}U \ncong [1, \mathrm{i}]$, and $S_{Z_1}U \ncong [0, 0]$. In this case, $F^{x_3=S_{Z_1}U}$ is nondegenerate, and we will modify the construction in Figure 10.3 by adding the binary gadget with signature S_{Z_1} to replace the three edges whose endpoints are small triangles both marked by 3 in the gadget. (This modification does change two edges in the construction of S; but it does not affect what has been proved for S, since S_{Z_1} is nondegenerate. The same is true for case (b.) and (c.) below.) The proof for $S_{Z_2}U$ is similar.

In case (b.), $\mathcal{U}_3 = \{[1, \mathrm{i}], [1, -\mathrm{i}]\}$.

We show in this case, one of $P_{Z_1}[1, \mathrm{i}]$ or $P_{Z_1}[1, -\mathrm{i}] \notin \mathcal{U}_3$. Because $P_{Z_1} \notin \langle Z_1\mathcal{E}\rangle$, in matrix form $P_{Z_1} = Z_1TZ_1^{\mathrm{T}}$ for some $T \notin \langle\mathcal{E}\rangle$. We claim that at least one of the two columns $\begin{bmatrix} e \\ f \end{bmatrix}$ of T have both entries nonzero. Otherwise, either T is degenerate or $T \in \mathcal{E}$, in either case $T \in \langle\mathcal{E}\rangle$, a contradictoion. If the first (resp. second) column has this property, $P_{Z_1}[1, -\mathrm{i}]$ (resp. $P_{Z_1}[1, \mathrm{i}]$) does not belong to $\mathcal{U}_3$. Indeed, if it is the first case,

$$P_{Z_1} \begin{bmatrix} 1 \\ -\mathrm{i} \end{bmatrix} = Z_1TZ_1^{\mathrm{T}} \begin{bmatrix} 1 \\ -\mathrm{i} \end{bmatrix} \cong Z_1T \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ \mathrm{i} & -\mathrm{i} \end{bmatrix} \begin{bmatrix} e \\ f \end{bmatrix} \ncong [1, \pm\mathrm{i}].$$

In this case, $F^{x_3=P_{Z_1}[1,-\mathrm{i}]}$ (resp. $F^{x_3=P_{Z_1}[1,\mathrm{i}]}$) is nondegenerate, and we will modify the construction in Figure 10.3 by adding the binary gadget with signature P_{Z_1} to replace the three edges whose endpoints are small triangles both marked by 3 in the gadget.

In case (c.), we have $|\mathcal{U}_3| = 2$ and $\mathcal{U}_3 = \{U, U^{\perp}\}$. Hence $U \ncong U^{\perp}$, and U and $U^{\perp}$ are linearly independent. Then $U \not\perp U$, otherwise $U \cong 0$. Hence the dot product $U^{\mathrm{T}}U \neq 0$, and we may assume $U = [a, b]$ and $U^{\perp} = [b, -a]$ are unit vectors: $a^2 + b^2 = 1$. Let $H = \begin{bmatrix} a & b \\ b & -a \end{bmatrix}$, then H is an orthogonal matrix. Then it follows that one of $P_H[a, b]$ or $P_H[b, -a]$ does not belong to $\mathcal{U}_3$. The proof is similar with case (b). In case (c) we will modify the construction in Figure 10.3 by adding the binary gadget with signature P_H to replace the three edges whose endpoints are small triangles both marked by 3 in the gadget.

https://doi.org/10.1017/9781107477063.011 Published online by Cambridge University Press

The proof for Condition (2) is similar to Condition (1). The replacement in the construction of Figure 10.3 happens at the three edges connecting the copy of F with U_2 and the corner F (the three edges whose endpoints are small triangles both marked by 2 in the gadget).

Now consider Condition (3). Suppose $R_L^{\mathrm{T}} L_R = 0$. We separate out the case $R_L \in \{[1, \mathrm{i}], [1, -\mathrm{i}]\}$, or not. If $R_L \in \{[1, \mathrm{i}], [1, -\mathrm{i}]\}$, and $R_L^{\mathrm{T}} L_R = 0$, then $R_L = L_R = [1, \mathrm{i}]$ or $[1, -\mathrm{i}]$. In this case $R_L^{\mathrm{T}} S_{Z_1} L_R \neq 0$ or $R_L^{\mathrm{T}} S_{Z_2} L_R \neq 0$ by a simple calculation as before. If $R_L \notin \{[1, \mathrm{i}], [1, -\mathrm{i}]\}$, and $R_L^{\mathrm{T}} L_R = 0$, then $L_R \notin \{[1, \mathrm{i}], [1, -\mathrm{i}]\}$ as well. We can assume $R_L = [a, b]$ and $L_R = [b, -a]$ and $a^2 + b^2 = 1$. Then one of $R_L^{\mathrm{T}} P_H L_R$ and $R_L^{\mathrm{T}} P_H^{\mathrm{T}} L_R$ is not zero, also by a simple calculation as before, where $H = \begin{bmatrix} a & b \\ b & -a \end{bmatrix}$. For condition (3), the replacement in the construction of Figure 10.3 happens at the three edges connecting the copy of F with U_1 with the copy of F with U_2 (the three edges whose endpoints are small triangles marked by 2 and 3 respectively in the gadget).

If conditions (1), (2), and (3) all hold, then the gadget satisfies the second property, and the theorem is proved. For each condition, if it does not hold, we have modified the gadget construction by adding some binary functions to rectify the construction, and these binary functions are available by the conditions of the theorem. With these modifications to the construction in Figure 10.3, the proof of the theorem is complete. □

Remark: In the proof of Theorem 10.24 we used the separation lemma to satisfy various conditions in isolation. In particular, for the second property (for the arity 4 signature P in the lower part of the gadget), we argued that we *could* satisfy the property by choosing unary functions U_1 and U_2 to make $F^{x_1=U_1}$ and $F^{x_1=U_2}$ degenerate. When the final gadget is produced by simultaneously satisfying both the first condition (for S) and the second condition (for P), there is no expectation that this degeneracy will persist. In fact, the we proved that the first condition (for S) *could* be satisfied by choosing U_1 and U_2 to make $F^{x_1=U_1}$ and $F^{x_1=U_2}$ nondegenerate. There is some nonconstructiveness by using the separation lemma to argue for the success of such constructions. If necessary, one can make it constructive (in the sense of Turing computability), by looking more closely at the sets of polynomial equations (which could be large, but specifiable by at most polynomially many bits). The reduction proved to exist is a polynomial time reduction, but the proof does not give it explicitly.

We will prove the #P-hardness part of Theorem 10.14 by appealing to Theorem 2.12 for symmetric Holant* problems. For that purpose we need to construct appropriate *symmetric* binary signatures as well.

https://doi.org/10.1017/9781107477063.011 Published online by Cambridge University Press

Theorem 10.25. *Let $\mathcal{F}$ denote any one of the function sets $\langle H\mathcal{E}\rangle$ (for an orthogonal matrix H), $\langle Z\mathcal{E}\rangle$ or $\langle Z\mathcal{M}\rangle$ (for the matrix $Z = Z_1$ or Z_2). Suppose we can realize a symmetric ternary function $F \in \mathcal{F} - \langle\mathcal{T}\rangle$ and a binary function $P \notin \mathcal{F}$. Then we can realize a* symmetric *binary function $Q \notin \mathcal{F}$.*

Proof. We define the *symmetric* binary function $Q = PF^{x_1=U}P^{\mathrm{T}}$ in matrix form, for some unary function U. This Q is realizable by a gadget consisting of a path of three signatures P, $F^{x_1=U}$ and P^{T}. At both ends we have a copy of P and we connect the second variable of each copy of P to the two remaining variables of the symmetric $F^{x_1=U}$.

The essence of the proof is an appropriate holographic transformation. We are given $F \in \langle H\mathcal{E}\rangle$ (for an orthogonal matrix H), or $\langle Z\mathcal{E}\rangle$ or $\langle Z\mathcal{M}\rangle$. Let $M = H$ or Z depending on the cases. Let $F = M^{\otimes 3}F_1$ and $P = MP_1M^{\mathrm{T}}$ in matrix form. $F \notin \langle\mathcal{T}\rangle$ implies that $F_1 \notin \langle\mathcal{T}\rangle$. We also have $F^{x_1=U} = MF_1^{x_1=MU}M^{\mathrm{T}}$ in matrix form for the binary $F^{x_1=U}$. Let $Q_1 = P_1M^{\mathrm{T}}MF_1^{x_1=MU}M^{\mathrm{T}}MP_1^{\mathrm{T}}$; then

$$Q = MQ_1M^{\mathrm{T}} = MP_1M^{\mathrm{T}}MF_1^{x_1=MU}M^{\mathrm{T}}MP_1^{\mathrm{T}}M^{\mathrm{T}}.$$

Case (1) $\mathcal{F}$ is $\langle H\mathcal{E}\rangle$.

We take $M = H$. Since H is orthogonal, $Q = HP_1F_1^{x_1=HU}P_1^{\mathrm{T}}H^{\mathrm{T}}$. We have $F_1 \notin \langle\mathcal{T}\rangle$. We also have $P_1 \notin \langle\mathcal{E}\rangle$, since $P \notin \langle H\mathcal{E}\rangle$. But by $F \in \langle H\mathcal{E}\rangle$, we have $F_1 \in \langle\mathcal{E}\rangle$. By the condition $F_1 \notin \langle\mathcal{T}\rangle$, we must actually have $F_1 \in \mathcal{E}$, because F_1 only has arity 3 and therefore any tensor factorization will put F_1 in $\langle\mathcal{T}\rangle$. Being symmetric and nondegenerate, $F_1 = [u, 0, 0, v]$, where $u \neq 0$ and $v \neq 0$. We only need to prove $Q_1 = P_1F_1^{x_1=HU}P_1^{\mathrm{T}} \notin \langle\mathcal{E}\rangle$, which is the same as $Q \notin \langle H\mathcal{E}\rangle$. Because we can pick any $U' = HU$, for any x, y, we can realize $F_1^{x_1=HU} = [x, 0, y]$. Suppose $P_1 = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$. Then $Q_1 = \begin{bmatrix} a^2x + b^2y & acx + bdy \\ acx + bdy & c^2x + d^2y \end{bmatrix}$.

We need $Q_1 \notin \langle\mathcal{E}\rangle$. This is translated into three conditions: (1) Q_1 is nondegenerate, (2) Q_1 is not of the form $[*, 0, *]$, and (3) Q_1 is not of the form $[0, *, 0]$. By the separation lemma, we only need to prove that there is some $[x, y]$ to satisfy each condition individually. If $x \neq 0$ and $y \neq 0$, then $F_1^{x_1=HU} = [x, 0, y]$ in nondegenerate. Also P_1 is nondegenerate because $P_1 \notin \langle\mathcal{E}\rangle$. Thus Q_1 is nondegenerate. Again because $P_1 \notin \langle\mathcal{E}\rangle$, either $ac \neq 0$ or $bd \neq 0$. There exists some $[x, y]$ such that $acx + bdy \neq 0$; thus Q_1 is not of the form $[*, 0, *]$. Similarly, it is easy find $[x, y]$ such that Q_1 is not of the form $[0, *, 0]$.

Case (2) $\mathcal{F}$ is $\langle Z\mathcal{E}\rangle$.

Take $M = Z$. Note that $Z^{\mathrm{T}}Z \cong (\neq_2)$. We have $Q \cong ZP_1(\neq_2)F_1^{x_1=ZU}(\neq_2)P_1^{\mathrm{T}}Z^{\mathrm{T}} = ZQ_1Z^{\mathrm{T}}$. We have $F_1 \notin \langle\mathcal{T}\rangle$, $F_1 \in \mathcal{E}$, and $P_1 \notin \langle\mathcal{E}\rangle$. We only need

https://doi.org/10.1017/9781107477063.011 Published online by Cambridge University Press

to prove $Q_1 \notin \langle \mathcal{E} \rangle$. For any x, y, we can pick U to realize $(\neq_2)F_1^{x_1=ZU}(\neq_2) = [x, 0, y]$. This is seen by the fact that $(\neq_2)[x, 0, y](\neq_2) = [y, 0, x]$. The remaining proof is the same as case (1) for $\langle H\mathcal{E} \rangle$.

Case (3) $\mathcal{F}$ is $\langle Z\mathcal{M} \rangle$.

Take $M = Z$. Since $Z^{\mathrm{T}}Z \cong (\neq_2)$, we have $Q_1 \cong P_1(\neq_2)F_1^{x_1=ZU}(\neq_2)P_1^{\mathrm{T}}$.

We have $F_1 \notin \langle \mathcal{T} \rangle$, and $F_1 \in \langle \mathcal{M} \rangle$. Again because F_1 has arity 3, from these two conditions we conclude $F_1 \in \mathcal{M}$, since any tensor factorization for F_1 would place it in $\langle \mathcal{T} \rangle$. Being symmetric and nondegenerate, F_1 has the form $F_1 \cong [f, 1, 0, 0]$. Let $P_1 = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, and we have $P_1 \notin \langle \mathcal{M} \rangle$, since $P \notin \langle Z\mathcal{M} \rangle$. For any x, there is some U, such that $F_1^{x_1=ZU} = [x, 1, 0]$. We only need to prove that $Q_1 = P_1(\neq_2)F_1^{x_1=ZU}(\neq_2)P_1^{\mathrm{T}} \notin \langle \mathcal{M} \rangle$.

Because $P_1 \notin \langle \mathcal{M} \rangle$, certainly P_1 is nondegenerate. Therefore $\det Q_1 \neq 0$, and Q_1 is nondegenerate. Also, the entry $P_1(1, 1) = d \neq 0$, because $P_1 \notin \mathcal{M}$. It follows that there exists x such that $Q_1(1, 1) = 2cd + d^2x \neq 0$. Hence, $Q_1 \notin \langle \mathcal{M} \rangle$. □

Now we are ready to finish the proof Theorem 10.14.

Proof of Theorem 10.14 (#P-hardness part)

Suppose $\mathcal{F} \not\subseteq \langle \mathcal{T} \rangle$, $\mathcal{F} \not\subseteq \langle H\mathcal{E} \rangle$ for all orthogonal H, $\mathcal{F} \not\subseteq \langle Z\mathcal{E} \rangle$ and $\mathcal{F} \not\subseteq \langle Z\mathcal{M} \rangle$ for $Z = Z_1$ and Z_2. By Lemma 10.19, we can realize functions of arity of 2 or 3 *not* belonging to these function sets respectively.

The conditions in Theorem 10.24 are satisfied, so we can realize a symmetric ternary function $Q_3 \notin \langle \mathcal{T} \rangle$ (with the help of those binary functions). Certainly Q_3 is nondegenerate. If Holant*$(\{Q_3\})$ is #P-hard, then the theorem is proved. Otherwise, by Theorem 2.12 for symmetric Holant* problems, Q_3 belongs to one of the special function families listed in the theorem. It can be shown that these are precisely restrictions of $\langle H\mathcal{E} \rangle$, $\langle Z\mathcal{E} \rangle$ or $\langle Z\mathcal{M} \rangle$ to symmetric signatures. By Theorem 10.25, we can realize a symmetric binary function Q_2 not in this family. By Theorem 2.12 again, Holant*$(\{Q_3, Q_2\})$ is #P-hard, and therefore Holant*$(\mathcal{F})$ is also #P-hard. □

https://doi.org/10.1017/9781107477063.011 Published online by Cambridge University Press

https://doi.org/10.1017/9781107477063.011 Published online by Cambridge University Press