2

Fibonacci Gates and Holant* Problems

We start with a special family of signatures called Fibonacci gates. They will be an important building block for tractable Holant problems.

2.1 Fibonacci Gates

Consider the signature $f = [1, 0, 1, 1]$. What is the complexity of Holant(f)? It is plain that f satisfies $f_{k+2} = f_{k+1} + f_k$, for $0 \le k \le 1$, and this turns out to be significant.

Definition 2.1. For any $n \ge 1$, a signature $f = [f_0, f_1, \ldots, f_n]$ is a Fibonacci gate if

$$f_{k+2} = f_{k+1} + f_k, \quad 0 \le k \le n-2.$$

A set of signatures $\mathcal{F}$ is called Fibonacci if every signature in $\mathcal{F}$ is a Fibonacci gate.

Each Fibonacci gate f has two degrees of freedom, as f_0 and f_1 are arbitrary; the rest of the f's the other f_k's are determined by the linear recurrence. In particular, any unary signature is a Fibonacci gate. As with the ordinary Fibonacci sequence, the linear recurrence satisfied by f gives the following formula

$$f_k = c_1\phi^k + c_2\overline{\phi}^k, \quad 0 \le k \le n, \tag{2.1}$$

where ϕ is the golden ratio $\frac{1+\sqrt{5}}{2}$ and $\overline{\phi}$ is its conjugate $\frac{1-\sqrt{5}}{2}$, both roots of the quadratic polynomial $X^2 - X - 1$. To match the initial values f_0 and f_1, we have $n = 3$, $c_1 = (-\overline{\phi}/\sqrt{5})f_0 + (1/\sqrt{5})f_1$ and $c_2 = (\phi/\sqrt{5})f_0 + (-1/\sqrt{5})f_1$. For example, for $f = [1, 0, 1, 1]$, we have $c_1 = -\overline{\phi}/\sqrt{5}$ and $c_2 = \phi/\sqrt{5}$.

For $f = [1, 0, 1, 1]$, Eq. (2.1) can be written as a sum of tensor products:

$$f = c_1 \begin{bmatrix} 1 \\ \phi \end{bmatrix}^{\otimes 3} + c_2 \begin{bmatrix} 1 \\ \overline{\phi} \end{bmatrix}^{\otimes 3}. \tag{2.2}$$

https://doi.org/10.1017/9781107477063.003 Published online by Cambridge University Press

This lists the values of f as a column vector indexed by its input assignments in lexicographic order.

Consider the matrix $M = \begin{bmatrix} 1 & 1 \\ \phi & \overline{\phi} \end{bmatrix}$. We have $M^{\otimes 3}\left(c_1 \begin{bmatrix} 1 \\ 0 \end{bmatrix}^{\otimes 3} + c_2 \begin{bmatrix} 0 \\ 1 \end{bmatrix}^{\otimes 3}\right) = f$. This suggests that we take a *holographic transformation* by the matrix M.

Recall that holographic transformations are defined in Section 1.3.2 of Chapter 1. By Theorem 1.5, for a bipartite Holant problem Holant($\mathcal{F} \mid \mathcal{G}$), the holographic transformation does not change the value

$$\text{Holant}(\Omega; \mathcal{F} \mid \mathcal{G}) = \text{Holant}(\Omega'; \mathcal{F}M \mid M^{-1}\mathcal{G}),$$

where Ω' is the transformed signature grid.

Let's review how we can treat a problem such as Holant(f) as a bipartite Holant problem, and in concrete terms what Theorem 1.5 says in this case. We can view Holant(f) as computing the following tensor *contraction* operation. Given a 3-regular graph Γ where each vertex is labeled f, first we replace each edge e by a path of length 2, and label the new vertex the binary EQUALITY function $(=_2) = [1, 0, 1]$. This transforms the Holant problem Holant(f) to a bipartite Holant problem Holant($=_2 \mid f$) on the Edge-Vertex incidence graph without changing its value on Γ. Next we form the tensor products $\mathcal{R}$ of all left-hand-side copies of $(=_2)$, and $\mathcal{G}$ of all right-hand-side copies of f, respectively. If Γ has m edges, these two tensors, $\mathcal{R}$ and $\mathcal{G}$, both have dimension 2^{2m}, each representing a function of arity $2m$, and there is a one-to-one correspondence between the variables according to the incidence relation in Γ. The contraction operation $\langle \mathcal{R}, \mathcal{G} \rangle$ is $\sum_{x_1, x_2, \ldots, x_{2m}=0,1} \mathcal{R}(x_1, x_2, \ldots, x_{2m})\mathcal{G}(x_1, x_2, \ldots, x_{2m})$. If we view $\mathcal{R}$ as a (row) vector of dimension 2^{2m} and $\mathcal{G}$ as a (column) vector of the same dimension, then the contraction $\langle \mathcal{R}, \mathcal{G} \rangle$ is the dot product of the two vectors.

Now we consider a transformation $(M^{-1})^{\otimes 2m}$ on $\mathcal{G}$ and $M^{\otimes 2m}$ on $\mathcal{R}$, namely

$$\mathcal{G} \mapsto \mathcal{G}' = (M^{-1})^{\otimes 2m}\mathcal{G}, \quad \text{and} \quad \mathcal{R} \mapsto \mathcal{R}' = \mathcal{R}M^{\otimes 2m}. \tag{2.3}$$

It is clear that

$$\langle \mathcal{R}', \mathcal{G}' \rangle = \langle \mathcal{R}, \mathcal{G} \rangle.$$

The transformation in (2.3) can be carried out by distributing M and M^{-1} to each left-hand side copy of $(=_2)$ and right-hand side copy of f. Thus we can consider $(=_2) \mapsto (=_2)M^{\otimes 2}$, where $(=_2)$ is written as a row vector $(1, 0, 0, 1)$, and $f \mapsto (M^{-1})^{\otimes 3} f$, where f is given in (2.2). It follows that

$$(M^{-1})^{\otimes 3} f = c_1 \begin{bmatrix} 1 \\ 0 \end{bmatrix}^{\otimes 3} + c_2 \begin{bmatrix} 0 \\ 1 \end{bmatrix}^{\otimes 3},$$

https://doi.org/10.1017/9781107477063.003 Published online by Cambridge University Press

and in symmetric signature notation this is $(M^{-1})^{\otimes 3} f = [c_1, 0, 0, c_2]$. Note that signatures of the form $[*, 0, \ldots, 0, *]$ force all incident edges to take the same value of 0 or 1. They are called Generalized Equalities, or Gen-Eq for short.

Recall that if we write a binary signature f as a matrix F, with entry $f(i, j)$ at row i and column j (where $i, j = 0, 1$), then MFM^{T} is the matrix form of $M^{\otimes 2} f$, where we write f as a column vector with entries ordered lexicographically. Similarly, $M^{\mathrm{T}} FM$ is the matrix form of $fM^{\otimes 2}$, where we write f as a row vector. It is clear that the identity matrix I is the matrix form of $(=_2)$. It follows that for $M = \begin{bmatrix} 1 & 1 \\ \phi & \overline{\phi} \end{bmatrix}$, $(=_2)M^{\otimes 2}$ takes the form $M^{\mathrm{T}} M = \begin{bmatrix} 1+\phi^2 & 0 \\ 0 & 1+\overline{\phi}^2 \end{bmatrix}$, which is $[1 + \phi^2, 0, 1 + \overline{\phi}^2]$ in symmetric signature notation. Thus both sides of the bipartite Holant problem $\mathrm{Holant}((=_2)M^{\otimes 2} \mid (M^{-1})^{\otimes 3} f)$ are Gen-Eq.

This means that the Holant value of $\mathrm{Holant}(=_2 \mid f)$, which is the same as the Holant value of $\mathrm{Holant}((=_2)M^{\otimes 2} \mid (M^{-1})^{\otimes 3} f)$, can be easily computed: On each connected component of the given graph Γ, any assignment on a single edge can be uniquely propagated to the entire component. The Holant value on Γ is the product over all connected components.

This can be generalized to any set $\mathcal{F}$ of Fibonacci gates. If $f = [f_0, f_1, \ldots, f_n] \in \mathcal{F}$ is a Fibonacci gate, then (2.1) can be written as

$$f = c_1 \begin{bmatrix} 1 \\ \phi \end{bmatrix}^{\otimes n} + c_2 \begin{bmatrix} 1 \\ \overline{\phi} \end{bmatrix}^{\otimes n}. \tag{2.4}$$

Thus, the same holographic transformation $f \mapsto (M^{-1})^{\otimes n} f$ by M transforms f to a Gen-Eq. This shows that $\mathrm{Holant}(\mathcal{F})$ is tractable.

Exercise: Show that there is an orthogonal matrix $M \in \mathbb{C}^{2\times 2}$ that transforms every signature of the form (2.4) to a Gen-Eq. Use this fact to prove that $\mathrm{Holant}(\mathcal{F})$ is tractable for a set of Fibonacci gates $\mathcal{F}$.

There is an alternative proof of the tractability of $\mathrm{Holant}(\mathcal{F})$ for any finite set of Fibonacci gates $\mathcal{F}$. We turn to this approach next.

Definition 2.2. For any $n \geq 1$, and a parameter $\lambda \in \mathbb{C}$, a signature $f = [f_0, f_1, \ldots, f_n]$ is a generalized Fibonacci gate (with parameter λ) if

$$f_{k+2} = \lambda f_{k+1} + f_k, \quad 0 \leq k \leq n-2. \tag{2.5}$$

A set of signatures $\mathcal{F}$ is called generalized Fibonacci if for some $\lambda \in \mathbb{C}$, every signature in $\mathcal{F}$ is a generalized Fibonacci gate with parameter λ.

https://doi.org/10.1017/9781107477063.003 Published online by Cambridge University Press

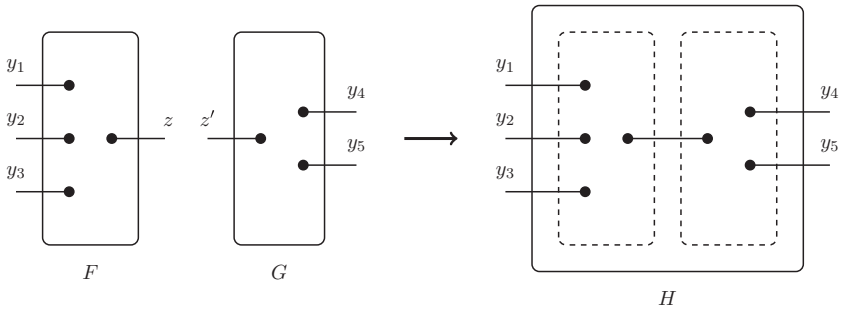

Figure 2.1. First operation.

Theorem 2.3. *For any finite set of generalized Fibonacci gates $\mathcal{F}$, the Holant problem Holant($\mathcal{F}$) is computable in polynomial time.*

Proof. If $\Gamma_1, \Gamma_2, \ldots, \Gamma_\kappa$ are the connected components of a graph Γ, then

$$\text{Holant}_\Gamma = \prod_{j=1}^{\kappa} \text{Holant}_{\Gamma_j}.$$

So we only need to consider connected graphs as inputs.

Suppose Γ has n nodes and m edges. First we cut all the edges in Γ. A node with degree d can be viewed as an $\mathcal{F}$-gate with d dangling edges. Now step by step we merge two dangling edges into one regular edge in the original graph, until we recover Γ after m steps. We prove that all the intermediate $\mathcal{F}$-gates still have generalized Fibonacci signatures with the same parameter λ, and at every step we can compute the intermediate signature in polynomial time. After m steps we get Γ as an $\mathcal{F}$-gate with no dangling edge; the only value of its signature is the Holant value we want. To carry this out, we only need to prove that it is true for one single step. There are two cases, depending on whether the two dangling edges to be merged are in the same component or not. These two operations are illustrated in Figures 2.1 and 2.2.

In the first case, the two dangling edges belong to two components before their merging (Figure 2.1). Let F have dangling edges $y_1, \ldots, y_s, z$ and G have dangling edges $y_{s+1}, \ldots, y_{s+t}, z'$. After merging z with z', we have a new $\mathcal{F}$-gate H with dangling edges $y_1, \ldots, y_{s+t}$. Inductively the signatures of F and G are both generalized Fibonacci gates with the same parameter λ. We show that this remains so for the resulting $\mathcal{F}$-gate H.

We first prove that H is symmetric. We only need to show that the value of H is unchanged if the values of two inputs are exchanged. Because F and G are symmetric, if both inputs are from $\{y_1, \ldots, y_s\}$ or from

https://doi.org/10.1017/9781107477063.003 Published online by Cambridge University Press

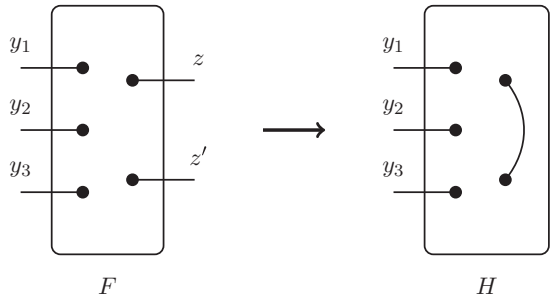

Figure 2.2. Second operation.

$\{y_{s+1}, \ldots, y_{s+t}\}$, the value of H is clearly unchanged. Suppose one input is from $\{y_1, \ldots, y_s\}$ and the other is from $\{y_{s+1}, \ldots, y_{s+t}\}$. By the symmetry of F and G we may assume these two inputs are y_1 and y_{s+1}. Thus we will fix an arbitrary assignment for $y_2, \ldots, y_s, y_{s+2}, \ldots, y_{s+t}$, and we want to show $H(0, y_2, \ldots, y_s, 1, y_{s+2}, \ldots, y_{s+t}) = H(1, y_2, \ldots, y_s, 0, y_{s+2}, \ldots, y_{s+t})$.

We will suppress the fixed values $y_2, \ldots, y_s, y_{s+2}, \ldots, y_{s+t}$ and denote

$$F_{xz} = F(x, y_2, \ldots, y_s, z),$$
$$G_{yz} = G(y, y_{s+2}, \ldots, y_{s+t}, z), \quad \text{and}$$
$$H_{xy} = H(x, y_2, \ldots, y_s, y, y_{s+2}, \ldots, y_{s+t}).$$

Then by the definition of Holant, $H_{ab} = F_{a0}G_{b0} + F_{a1}G_{b1}$, for $a, b \in \{0, 1\}$.

Because F and G are generalized Fibonacci gates with parameter λ, we have

$$H_{01} = F_{00}G_{10} + F_{01}(\lambda G_{01} + G_{00}), \quad \text{and} \quad H_{10} = F_{10}G_{00} + (\lambda F_{01} + F_{00})G_{01}.$$

By the symmetry of F and G, we have $H_{01} = H_{10}$.

Now we show that $H(y_1, \ldots, y_{s+t})$ is also a generalized Fibonacci gate with parameter λ. Since we have proved that H is symmetric, we can choose any two input variables to prove it being Fibonacci. Again, we choose y_1 and y_{s+1}. (This assumes that y_1 and y_{s+1} exist, i.e., F and G are not unary functions. If either one of them is unary, the proof is just as easy.) For any fixed values of all other variables, we have $H_{00} = F_{00}G_{00} + F_{01}G_{01}$, $H_{01} = F_{00}G_{10} + F_{01}G_{11}$, and $H_{11} = F_{10}G_{10} + F_{11}G_{11}$. Now using the fact that both F and G are generalized Fibonacci gates with parameter λ, it follows that $H_{11} = \lambda H_{01} + H_{00}$.

If the first two terms of the signatures of F and G are f_0, f_1 and g_0, g_1 respectively, then the first two terms of the signature H are $h_0 = f_0g_0 + f_1g_1$ and $h_1 = f_0g_1 + f_1g_0 + \lambda f_1g_1$.

https://doi.org/10.1017/9781107477063.003 Published online by Cambridge University Press

Next we consider the second case, where the two dangling edges to be merged are in the same component (Figure 2.2). Obviously, the signature for the new gate H is also symmetric. If $F = [f_0, f_1, \ldots, f_n]$ is the signature before the merging operation, then the signature after the merging operation is $H = [f_0 + f_2, f_1 + f_3, \ldots, f_{n-2} + f_n]$. Such an operation preserves the linear recurrence. It follows that H is also a generalized Fibonacci gate with parameter λ, and the signature has already been computed. □

We remark that the proof of tractability in Theorem 2.3 is also valid for an infinite set $\mathcal{F}$ of generalized Fibonacci gates with the same parameter λ; however, for an infinite set $\mathcal{F}$ the input size should include the description of the signatures used in the signature grid. See Remark 1 after Theorem 2.12.

2.2 Orthogonal Transformation of Fibonacci Gates

By Definition 2.1, Fibonacci gates correspond to the case $\lambda = 1$ in Definition 2.2. How much generality does an arbitrary parameter λ provide? Can we recapture the tractability of generalized Fibonacci gates by holographic transformation as we did for $\lambda = 1$?

The recurrence relation $f_{k+2} = \lambda f_{k+1} + f_k$ has the characteristic polynomial $X^2 - \lambda X - 1$. If $\lambda \neq \pm 2\mathrm{i}$ then there are two distinct eigenvalues $\alpha, \beta \neq \pm \mathrm{i}$, and a general solution to the recurrence has the form $f_k = a\alpha^k + b\beta^k$, for some constants a and b determined by f_0 and f_1. Note that $\alpha\beta = -1$. More revealing is the following expression in terms of tensor products:

$$f = a \begin{bmatrix} 1 \\ \alpha \end{bmatrix}^{\otimes n} + b \begin{bmatrix} 1 \\ \beta \end{bmatrix}^{\otimes n}. \tag{2.6}$$

Note that the two vectors $\begin{pmatrix} 1 & \alpha \end{pmatrix}^{\mathrm{T}}$ and $\begin{pmatrix} 1 & \beta \end{pmatrix}^{\mathrm{T}}$ are orthogonal, i.e., their dot product is zero, and $1 + \alpha^2, 1 + \beta^2 \neq 0$. We can scale the entries as follows. Let

$$T = \begin{bmatrix} t_{00} & t_{01} \\ t_{10} & t_{11} \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{1+\alpha^2}} & \frac{1}{\sqrt{1+\beta^2}} \\ \frac{\alpha}{\sqrt{1+\alpha^2}} & \frac{\beta}{\sqrt{1+\beta^2}} \end{bmatrix}.$$

Then T is an orthogonal matrix. Let $a' = a(1 + \alpha^2)^{n/2}$ and $b' = b(1 + \beta^2)^{n/2}$; then we have

$$f = a' \begin{bmatrix} t_{00} \\ t_{10} \end{bmatrix}^{\otimes n} + b' \begin{bmatrix} t_{01} \\ t_{11} \end{bmatrix}^{\otimes n} = T^{\otimes n} \left(a' \begin{bmatrix} 1 \\ 0 \end{bmatrix}^{\otimes n} + b' \begin{bmatrix} 0 \\ 1 \end{bmatrix}^{\otimes n} \right). \tag{2.7}$$

https://doi.org/10.1017/9781107477063.003 Published online by Cambridge University Press

Hence the holographic transformation $(T^{-1})^{\otimes n} f$ is the Gen-Eq signature $a' \begin{bmatrix} 1 \\ 0 \end{bmatrix}^{\otimes n} + b' \begin{bmatrix} 0 \\ 1 \end{bmatrix}^{\otimes n} = [a', 0, \ldots, 0, b']$. Meanwhile the holographic transformation on the left-hand side keeps the binary Equality $(=_2)$ unchanged by T:

$$(=_2) \mapsto (=_2) T^{\otimes 2} = (=_2), \tag{2.8}$$

because T is an orthogonal matrix, and in matrix form $(=_2)T^{\otimes 2}$ is $T^{\mathrm{T}} I T = I$.

Thus for any finite set of generalized Fibonacci gates $\mathcal{F}$ with parameter $\lambda \neq \pm 2\mathrm{i}$, the Holant problem Holant($\mathcal{F}$), being equivalent to Holant($=_2 | \mathcal{F}$), is tractable. This is an alternative proof of Theorem 2.3, when $\lambda \neq \pm 2\mathrm{i}$.

By a limiting argument, one can give a proof of tractability by this *holographic* approach when $\lambda = \pm 2\mathrm{i}$ as well. We will, however, refrain from that, because intrinsically there is a difference for the signatures when $\lambda = \pm 2\mathrm{i}$.

Every generalized Fibonacci gate with $\lambda \neq \pm 2\mathrm{i}$ can be transformed to a Gen-Eq by some orthogonal matrix. By composition, for any $\lambda, \mu \neq \pm 2\mathrm{i}$, any generalized Fibonacci gate with parameter λ can also be transformed by some orthogonal matrix to another generalized Fibonacci gate with parameter μ. We can take the basic Fibonacci gate with $\lambda = 1$, as well as Gen-Eq, as a normal form for this class. The orthogonal transformation between these two is the normalized version of the matrix M from Section 2.1: $\begin{bmatrix} \frac{1}{\sqrt{1+\phi^2}} & \frac{1}{\sqrt{1+\bar{\phi}^2}} \\ \frac{\phi}{\sqrt{1+\phi^2}} & \frac{\bar{\phi}}{\sqrt{1+\bar{\phi}^2}} \end{bmatrix}$.

An invertible holographic transformation maps nondegenerate signatures to nondegenerate signatures. This is because if $fM^{\otimes n} = (a_1, b_1) \otimes \cdots \otimes (a_n, b_n)$, then $f = (a'_1, b'_1) \otimes \cdots \otimes (a'_n, b'_n)$, where $(a'_k, b'_k) = (a_k, b_k)M^{-1}$. The same argument works for contravariant tensors. Suppose f is a generalized Fibonacci gate satisfying (2.6). The following are equivalent: (1) f is nondegenerate, (2) $ab \neq 0$ in (2.6), (3) its Gen-Eq form $[a', 0, \ldots, 0, b']$ in (2.7) is nondegenerate, and (4) $a'b' \neq 0$.

Define

$$\mathscr{F} = \{f \mid f \text{ satisfies (2.5) for some } \lambda \neq \pm 2\mathrm{i}\} \cup \text{Gen-Eq}. \tag{2.9}$$

Theorem 2.4. *For any $f \in \mathscr{F}$ in (2.9),*

1. *There exists an orthogonal T such that Tf is a* Gen-Eq.
2. *There exists an orthogonal T such that Tf is a Fibonacci gate satisfying Definition 2.1.*
3. *For all orthogonal T, $Tf \in \mathscr{F}$.*

Proof. We only need to prove the last item. By composition and the invertibility of the transformation, we may assume f is a Gen-Eq function $[a', 0, \ldots, 0, b']$.

https://doi.org/10.1017/9781107477063.003 Published online by Cambridge University Press

If $T = \begin{bmatrix} t_{00} & t_{01} \\ t_{10} & t_{11} \end{bmatrix}$, then

$$Tf = a' \begin{bmatrix} t_{00} \\ t_{10} \end{bmatrix}^{\otimes n} + b' \begin{bmatrix} t_{01} \\ t_{11} \end{bmatrix}^{\otimes n}.$$

If any entry $t_{ij} = 0$, then by orthogonality $t_{\bar{i}\bar{j}} = 0$, where $\bar{x} = 1 - x$, and Tf is a Gen-Eq. Suppose T has no zero entries; then Tf satisfies a recurrence relation with the characteristic roots $\lambda_1 = t_{10}/t_{00}$ and $\lambda_2 = t_{11}/t_{01}$. We have $\lambda_1 \neq \lambda_2$ and $\lambda_1\lambda_2 = -1$, by the orthogonality of T. Hence the characteristic polynomial of the recurrence is $X^2 - (\lambda_1 + \lambda_2)X - 1$. Since $\lambda_1 \neq \lambda_2$, we have $\lambda_1 + \lambda_2 \neq \pm 2\mathrm{i}$. Thus $Tf \in \mathscr{F}$. □

We may consider Gen-Eq gates provide a normal form for generalized Fibonacci gates (with $\lambda \neq \pm 2\mathrm{i}$) under orthogonal transformations as in (2.7). Curiously, any Gen-Eq of arity at least 3 does not satisfy any recurrence $f_{k+2} = \lambda f_{k+1} + f_k$. (We can generalize the form of a second-order recurrence to $af_k + bf_{k+1} - af_{k+2} = 0$ for some $(a, b) \neq 0$. Then Gen-Eq does satisfy the recurrence with $(a, b) = (0, 1)$.) The proof of Theorem 2.4 also establishes the following: Other than Gen-Eq itself, any signature f expressible as (2.7), i.e., an orthogonal transformation of Gen-Eq, is in fact a generalized Fibonacci gate.

Exercise: Prove that generalized Fibonacci gates with parameter $\lambda = \pm 2\mathrm{i}$ cannot be transformed to any member of $\mathscr{F}$ in (2.9).

We now ask, what is the right perspective for generalized Fibonacci gates with parameter $\lambda = \pm 2\mathrm{i}$ under holographic transformations?

Let $\mathcal{F}$ be such a set of signatures. To show Holant($\mathcal{F}$) is tractable, we first get rid of degenerate signatures as follows. Each degenerate signature can be replaced by a set of unary signatures. Then combine a unary signature with any $f \in \mathcal{F}$ satisfying (2.5) produces another signature f' with $\mathrm{arity}(f') = \mathrm{arity}(f) - 1$, which also satisfies (2.5) with the same λ. So we may assume every signature in $\mathcal{F}$ is nondegenerate and has arity ≥ 2.

If every vertex in the given instance graph H has degree ≤ 2, then the Holant value Holant($\mathcal{F}$) can be obtained by computing it as a matrix product and then taking the trace, per each connected component.

Each $f \in \mathcal{F}$ has the form $f_k = ck\mu^{k-1} + d\mu^k$ ($0 \leq k \leq n$ and $\mu = \pm\mathrm{i}$), where $X^2 - 2\mu X - 1$ has double root μ. We will suppose $\mu = \mathrm{i}$. The case

https://doi.org/10.1017/9781107477063.003 Published online by Cambridge University Press

$\mu = -\mathrm{i}$ is similar. Consider

$$f' = [c', d', 0, \ldots, 0] = c' \begin{bmatrix} 1 \\ 0 \end{bmatrix}^{\otimes n} + d' \sum_{j=1}^{n} \left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}^{\otimes (j-1)} \otimes \begin{bmatrix} 0 \\ 1 \end{bmatrix} \otimes \begin{bmatrix} 1 \\ 0 \end{bmatrix}^{\otimes (n-j)} \right\}, \tag{2.10}$$

where we set $c' = d - cn\mathrm{i}/2$ and $d' = c\mathrm{i}/2$. Let

$$Z = \begin{bmatrix} 1 & 1 \\ \mathrm{i} & -\mathrm{i} \end{bmatrix}. \tag{2.11}$$

We apply $Z^{\otimes n}$ to f' by distributing each Z among the factors of the tensor product in (2.10):

$$Z^{\otimes n} f' = c' \begin{bmatrix} 1 \\ \mathrm{i} \end{bmatrix}^{\otimes n} + d' \sum_{j=1}^{n} \left\{ \begin{bmatrix} 1 \\ \mathrm{i} \end{bmatrix}^{\otimes (j-1)} \otimes \begin{bmatrix} 1 \\ -\mathrm{i} \end{bmatrix} \otimes \begin{bmatrix} 1 \\ \mathrm{i} \end{bmatrix}^{\otimes (n-j)} \right\}. \tag{2.12}$$

Then $Z^{\otimes n} f' = f$. To wit, any entry of $Z^{\otimes n} f'$ indexed by a bit pattern in $\{0, 1\}^n$ of Hamming weight k is $c'\mathrm{i}^k + d'[k(-\mathrm{i})\mathrm{i}^{k-1} + (n-k)\mathrm{i}^k] = ck\mathrm{i}^{k-1} + d\mathrm{i}^k$. Thus the contravariant transformation $(Z^{-1})^{\otimes n} f = f'$. Let $\mathcal{F}' = Z^{-1}\mathcal{F}$. Under the covariant holographic transformation by Z, $(=_2)$ is transformed to $Z^{\mathrm{T}} I Z = 2\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$, which is two times the binary DISEQUALITY $(\neq_2)$ function $[0, 1, 0]$ in matrix form. Holant$(\mathcal{F})$, which is the same as Holant$(=_2 | \mathcal{F})$, is equivalent to Holant$(\neq_2 | \mathcal{F}')$ by Theorem 1.5. The bipartite problem Holant$(\neq_2 | \mathcal{F}')$, where each $f' \in \mathcal{F}'$ has the form in (2.10), and may have different parameters c' and d', is tractable by the following argument. In any nonzero evaluation, on account of $(\neq_2)$, an assignment must assign exactly half the edges 0 and the other half 1. For any f' in (2.10) with arity greater than 2, to be nonzero it must have strictly more incident edges assigned 0. Any f' with arity 2 requires at least half of the input bits be set to 0. Therefore the Holant value is 0 unless every f' has arity 2. Hence the computation reduces to the case where each function has arity ≤ 2, and the Holant value is computed by matrix product.

What is the class of Holant problems that can be solved by Fibonacci gates, possibly after a suitable holographic transformation? We now characterize this class of problems in terms of signatures that can be transformed to a Fibonacci gate under a holographic transformation.

Theorem 2.5. *A symmetric signature $[f_0, f_1, \ldots, f_n]$ can be transformed by some invertible holographic transformation to a Fibonacci gate according to Definition 2.1 (equivalently to a signature in $\mathscr{F}$ defined in (2.9)) iff there exist three constants a, b and c, such that $b^2 - 4ac \neq 0$, and for all $0 \leq k \leq n-2$,*

$$af_k + bf_{k+1} + cf_{k+2} = 0. \tag{2.13}$$

https://doi.org/10.1017/9781107477063.003 Published online by Cambridge University Press

Proof. For an invertible $M = \begin{bmatrix} \alpha & \gamma \\ \beta & \delta \end{bmatrix}$, we have

$$M^{\otimes n}[u, 0, \ldots, 0, v] = u \begin{bmatrix} \alpha \\ \beta \end{bmatrix}^{\otimes n} + v \begin{bmatrix} \gamma \\ \delta \end{bmatrix}^{\otimes n}.$$

If f has this form, then f satisfies a recurrence relation $af_k + bf_{k+1} + cf_{k+2} = 0$ $(0 \le k \le n-2)$, where a, b, and c satisfy $b^2 - 4ac \neq 0$. In particular, they are not all 0. It can be directly verified that the following choice works: $a = \beta\delta$, $b = -(\alpha\delta + \beta\gamma)$ and $c = \alpha\gamma$.

Conversely, from $af_k + bf_{k+1} + cf_{k+2} = 0$, if $c \neq 0$ we have a second-order linear recurrence with distinct eigenvalues λ and μ, which gives us the expression

$$f = u \begin{bmatrix} 1 \\ \lambda \end{bmatrix}^{\otimes n} + v \begin{bmatrix} 1 \\ \mu \end{bmatrix}^{\otimes n} = M[u, 0, \ldots, 0, v],$$

where $M = \begin{bmatrix} 1 & 1 \\ \lambda & \mu \end{bmatrix}$. If $a \neq 0$, then we have a reverse linear recurrence. If $a = c = 0$, then $b \neq 0$, and we have a Gen-Eq. □

Theorem 2.6. Holant $(\{R_1, R_2, \ldots, R_t\} \mid \{G_1, G_2, \ldots, G_s\})$ *is tractable, where G_i and R_j have arities n_i and m_j respectively, if there exists an invertible M such that all $M^{-1}G_i$ and R_jM are Fibonacci gates. This happens iff there exist a, b, and c, such that $b^2 - 4ac \neq 0$ and the following two conditions are satisfied:*

1. *For any $G_i = [x_1^{(i)}, x_2^{(i)}, \ldots, x_{n_i}^{(i)}]$ and any $0 \le k \le n_i - 2$, $ax_k^{(i)} + bx_{k+1}^{(i)} + cx_{k+2}^{(i)} = 0$.*
2. *For any $R_j = [y_1^{(j)}, y_2^{(j)}, \ldots, y_{m_j}^{(j)}]$ and any $0 \le k \le m_j - 2$, $cy_k^{(j)} - by_{k+1}^{(j)} + ay_{k+2}^{(j)} = 0$.*

Proof. We may take the normal form Gen-Eq, and assume $(M^{-1})^{\otimes n_i}G_i = [*, 0, \ldots, 0, *]$ and $R_jM^{\otimes m_j} = [*, 0, \ldots, 0, *]$, for some $M = \begin{bmatrix} \alpha & \gamma \\ \beta & \delta \end{bmatrix}$. By the proof of Theorem 2.5, there exist $a = \beta\delta$, $b = -(\alpha\delta + \beta\gamma)$ and $c = \alpha\gamma$, such that $b^2 - 4ac \neq 0$, and for all $G_i = [x_1^{(i)}, x_2^{(i)}, \ldots, x_{n_i}^{(i)}]$, $ax_k^{(i)} + bx_{k+1}^{(i)} + cx_{k+2}^{(i)} = 0$ (for $0 \le k \le n_i - 2$). The characteristic polynomial is $cX^2 + bX + a$. If we write R_j in its transpose, then Theorem 2.5 applies to R_j, with $(M^{-1})^{\mathrm{T}} = \frac{1}{\det(M)} \begin{bmatrix} \delta & -\beta \\ -\gamma & \alpha \end{bmatrix}$ replacing M. It follows that each R_j satisfies $cy_k^{(j)} - by_{k+1}^{(j)} + ay_{k+2}^{(j)} = 0$ $(0 \le k \le m_j - 2)$.

https://doi.org/10.1017/9781107477063.003 Published online by Cambridge University Press

Conversely, suppose all G_i and R_j satisfy the linear recurrences. Suppose $c \neq 0$. Then for each G_i the characteristic polynomial of the recurrence is $X^2 + \frac{b}{c}X + \frac{a}{c}$. Let Δ denote the discriminant $b^2 - 4ac$. Let $M = \begin{bmatrix} 2c & 2c \\ -b+\sqrt{\Delta} & -b-\sqrt{\Delta} \end{bmatrix}$. Obviously this is nonsingular, and it can be verified that $G_i = a_i \begin{bmatrix} 2c \\ -b+\sqrt{\Delta} \end{bmatrix}^{\otimes n_i} + b_i \begin{bmatrix} 2c \\ -b-\sqrt{\Delta} \end{bmatrix}^{\otimes n_i}$, for some constants a_i and b_i depending on G_i. Similarly, each R_j satisfies $R_j = c_j \begin{bmatrix} b+\sqrt{\Delta} & 2c \end{bmatrix}^{\otimes m_j} + d_j \begin{bmatrix} b-\sqrt{\Delta} & 2c \end{bmatrix}^{\otimes m_j}$, for some c_j and d_j depending on R_j. Then both $(M^{-1})^{\otimes n_i} G_i$ and $R_j M^{\otimes m_j}$ are Gen-Eq. The proof is similar if $a \neq 0$. If $a = c = 0$, then $b \neq 0$, and both G_i and R_j are already Gen-Eq. □

2.3 A Dichotomy Theorem for Holant$^*(\mathcal{F})$

Every degenerate signature is a tensor product of unary signatures. If $\mathcal{F}$ consists of degenerate signatures only, each signature grid Ω on Holant$(\mathcal{F})$ is decomposed into a disjoint union of components, one for each vertex, and the Holant value is trivially computable. To isolate more interesting tractable families of signatures, we consider the following Holant* problems, where all unary functions $\mathcal{U}$ are assumed to be free.

Definition 2.7. For any signature set $\mathcal{F}$,

$$\text{Holant}^*(\mathcal{F}) = \text{Holant}(\mathcal{F} \cup \mathcal{U}).$$

For bipartite graphs we also denote

$$\text{Holant}^*(\mathcal{F} \mid \mathcal{G}) = \text{Holant}(\mathcal{F} \cup \mathcal{U} \mid \mathcal{G} \cup \mathcal{U}),$$

for any signature sets $\mathcal{F}$ and $\mathcal{G}$.

We begin with a technical proposition.

Proposition 2.8. *Let $n \geq 3$ and let $[x_0, x_1, \ldots, x_n]$ be a nondegenerate symmetric signature. Then for any $2 \leq m \leq n-1$, there exists a nondegenerate subsignature of arity m, unless the signature is of the form $[x_0, 0, \ldots, 0, x_n]$.*

Proof. We assume $(x_1, \ldots, x_{n-1}) \neq 0$, otherwise we are done. We first prove that

$$\text{rank} \begin{bmatrix} x_0 & \ldots & x_{n-2} \\ x_1 & \ldots & x_{n-1} \end{bmatrix} = 2, \quad \text{or} \quad \text{rank} \begin{bmatrix} x_1 & \ldots & x_{n-1} \\ x_2 & \ldots & x_n \end{bmatrix} = 2. \tag{2.14}$$

https://doi.org/10.1017/9781107477063.003 Published online by Cambridge University Press

Suppose otherwise. Then there are λ, μ such that $(x_0, \ldots, x_{n-2}) = \lambda(x_1, \ldots, x_{n-1})$ and $(x_2, \ldots, x_n) = \mu(x_1, \ldots, x_{n-1})$. If $\lambda = 0$ then $x_0 = \ldots = x_{n-2} = 0$ and $x_{n-1} \neq 0$, and $\det \begin{bmatrix} x_{n-2} & x_{n-1} \\ x_{n-1} & x_n \end{bmatrix} \neq 0$. Hence $\lambda \neq 0$. Similarly $\mu \neq 0$. Then any $x_i \neq 0$ $(0 < i < n)$ implies that $x_{i-1} \neq 0$ and $x_{i+1} \neq 0$. Hence $x_i \neq 0$ for all $0 \leq i \leq n$. This implies that $\lambda\mu = 1$, and thus $\operatorname{rank} \begin{bmatrix} x_0 & \ldots & x_{n-1} \\ x_1 & \ldots & x_n \end{bmatrix} = 1$, a contradiction to the hypothesis of the lemma. Hence (2.14) holds. Then an induction completes the proof. Note that if $[x_0, \ldots, x_{n-1}]$ is nondegenerate and has the form $[x_0, 0, \ldots, 0, x_{n-1}]$, then $x_{n-1} \neq 0$. Then for any $2 \leq m \leq n-1$, we have a nondegenerate subsignature $[0, \ldots, 0, x_{n-1}, x_n]$ of arity m. □

The next definition defines some explicit families of signatures. It turns out that they essentially capture the tractable classes of Holant* problems, as will be shown in Theorem 2.12.

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

For arity ≥ 2, type I is the union of generalized Fibonacci gates with Gen-Eq. For any nondegenerate signature of arity ≥ 3, if it is of type I(a, b), then (a, b) is unique up to a scalar multiplier; i.e., this $(a, b) \in \mathrm{P}\mathbb{C}^1$ is uniquely determined as a point on the projective line. Similarly, any nondegenerate signature of arity ≥ 3 cannot be both of type I and type II. These can be directly verified.

Proposition 2.10. *$[x_0, x_1, x_2, x_3]$ is of type I iff* $\det \begin{bmatrix} x_0 - x_2 & x_1 \\ x_1 - x_3 & x_2 \end{bmatrix} = 0.$

Let $M = \begin{bmatrix} \alpha & \gamma \\ \beta & \delta \end{bmatrix}$, *and define* $[x_0, x_1, x_2, x_3] = M^{\otimes 3}[c, 0, 0, d] = c \begin{bmatrix} \alpha \\ \beta \end{bmatrix}^{\otimes 3} + d \begin{bmatrix} \gamma \\ \delta \end{bmatrix}^{\otimes 3}$. *Then* $\det \begin{bmatrix} x_0 & x_1 \\ x_1 & x_2 \end{bmatrix} = cd\alpha\gamma(\det M)^2$, $\det \begin{bmatrix} x_1 & x_2 \\ x_2 & x_3 \end{bmatrix} = cd\beta\delta(\det M)^2$, *and* $\det \begin{bmatrix} x_0 - x_2 & x_1 \\ x_1 - x_3 & x_2 \end{bmatrix}$ *is their sum.*

Suppose $[x_0, x_1, x_2, x_3] = M^{\otimes 3}[c, 0, 0, d]$ *is nondegenerate. Then,* $[x_0, x_1, x_2, x_3]$ *is of type I iff the two columns of M are orthogonal* $\alpha\gamma + \beta\delta = 0$. *In this case, it belongs to* I(a, b), *where* $a = \alpha\gamma = -\beta\delta$ *and* $b = \alpha\delta + \beta\gamma$.

https://doi.org/10.1017/9781107477063.003 Published online by Cambridge University Press

Proof. A straightforward calculation. Note that $[x_0, x_1, x_2]$ is a binary signature having the matrix form $M\begin{bmatrix} c\alpha & 0 \\ 0 & d\gamma \end{bmatrix}M^{\mathrm{T}}$. Similarly, $[x_1, x_2, x_3]$ has the form $M\begin{bmatrix} c\beta & 0 \\ 0 & d\delta \end{bmatrix}M^{\mathrm{T}}$. □

Proposition 2.11. *Let* $[x_0, x_1, x_2, x_3] = M^{\otimes 3}[c, 0, 0, d]$ *be nondegenerate. Then,* $[x_0, x_1, x_2, x_3]$ *is of type II iff the columns of* M *are multiples of* $(1, \mathrm{i})^{\mathrm{T}}$ *and* $(1, -\mathrm{i})^{\mathrm{T}}$ *respectively.*

Proof. By nondegeneracy, $cd \neq 0$ and $\det M \neq 0$. The equations $x_0 + x_2 = 0$ and $x_1 + x_3 = 0$ are respectively $c\alpha(\alpha^2 + \beta^2) + d\gamma(\gamma^2 + \delta^2) = 0$ and $c\beta(\alpha^2 + \beta^2) + d\delta(\gamma^2 + \delta^2) = 0$. Viewed as a linear equation system on c and d, there are nonzero solutions iff its determinant is zero:

$$(\alpha^2 + \beta^2)(\gamma^2 + \delta^2)\det M = 0.$$

It follows that at least one of the factors $\alpha^2 + \beta^2 = 0$ or $\gamma^2 + \delta^2 = 0$. If $\alpha^2 + \beta^2 = 0$, by the linear equation system, since (γ, δ) is a nonzero vector, we get the second factor, $\gamma^2 + \delta^2 = 0$, as well. Similarly, starting with $\gamma^2 + \delta^2 = 0$ we also get $\alpha^2 + \beta^2 = 0$.

Conversely, if $\beta = \pm\mathrm{i}\alpha$ and $\delta = \mp\mathrm{i}\gamma$, then $c\begin{bmatrix} \alpha \\ \beta \end{bmatrix}^{\otimes 3} + d\begin{bmatrix} \gamma \\ \delta \end{bmatrix}^{\otimes 3}$ has the form $[x, y, -x, -y]$. □

The essence of Proposition 2.11 is that $c\begin{bmatrix} \alpha \\ \beta \end{bmatrix}^{\otimes 3} + d\begin{bmatrix} \gamma \\ \delta \end{bmatrix}^{\otimes 3}$ satisfies the second-order linear recurrence with eigenvalues β/α and δ/γ, and it is of type II iff these eigenvalues are $\pm\mathrm{i}$.

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

Since all unary signatures can be used for free in Holant$^*(\mathcal{F})$, we assume the arity of every signature in $\mathcal{F}$ is greater than 1. And since any degenerate signature can be decomposed to unary signatures, we also assume that every signature in $\mathcal{F}$ is nondegenerate.

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

$\mathfrak{A}$. *Every signature in* $\mathcal{F}$ *is of arity* ≤ 2;

https://doi.org/10.1017/9781107477063.003 Published online by Cambridge University Press

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

$\mathfrak{C}$. *Every signature in $\mathcal{F}$ either (1) has type II or (2) has arity 2 and is of the form* $[\lambda, 0, \lambda]$.

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

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

Remark: For a more conceptual characterization, the reader is encouraged to compare this dichotomy theorem with Theorem 7.19. To formulate and prove Theorem 7.19, however, we need to develop the necessary machinery and preparatory results, of which Theorem 2.12 is the first one. The reader is also encouraged to compare Theorem 2.12 with Theorem 3.5, which is a more abstract statement of Theorem 2.12.

This dichotomy theorem essentially says that for Holant$^*(\mathcal{F})$ the only tractable symmetric signature sets $\mathcal{F}$ are those when, after all degenerate signatures are removed, $\mathcal{F}$ consists of functions of arity ≤ 2, or consists of generalized Fibonacci gates under a holographic transformation and some specific binary functions, or one additional class.

Proof of Tractability: Class $\mathfrak{A}$ is when every signature in $\mathcal{F}$ has arity ≤ 2. Then the graph of the signature grid Ω is a disjoint union of paths and cycles (isolated points contribute a constant; we may assume there are no isolated points). By matrix multiplication, we can compute the Holant value for a path. The Holant value for a cycle is obtained by taking the trace of a path. The value Holant_Ω is the product over connected components.

Next we consider class $\mathfrak{B}$. If $a = 0$, all functions in $\mathcal{F}$ are Gen-Eq or binary Disequality, and Holant$^*(\mathcal{F})$ is obviously computable in polynomial time. Suppose $a \neq 0$. All functions in $\mathcal{F}$ having arity > 2 are generalized Fibonacci gates. We first assume $X^2 - (b/a)X - 1$ has two distinct eigenvalues two distinct eigenvalues λ and μ, where $\lambda \neq \mu$. Then $\lambda\mu = -1$ and $\lambda + \mu = b/a$.

https://doi.org/10.1017/9781107477063.003 Published online by Cambridge University Press

All functions in $\mathcal{F}$ have the form $f = u\begin{bmatrix}1\\ \lambda\end{bmatrix}^{\otimes n} + v\begin{bmatrix}1\\ \mu\end{bmatrix}^{\otimes n}$ or multiples of $g = [2, b/a, -2]$. Under the (inverse) holographic transformation by $M = \begin{bmatrix}1 & \lambda\\ 1 & \mu\end{bmatrix}$, $f \mapsto M^{\otimes n} f$ is a Gen-Eq, $g \mapsto M^{\otimes 2} g = [0, 4 + (b/a)^2, 0]$, and $(=_2) = [1, 0, 1]$ is turned into $[1, 0, 1](M^{-1})^{\otimes 2}$, which is a multiple of $[1 + \mu^2, 0, 1 + \lambda^2]$, a Gen-Eq. In the meantime, all unary functions are transformed into unary functions. Connecting any unary function to a Gen-Eq creates another Gen-Eq. Connecting any unary function to any function of arity 2 creates another unary function. Hence Holant$^*(\mathcal{F})$ is tractable.

Continuing for class $\mathfrak{B}$, if the eigenvalues $\lambda = \mu$, then $b = \pm 2a\mathrm{i}$. Then $a \neq 0$ and we can normalize to $a = 1$, $b = \pm 2\mathrm{i}$, $\lambda = b/2$. The signature $g = [2, b, -2] = [2, \pm 2\mathrm{i}, -2]$ is degenerate. We have proved using the holographic transformation Z that in this case Holant$^*(\mathcal{F})$ is tractable.

For class $\mathfrak{C}$, we first get rid of all unary functions and binary functions of the form $\lambda[1, 0, 1]$. If there is a function $\lambda[1, 0, 1]$ applied to two variables (edges) e and e', we just merge them into one, and remove this function with a global factor λ. Combining a unary function with a function of the type $[x, y, -x, -y, \ldots]$ does not change its type. Hence, we may assume all functions satisfy $x_k + x_{k+2} = 0$. If the input graph $G(V, E)$ is not bipartite, there is a cycle $v_1, e_1, v_2, \ldots, v_k, e_k, v_1$ of odd length. We partition all 0-1 assignments for E into two parts, with a 1-1 correspondence between them. An assignment σ is mapped to σ', which assigns the same values on $E - \{e_1, \ldots, e_k\}$, but opposite values on $\{e_1, \ldots, e_k\}$. Under σ and σ', all functions on $V - \{v_1, \ldots, v_k\}$ evaluate to the same value, and if $\sigma(e_{j-1}) \neq \sigma(e_j)$ (where the index j is counted mod k), the function at v_j evaluates to the same value under σ and σ', and if $\sigma(e_{j-1}) = \sigma(e_j)$, the function at v_j evaluates to opposite values under σ and σ'. Consider $\sigma(e_1), \sigma(e_2), \ldots, \sigma(e_k), \sigma(e_1)$. There must be an even number of times where the value changes in this sequence. Since k is odd, there are an odd number of v_j's, where the functions at v_j give opposite values under σ and σ'. Hence, in the summation, the contributions of σ and σ' are canceled.

If the input graph is bipartite, the problem is Holant$(\mathcal{F} \mid \mathcal{F})$. It is turned into Holant$(\mathcal{F}' \mid \mathcal{F}')$, where $\mathcal{F}' \subseteq$ Gen-Eq, by the holographic reduction $Z = \begin{bmatrix}1 & 1\\ \mathrm{i} & -\mathrm{i}\end{bmatrix}$. This is easily seen by writing each contravariant $f \in \mathcal{F}$ as $u\begin{bmatrix}1\\ \mathrm{i}\end{bmatrix}^{\otimes n} + v\begin{bmatrix}1\\ -\mathrm{i}\end{bmatrix}^{\otimes n} = Z[u, 0, \ldots, 0, v]$, and each covariant $g \in \mathcal{F}$ as $u'\begin{bmatrix}1 & \mathrm{i}\end{bmatrix}^{\otimes m} + v'\begin{bmatrix}1 & -\mathrm{i}\end{bmatrix}^{\otimes m}$, so that $gZ = u'\begin{bmatrix}0 & 2\end{bmatrix}^{\otimes m} + v'\begin{bmatrix}2 & 0\end{bmatrix}^{\otimes m} = [2^m v', 0, \ldots, 0, 2^m u']$. Hence Holant$(\mathcal{F} \mid \mathcal{F})$ is tractable.

https://doi.org/10.1017/9781107477063.003 Published online by Cambridge University Press

Proof Outline for Hardness: Now for the proof of hardness, we first prove in Lemma 2.13 that the theorem holds if $\mathcal{F}$ contains a single symmetric signature of arity 3. The main technique is holographic reduction. In Lemma 2.14, we prove that if one signature of arity 3 has the form in class $\mathfrak{B}$ of Theorem 2.12, and we combine it with another signature of arity 2 that is *not* in class $\mathfrak{B}$, then the Holant* problem is #P-hard. The idea of the proof of Lemma 2.14 is to reduce it to Lemma 2.13 with holographic reductions. In Lemma 2.15, we prove the same thing for class $\mathfrak{C}$. In Lemma 2.16 we extend Lemma 2.13 to a single signature of arbitrary arity. Finally, we extend the proof to a set of signatures of arbitrary arities.

2.3.1 The First Lemma: A Single Ternary Signature

The following lemma says that Theorem 2.12 holds if $\mathcal{F}$ consists of only one signature of arity 3. It serves as the foundation for subsequent lemmas.

Lemma 2.13. *Let $f = [x_0, x_1, x_2, x_3]$ be a nondegenerate symmetric signature with arity three, then* Holant*(f) *is #P-hard unless f has type I or type II.*

Proof. We assume f is of neither type I nor type II, and prove that Holant*(f) is #P-hard. Our starting point is that Holant([0, 1, 1]|[1, 0, 0, 1]) and Holant([1, 0, 1]|[1, 1, 0, 0]) are both #P-complete [XZZ07]. The first problem is to count the number of vertex covers for 3-regular graphs; the second is to count the number of (not necessarily perfect) matchings for 3-regular graphs. Both problems remain #P-complete even for planar graphs [DL92, XZZ07].

Every nondegenerate signature $[x_0, x_1, x_2, x_3]$ belongs to one of the following categories: For $0 \le k \le 3$,

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

This parametrization is obtained by considering the solutions $(a, b, c)^{\mathrm{T}}$ to $\begin{bmatrix} x_0 & x_1 & x_2 \\ x_1 & x_2 & x_3 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = 0$, which form a one-dimensional vector space. Suppose $(a, b, c)^{\mathrm{T}}$ is a nonzero solution. If $a = c = 0$, we have $x_1 = x_2 = 0$ and a diagonal matrix $\begin{bmatrix} \sqrt[3]{x_0} & 0 \\ 0 & \sqrt[3]{x_3} \end{bmatrix}$ can be used in category 1. If a and c are not both 0, then we may consider $[x_0, x_1, x_2, x_3]$ satisfies a second-order linear recurrence relation (either forward or backward). Depending on whether the characteristic

https://doi.org/10.1017/9781107477063.003 Published online by Cambridge University Press

equation has two distinct roots or a double root, we have a case in category 1 or the other two. Category 3 can be viewed as the reversal of category 2, so we will omit the proof for category 3. We take the convention that $k0^{k-1} = 0, 1, 0, 0$ for $k = 0, 1, 2, 3$ respectively.

For category 1, we have

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

Nondegeneracy of f implies that $\alpha_1\beta_2 - \alpha_2\beta_1 \neq 0$. The fact that f is not of type I or II implies that $\alpha_1\beta_1 + \alpha_2\beta_2 \neq 0$ by Proposition 2.10, and $\alpha_1^2 + \alpha_2^2 \neq 0$ or $\beta_1^2 + \beta_2^2 \neq 0$ by Proposition 2.11. By symmetry, we can assume that $\alpha_1^2 + \alpha_2^2 \neq 0$.

We apply an orthogonal transformation $T = \frac{1}{\sqrt{\alpha_1^2+\alpha_2^2}}\begin{bmatrix}\alpha_1 & \alpha_2\\ -\alpha_2 & \alpha_1\end{bmatrix}$ to map the vector $(\alpha_1, \alpha_2)^{\mathrm{T}}$ to $(\alpha_1', 0)^{\mathrm{T}}$, where $\alpha_1' = \sqrt{\alpha_1^2 + \alpha_2^2} \neq 0$.

$$f' = T^{\otimes 3} f = [x_0', x_1', x_2', x_3'] = \begin{bmatrix}\alpha_1'\\ 0\end{bmatrix}^{\otimes 3} + \begin{bmatrix}\beta_1'\\ \beta_2'\end{bmatrix}^{\otimes 3}.$$

By Theorem 1.6, this transformation does not change the complexity of the Holant problem. So it suffices to prove #P-hardness for f'. By a scalar multiplication we may assume $\alpha_1' = 1$. So, reusing the notation f, we can assume the given signature is

$$f = [x_0, x_1, x_2, x_3] = \begin{bmatrix}1\\ 0\end{bmatrix}^{\otimes 3} + \begin{bmatrix}\alpha\\ \beta\end{bmatrix}^{\otimes 3}. \tag{2.15}$$

The conditions from Lemma 2.13 become simply both $\alpha, \beta \neq 0$.

Consider $M = \begin{bmatrix}1 & \alpha\\ 0 & \beta\end{bmatrix}$. We have

$$f = M^{\otimes 3}[1, 0, 0, 1] = M^{\otimes 3}\left(\begin{bmatrix}1\\ 0\end{bmatrix}^{\otimes 3} + \begin{bmatrix}0\\ 1\end{bmatrix}^{\otimes 3}\right),$$

and

$$(0, 1, 1, 1)(M^{-1})^{\otimes 2} = \frac{1}{\beta^2}(0, \beta, \beta, 1 - 2\alpha).$$

One can check this by computing $M^{\mathrm{T}}\begin{bmatrix}0 & \beta\\ \beta & 1-2\alpha\end{bmatrix}M$. Ignoring $1/\beta^2$, by a holographic reduction the complexity of Holant($[0, \beta, 1 - 2\alpha] \mid f$) is the same as the #P-complete vertex cover problem Holant($[0, 1, 1] \mid [1, 0, 0, 1]$). In particular, Holant($[0, \beta, 1 - 2\alpha], f$) is #P-hard. In order to prove the #P-hardness

https://doi.org/10.1017/9781107477063.003 Published online by Cambridge University Press

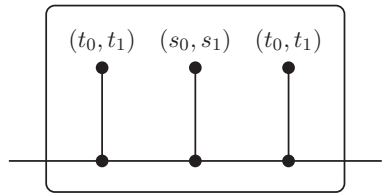

Figure 2.3. We use this gadget to realize the signature $[0, \beta, 1-2\alpha]$. All (three) nodes of degree 3 in this gadget have the signature $f = [x_0, x_1, x_2, x_3]$.

of Holant$^*(f)$, we only need to show that $[0, \beta, 1-2\alpha]$ can be realized by f with some unary signatures.

We use the gadget in Figure 2.3 to realize $[0, \beta, 1-2\alpha]$, where the two unary signatures (t_0, t_1) and (s_0, s_1) will be determined later. Let

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

For f in (2.15), if one input is 0, the induced binary signature has its matrix form $A + \alpha B$. If one input is 1, the induced binary signature has matrix form βB. It follows that the signature in Figure 2.3, as a binary function in matrix form, is

$$\begin{aligned} &(t_0(A + \alpha B) + t_1\beta B)(s_0(A + \alpha B) + s_1\beta B)(t_0(A + \alpha B) + t_1\beta B) \\ &\quad = (t_0 A + (t_0\alpha + t_1\beta)B)(s_0 A + (s_0\alpha + s_1\beta)B)(t_0 A + (t_0\alpha + t_1\beta)B). \end{aligned}$$

Exercise: Show that $t_0(A + \alpha B) + t_1\beta B$ is the signature matrix of the binary gadget obtained by connecting the unary signature $[t_0, t_1]$ to f of (2.15). Deduce that the above matrix is the signature matrix of the binary gadget in Figure 2.3.

Now we use a new set of variables:

$$x = t_0, \quad y = t_0\alpha + t_1\beta, \quad z = s_0, \quad w = s_0\alpha + s_1\beta, \tag{2.16}$$

and write the above matrix as $(xA + yB)(zA + wB)(xA + yB)$. We note that, since $\beta \neq 0$, for any given x, y, z, w, we can find t_0, t_1, s_0, s_1 to satisfy the relations (2.16). Then, to realize $[0, \beta, 1-2\alpha]$, we will choose some x, y, z, and w such that

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

https://doi.org/10.1017/9781107477063.003 Published online by Cambridge University Press

Substituting A and B, and denoting $\alpha^2 + \beta^2$ by γ, we have the following:

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

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

$$g_1 = xy\alpha\beta^3(x + y\alpha^2)(x + y\gamma);$$

and the $(2, 2)$ entry is

$$g_2 = xy^2\beta^4(x(2\alpha^2 + \beta^2) + 2y\alpha^2\gamma).$$

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

$$\alpha(2\alpha - 1)x^2 + (2\alpha^2 - \alpha + \beta^2)(2\alpha^2 + \beta^2)x + \alpha^2(\alpha^2 + \beta^2)(2\alpha^2 - \alpha + 2\beta^2) = 0 \tag{2.17}$$

is equivalent to $\beta g_2 = (1 - 2\alpha)g_1$ after removing $xy\beta^3$ and setting $y = 1$. What we have to prove is that (2.17) has a root $x \neq 0$ that is not a root of $g_1 = g_1(x, 1) = 0$. The roots of $g_1(x, 1) = 0$ are $x = 0, x = -\alpha^2$ and $x = -\gamma$. First, when $x = -\alpha^2$, the left-hand side expression in (2.17) can be simplified to $\alpha^2\beta^4 \neq 0$. (Note that when $g_1 = 0$, the left-hand side expression in (2.17) is βg_2, after removing $xy\beta^3$, and $y = 1$.) So $x = -\alpha^2$ is not a root of (2.17). Second, when $x = -\gamma$ the left-hand side expression in (2.17) can be simplified to $-\beta^4\gamma$, and if $x = -\gamma$ is a root, this would force $\gamma = 0$. So, assuming the left-hand side expression in (2.17) as a polynomial in x is indeed of degree 2, and it does not have a double root 0, then we can find a root $\xi \neq 0$ of (2.17) that is not a root of g_1. Indeed, This $\xi \neq -\alpha^2$, because $-\alpha^2$ is not a root of (2.17); ξ cannot be $-\gamma$ either, for otherwise $-\gamma$ would be a root of (2.17), which would force $\gamma = 0$, and thus $\xi = -\gamma = 0$, a contradiction. Thus ξ is a nonzero root of (2.17) but not a root of $g_1(x, 1)$, as is needed.

https://doi.org/10.1017/9781107477063.003 Published online by Cambridge University Press

Now let us consider the exceptional cases: either $x = 0$ is a double root of (2.17), or (2.17) has degree < 2. If $x = 0$ is a double root of (2.17), we have

$$(2\alpha^2 - \alpha + \beta^2)(2\alpha^2 + \beta^2) = \alpha^2(\alpha^2 + \beta^2)(2\alpha^2 - \alpha + 2\beta^2) = 0.$$

To satisfy this, since $\alpha\beta \neq 0$, there are only four exceptional cases (A1 to A4): $\alpha = 1, \beta = \pm\mathrm{i}$ or $\alpha = -\frac{1}{2}, \beta = \pm\frac{\mathrm{i}}{\sqrt{2}}$. On the other hand, if the polynomial in (2.17) has degree < 2, by $\alpha \neq 0$, we get $\alpha = \frac{1}{2}$. In this case, the polynomial becomes

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

This gives us four additional exceptional cases (B1 to B4): $\alpha = \frac{1}{2}, \beta = \pm\frac{\mathrm{i}}{2}$, in which case the polynomial is linear with root $x = 0$; or $\alpha = \frac{1}{2}, \beta = \pm\frac{\mathrm{i}}{\sqrt{2}}$, in which case the polynomial degenerates to a (nonzero) constant. In all other cases, there is a nonzero root of (2.17) that is not a root of $g_1(x, 1)$, completing the #P-hardness proof.

For the cases A1 and A2, we use a new starting problem Holant([1, 1, 0] | [1, 0, 0, 1]), which is the reversal of the previous problem and therefore it is also #P-complete. The problem can also be understood as counting independent sets over 3-regular graphs. Then all previous parts of the proof are still valid, except that the signature of arity 2 to be realized is

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

Substituting $\alpha = 1, \beta = \pm\mathrm{i}$, the signature is $[1, 0, 1]$ which is trivially realizable by one edge. So we have proved that it is #P-hard in the cases A1 and A2. Now consider the cases A3 and A4: $\alpha = -\frac{1}{2}, \beta = \pm\frac{\mathrm{i}}{\sqrt{2}}$. We will give a different parametrization. For the case A3, we apply an orthogonal transformation $M = \begin{bmatrix} -\mathrm{i} & -\sqrt{2} \\ \sqrt{2} & -\mathrm{i} \end{bmatrix}$ and a scalar multiplier $(2\mathrm{i})^3$ on the signature f in (2.15), to get $f \mapsto (2\mathrm{i})^3 M^{\otimes 3} f = \begin{bmatrix} 1 \\ 0 \end{bmatrix}^{\otimes 3} + \begin{bmatrix} 2 \\ 2\sqrt{2}\mathrm{i} \end{bmatrix}^{\otimes 3}$. This is not one of the exceptional cases and we have proved that it is #P-hard. For the case A4, we apply another orthogonal transformation $M' = \begin{bmatrix} \mathrm{i} & -\sqrt{2} \\ \sqrt{2} & \mathrm{i} \end{bmatrix}$ and a scalar multiplier $(-2\mathrm{i})^3$ on the signature and it becomes $\begin{bmatrix} 1 \\ 0 \end{bmatrix}^{\otimes 3} + \begin{bmatrix} 2 \\ -2\sqrt{2}\mathrm{i} \end{bmatrix}^{\otimes 3}$.

The cases B3 and B4 can be shown by the same method as in A4 and A3, using M' and M respectively. The only remaining cases are B1 and B2. Here we will use another gadget similar to the one in Figure 2.3 except we remove the

https://doi.org/10.1017/9781107477063.003 Published online by Cambridge University Press

middle edge (including the node labeled (s_0, s_1) and the middle node of degree 3). For B1, $\alpha = \frac{1}{2}$, $\beta = \frac{\mathrm{i}}{2}$, the signature of this gadget is

$$(t_0(A + \alpha B) + t_1\beta B)^2 = (xA + yB)^2,$$

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

Exercise: Prove cases B3 and B4 using the matrices M' and M in the cases A4 and A3, respectively.

Now we prove for category 2. In this case $x_k = ck\alpha^{k-1} + d\alpha^k$, and it satisfies the recurrence $x_{k+2} = 2\alpha x_{k+1} - \alpha^2 x_k$, for $k = 0, 1$. The characteristic polynomial $X^2 - 2\alpha X + \alpha^2$ has a double root α. If $\alpha = \pm\mathrm{i}$, then f has type I$(1, 2\alpha)$. Since we assumed f is not of type I, $\alpha \neq \pm\mathrm{i}$. Then we can choose some orthogonal matrix to transform the signature to the form $[x, y, 0, 0]$ where $y \neq 0$, as follows.

Let $M = \begin{bmatrix} 1 & \frac{d-1}{3} \\ \alpha & c + \frac{d-1}{3}\alpha \end{bmatrix}$ with $\det M = c \neq 0$, then the signature $[x_0, x_1, x_2, x_3]$ can be expressed as

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

In symmetric signature notation $[x_0, x_1, x_2, x_3] = M^{\otimes 3}[1, 1, 0, 0]$.

Exercise: Verify that $[x_0, x_1, x_2, x_3] = M^{\otimes 3}[1, 1, 0, 0]$. Use M to show that there is another matrix M' such that $[x_0, x_1, x_2, x_3] = M'^{\otimes 3}[0, 1, 0, 0]$.

Let $M = QR$ be its QR-factorization, i.e., where Q is orthogonal and R is upper triangular. In fact, for $M = \begin{bmatrix} 1 & * \\ \alpha & * \end{bmatrix}$, recalling that $1 + \alpha^2 \neq 0$, we can choose our Q as the (complex) orthogonal matrix $Q = Q^{\mathrm{T}} = \frac{1}{\sqrt{1+\alpha^2}}\begin{bmatrix} 1 & \alpha \\ \alpha & -1 \end{bmatrix}$. Then $QM = R = \begin{bmatrix} u & w \\ 0 & v \end{bmatrix}$ is upper triangular, where $u = \sqrt{1 + \alpha^2}$. Because $\det Q = -1$ and $\det R = -\det M = -c \neq 0$, we have $uv \neq 0$. This Q is our

https://doi.org/10.1017/9781107477063.003 Published online by Cambridge University Press

choice of the orthogonal transformation. It follows that

$$
\begin{aligned}
&Q^{\otimes 3}[x_0, x_1, x_2, x_3] \\
&= (QM)^{\otimes 3}[1, 1, 0, 0] \\
&= R^{\otimes 3}\left\{\begin{bmatrix}1\\0\end{bmatrix}^{\otimes 3} + \begin{bmatrix}1\\0\end{bmatrix}\otimes\begin{bmatrix}1\\0\end{bmatrix}\otimes\begin{bmatrix}0\\1\end{bmatrix} + \begin{bmatrix}1\\0\end{bmatrix}\otimes\begin{bmatrix}0\\1\end{bmatrix}\otimes\begin{bmatrix}1\\0\end{bmatrix} + \begin{bmatrix}0\\1\end{bmatrix}\otimes\begin{bmatrix}1\\0\end{bmatrix}\otimes\begin{bmatrix}1\\0\end{bmatrix}\right\} \\
&= \begin{bmatrix}u\\0\end{bmatrix}^{\otimes 3} + \begin{bmatrix}u\\0\end{bmatrix}\otimes\begin{bmatrix}u\\0\end{bmatrix}\otimes\begin{bmatrix}w\\v\end{bmatrix} + \begin{bmatrix}u\\0\end{bmatrix}\otimes\begin{bmatrix}w\\v\end{bmatrix}\otimes\begin{bmatrix}u\\0\end{bmatrix} + \begin{bmatrix}w\\v\end{bmatrix}\otimes\begin{bmatrix}u\\0\end{bmatrix}\otimes\begin{bmatrix}u\\0\end{bmatrix}.
\end{aligned}
\tag{2.18}
$$

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

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

For this signature, we can apply a holographic transformation defined by the matrix $M' = \begin{bmatrix}1 & -\frac{v-1}{3}\\ 0 & 1\end{bmatrix}$ with inverse $M'^{-1} = \begin{bmatrix}1 & \frac{v-1}{3}\\ 0 & 1\end{bmatrix}$. To prove #P-hardness, we will reduce from the MATCHING problem Holant($[1, 0, 1] \mid [1, 1, 0, 0]$). Under a contravariant transformation $(M'^{-1})^{\otimes 3}[1, 1, 0, 0] = [v, 1, 0, 0]$, the signature $[1, 1, 0, 0]$ becomes $[v, 1, 0, 0]$. By the corresponding covariant transformation, $[1, 0, 1]$ becomes $[1, \frac{1-v}{3}, 1 + \frac{(1-v)^2}{9}]$:

$$
\begin{aligned}
(1, 0, 0, 1)M'^{\otimes 2} &= ((1, 0)^{\otimes 2} + (0, 1)^{\otimes 2})M'^{\otimes 2} \\
&= \left(1, \frac{1-v}{3}, \frac{1-v}{3}, 1 + \frac{(1-v)^2}{9}\right).
\end{aligned}
$$

We complete this proof by using the same gadget in Figure 2.3 to realize this binary signature, using unary signatures and $[v, 1, 0, 0]$.

We will rename the values $x = t_0$, $y = t_1$, $z = s_0$ and $w = s_1$ in Figure 2.3. The signature of this gadget in matrix form is $(xA + yB)(zA + wB)(xA + yB)$, where $A = \begin{bmatrix}v & 1\\ 1 & 0\end{bmatrix}$ and $B = \begin{bmatrix}1 & 0\\ 0 & 0\end{bmatrix}$.

We set $x = 1$. Then $(A + yB)(zA + wB)(A + yB)$ is

$$
\begin{bmatrix} w \cdot (v+y)^2 + z \cdot (v(v+y)^2 + 2(v+y) & w \cdot (v+y) + z \cdot (v(v+y)+1) \\ w \cdot (v+y) + z \cdot (v(v+y)+1) & w + z \cdot v \end{bmatrix}.
$$

Our goal is to choose y, z and w such that it is equal to $\begin{bmatrix}1 & (1-v)/3\\ (1-v)/3 & 1 + (1-v)^2/9\end{bmatrix}$. We can write this requirement as a system of three linear equations in z and w, whose coefficient matrix depends on y.

https://doi.org/10.1017/9781107477063.003 Published online by Cambridge University Press

Then we can complete the proof, if we can choose y such that the following matrix has determinant 0, yet the first two columns have rank 2.

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

After some elementary column operations on the first two columns it becomes

$$\begin{bmatrix} -(v+y)^2 & 2(v+y) & 1 \\ 0 & 1 & (1-v)/3 \\ 1 & 0 & 1+(1-v)^2/9 \end{bmatrix},$$

whose first two columns obviously have rank 2. Its determinant is

$$-(v+y)^2(1+(1-v)^2/9)+2(v+y)(1-v)/3-1.$$

Clearly $1+(1-v)^2/9$ and $(1-v)/3$ cannot be both 0. Thus we can set y so that this determinant is 0. This completes the proof of Lemma 2.13. □

2.3.2 A Pair of Signatures

Lemma 2.13 shows what happens when there is a single nondegenerate symmetric signature f of arity 3. It explicitly isolates two tractable cases: If Holant*(f) is *not* #P-hard, then f must be of type I or type II. The next lemma addresses what happens when we have more than one signature, where one signature happens to be of type I but some other signature does not *go along* with it.

Lemma 2.14. *Let $f=[x_0,x_1,x_2,x_3]$ and $g=[y_0,y_1,y_2]$ be nondegenerate symmetric signatures with arity 3 and 2 respectively. Suppose f is of type $I(a,b)$ for some a,b (not both zero), but g is not of type $I(a,b)$, i.e., $ay_0+by_1-ay_2\neq 0$, and g is not of the form $[2a\lambda,b\lambda,-2a\lambda]$. Then* Holant*($g \mid f$) *is #P-hard. It follows that* Holant*($\{f,g\}$) *is also #P-hard.*

Proof. Since f is nondegenerate, the pair (a,b) is unique up to a scalar factor. Our proof plan is as follows. We will show that Holant*($g \mid f$) is #P-hard by a holographic reduction where g is transformed to the binary EQUALITY ($=_2$). Hence we want an invertible matrix M such that $f \mapsto M^{\otimes 3}f = f'$ and $g \mapsto g(M^{-1})^{\otimes 2} = [1,0,1]$. Note that $[1,0,1]$ can be replaced by an edge, and the unary signatures are transformed to unary signatures. Thus the complexity of the problem Holant*($g \mid f$) is the same as Holant*(f'). We then apply Lemma 2.13 to f'. Note that f' is nondegenerate. Otherwise, f' would be a

https://doi.org/10.1017/9781107477063.003 Published online by Cambridge University Press

tensor product of unary signatures, and then $f = (M^{-1})^{\otimes 3} f'$ would also be a tensor product of unary signatures, and thus degenerate.

The requirement that $g(M^{-1})^{\otimes 2} = [1, 0, 1]$ is the same as $G = M^{\mathrm{T}} M$ for some matrix $M = \begin{bmatrix} \alpha & \gamma \\ \beta & \delta \end{bmatrix}$, where $G = \begin{bmatrix} y_0 & y_1 \\ y_1 & y_2 \end{bmatrix}$ is the matrix form of g. This is

$$\begin{bmatrix} y_0 & y_1 \\ y_1 & y_2 \end{bmatrix} = M^{\mathrm{T}} M = \begin{bmatrix} \alpha^2 + \beta^2 & \alpha\gamma + \beta\delta \\ \alpha\gamma + \beta\delta & \gamma^2 + \delta^2 \end{bmatrix}. \tag{2.19}$$

Such a factorization for a symmetric complex matrix G exists. If $y_0 \neq 0$, we take $M = \frac{1}{\sqrt{y_0}} \begin{bmatrix} y_0 & y_1 \\ 0 & \Delta \end{bmatrix}$, where $\Delta = \sqrt{\det G}$. Note that $\det G \neq 0$ since g is nondegenerate. If $y_0 = 0$ but $y_2 \neq 0$, we can first exchange y_0 and y_2 by $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$, and take $M = \frac{1}{\sqrt{y_2}} \begin{bmatrix} \Delta & 0 \\ y_1 & y_2 \end{bmatrix}$. If $y_0 = y_2 = 0$, then we take $M = \sqrt{y_1} Z$ where $Z = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ \mathrm{i} & -\mathrm{i} \end{bmatrix}$.

Let $f' = [u_0, u_1, u_2, u_3] = M^{\otimes 3} f$. We want to show that f' is not of type I or II, the two tractable cases of Lemma 2.13.

First we assume for a contradiction that f' is of type II. Then it satisfies the recurrence $u_{k+2} + u_k = 0$, for $k = 0, 1$, with characteristic polynomial $X^2 + 1$. Therefore $f' = c \begin{bmatrix} 1 \\ \mathrm{i} \end{bmatrix}^{\otimes 3} + d \begin{bmatrix} 1 \\ -\mathrm{i} \end{bmatrix}^{\otimes 3}$, for some c and d. We have $M^{-1} = \Delta^{-1} \begin{bmatrix} \delta & -\gamma \\ -\beta & \alpha \end{bmatrix}$, where $\Delta = \det M \neq 0$. Then it follows that

$$f = (M^{-1})^{\otimes 3} f' = c\Delta^{-3} \begin{bmatrix} \delta - \gamma\mathrm{i} \\ -\beta + \alpha\mathrm{i} \end{bmatrix}^{\otimes 3} + d\Delta^{-3} \begin{bmatrix} \delta + \gamma\mathrm{i} \\ -\beta - \alpha\mathrm{i} \end{bmatrix}^{\otimes 3}.$$

Since f is nondegenerate, $cd \neq 0$. By the assumption that f has type I(a, b), and by Proposition 2.10, we have

$$\alpha^2 + \beta^2 + \gamma^2 + \delta^2 = 0,$$

and $a = \gamma^2 + \delta^2 = -\alpha^2 - \beta^2$, and $b = -2(\alpha\gamma + \beta\delta)$ after a scaling. By (2.19), $[y_0, y_1, y_2] = [-a, -b/2, a]$ is of the form $[2a\lambda, b\lambda, -2a\lambda]$, a contradiction.

This completes our first step. Our second step is to show that f' cannot be of type I either. For a contradiction, suppose f' is of type I.

https://doi.org/10.1017/9781107477063.003 Published online by Cambridge University Press

We are given that f is of type I(a, b). If $a = 0$ then $b \neq 0$ and $x_1 = x_2 = 0$. We can write

$$f = x_0 \begin{bmatrix} 1 \\ 0 \end{bmatrix}^{\otimes 3} + x_3 \begin{bmatrix} 0 \\ 1 \end{bmatrix}^{\otimes 3},$$

and

$$f' = [u_0, u_1, u_2, u_3] = M^{\otimes 3} f = x_0 \begin{bmatrix} \alpha \\ \beta \end{bmatrix}^{\otimes 3} + x_3 \begin{bmatrix} \gamma \\ \delta \end{bmatrix}^{\otimes 3}.$$

By Proposition 2.10 the dot product of $(\alpha, \beta)^{\mathrm{T}}$ and $(\gamma, \delta)^{\mathrm{T}}$ is zero, i.e.,

$$y_1 = \alpha\gamma + \beta\delta = 0.$$

This implies that $ay_0 + by_1 - ay_2 = 0$, a contradiction.

Finally we consider the case $a \neq 0$. By a scaling we can assume $a = 1$. Assume temporarily that $b \neq \pm 2\mathrm{i}$, then the recurrence $x_{k+2} = bx_{k+1} + x_k$ has two distinct eigenvalues λ and μ with $\lambda + \mu = b$ and $\lambda\mu = -1$. So we can write $[x_0, x_1, x_2, x_3] = c \begin{bmatrix} 1 \\ \lambda \end{bmatrix}^{\otimes 3} + d \begin{bmatrix} 1 \\ \mu \end{bmatrix}^{\otimes 3} = \begin{bmatrix} 1 & 1 \\ \lambda & \mu \end{bmatrix}^{\otimes 3} [c, 0, 0, d]$. Then $[u_0, u_1, u_2, u_3] = \left(M \begin{bmatrix} 1 & 1 \\ \lambda & \mu \end{bmatrix} \right)^{\otimes 3} [c, 0, 0, d]$. By Proposition 2.10 we have

$$\det \begin{bmatrix} u_0 - u_2 & u_1 \\ u_1 - u_3 & u_2 \end{bmatrix} = \left(\det M \cdot \det \begin{bmatrix} 1 & 1 \\ \lambda & \mu \end{bmatrix} \right)^2 cd \begin{bmatrix} 1 & \lambda \end{bmatrix} M^{\mathrm{T}} M \begin{bmatrix} 1 \\ \mu \end{bmatrix}. \quad (2.20)$$

Note that the dot product of the two columns of $M \begin{bmatrix} 1 & 1 \\ \lambda & \mu \end{bmatrix}$ is $(1 \ \lambda) M^{\mathrm{T}} M \begin{pmatrix} 1 \\ \mu \end{pmatrix} = (1 \ \lambda) \begin{bmatrix} y_0 & y_1 \\ y_1 & y_2 \end{bmatrix} \begin{pmatrix} 1 \\ \mu \end{pmatrix}$, which is exactly $y_0 + by_1 - y_2 \neq 0$. Also $cd(\lambda - \mu)^2 = \det \begin{bmatrix} x_0 & x_1 \\ x_1 & x_2 \end{bmatrix}$. Indeed the binary signature $[x_0, x_1, x_2] = c \begin{bmatrix} 1 \\ \lambda \end{bmatrix}^{\otimes 2} + d \begin{bmatrix} 1 \\ \mu \end{bmatrix}^{\otimes 2}$ has matrix form $\begin{bmatrix} 1 & 1 \\ \lambda & \mu \end{bmatrix} \begin{bmatrix} c & 0 \\ 0 & d \end{bmatrix} \begin{bmatrix} 1 & 1 \\ \lambda & \mu \end{bmatrix}^{\mathrm{T}}$, with determinant $cd(\lambda - \mu)^2$. Hence (2.20) is equivalent to

$$\det \begin{bmatrix} u_0 - u_2 & u_1 \\ u_1 - u_3 & u_2 \end{bmatrix} = (\det M)^2 \det \begin{bmatrix} x_0 & x_1 \\ x_1 & x_2 \end{bmatrix} (y_0 + by_1 - y_2). \quad (2.21)$$

Equation (2.21) is valid for $b \neq \pm 2\mathrm{i}$. However, we show that it is valid for $b = \pm 2\mathrm{i}$ by a limiting argument. We take a sequence $b^{(n)} \to b$, where $b^{(n)} \neq 0, \pm 2\mathrm{i}$. For the given $g = [y_0, y_1, y_2]$ with $y_0 + by_1 - y_2 \neq 0$ we may define a sequence $g^{(n)} = [y_0, \frac{b}{b^{(n)}} y_1, y_2] \to g$, satisfying $y_0 + b^{(n)}(\frac{b}{b^{(n)}} y_1) - y_2 \neq 0$. From $g^{(n)}$ we may define $M^{(n)}$, with $(M^{(n)})^{\mathrm{T}} M^{(n)}$ being the matrix form of

https://doi.org/10.1017/9781107477063.003 Published online by Cambridge University Press

$g^{(n)}$, and satisfying $M^{(n)} \to M$. Then we can define $x^{(n)} = [x_0, x_1, x_2^{(n)}, x_3^{(n)}]$ by the same recurrence with parameter $b^{(n)}$ and then define $u^{(n)}$ by $M^{(n)}$ from $x^{(n)}$ accordingly. By taking limit in (2.21) for this sequence we get (2.21) valid for $b = \pm 2\mathrm{i}$ as well.

The condition that f' is of type I is equivalent to $\det \begin{bmatrix} u_0 - u_2 & u_1 \\ u_1 - u_3 & u_2 \end{bmatrix} = 0$. We know that $\det M \neq 0$, $y_0 + by_1 - y_2 \neq 0$, hence $\det \begin{bmatrix} x_0 & x_1 \\ x_1 & x_2 \end{bmatrix} = 0$. As the third column in $\begin{bmatrix} x_0 & x_1 & x_2 \\ x_1 & x_2 & x_3 \end{bmatrix}$ is linearly dependent on the first two columns by the recurrence, this says that f is degenerate, a contradiction. □

Lemma 2.15 does the same thing as Lemma 2.14 for tractable cases of type II in Lemma 2.13.

Lemma 2.15. *Let $f = [x, y, -x, -y]$ be of type II and $g = [y_0, y_1, y_2]$, both nondegenerate. If g is not of type II, i.e., $y_0 + y_2 \neq 0$, and g is not of the form $[\lambda, 0, \lambda]$, then* Holant*$(g \mid f)$ *is #P-hard. It follows that* Holant*$(\{f, g\})$ *is also #P-hard.*

Proof. The idea of the proof is the same as for Lemma 2.14.

Let $M = \begin{bmatrix} \alpha & \gamma \\ \beta & \delta \end{bmatrix}$ be the matrix such that $g(M^{-1})^{\otimes 2} = [1, 0, 1]$, then $G = \begin{bmatrix} y_0 & y_1 \\ y_1 & y_2 \end{bmatrix} = M^{\mathrm{T}}M$. Let $f' = M^{\otimes 3} f$. Then f' is nondegenerate since f is. Note that f has the form

$$[x, y, -x, -y] = c \begin{bmatrix} 1 \\ \mathrm{i} \end{bmatrix}^{\otimes 3} + d \begin{bmatrix} 1 \\ -\mathrm{i} \end{bmatrix}^{\otimes 3},$$

where $cd \neq 0$ as f is nondegenerate. We get

$$f' = c \begin{bmatrix} \alpha + \gamma\mathrm{i} \\ \beta + \delta\mathrm{i} \end{bmatrix}^{\otimes 3} + d \begin{bmatrix} \alpha - \gamma\mathrm{i} \\ \beta - \delta\mathrm{i} \end{bmatrix}^{\otimes 3}.$$

We want to show that Holant*(f') is #P-hard. By Lemma 2.13, we have to consider two cases. If f' is of type II, then by Proposition 2.11

$$0 = (\alpha + \gamma\mathrm{i})^2 + (\beta + \delta\mathrm{i})^2 = (\alpha - \gamma\mathrm{i})^2 + (\beta - \delta\mathrm{i})^2.$$

This gives $\alpha\gamma + \beta\delta = 0$ and $\alpha^2 + \beta^2 = \gamma^2 + \delta^2$. By $G = M^{\mathrm{T}}M$, this is precisely $y_1 = 0$ and $y_0 = y_2$. Then $[y_0, y_1, y_2]$ has the form $[\lambda, 0, \lambda]$, a contradiction.

https://doi.org/10.1017/9781107477063.003 Published online by Cambridge University Press

Suppose f' has type I. By Proposition 2.10 we have

$$0 = (\alpha + \gamma \mathrm{i})(\alpha - \gamma \mathrm{i}) + (\beta + \delta \mathrm{i})(\beta + \delta \mathrm{i}) = \alpha^2 + \beta^2 + \gamma^2 + \delta^2.$$

This is precisely $y_0 + y_2 = 0$, a contradiction. □

2.3.3 A Single Signature of Arity n

The following lemma extends Lemma 2.13 to a signature with an arbitrary arity.

Lemma 2.16. *Let $f = [x_0, x_1, \ldots, x_n]$ be a nondegenerate symmetric signature with arity $n > 3$, then* Holant*(f) *is #P-hard unless f has type I or type II.*

Proof. Assume that Holant*(f) is not #P-hard, we prove that f must be of type I or type II. Using unary signatures $\Delta_0 = [1, 0]$ and $\Delta_1 = [0, 1]$, for any subset S of subsignatures of f, Holant*$(S) \leq_{\mathrm{T}}^{\mathrm{p}}$ Holant*(f). Hence Holant*(S) is not #P-hard.

Case A: For every $0 \leq k \leq n-2$, the subsignature $[x_k, x_{k+1}, x_{k+2}]$ is nondegenerate.

By Lemma 2.13, there are two tractable types to be considered for $[x_0, x_1, x_2, x_3]$.

(i) $[x_0, x_1, x_2, x_3]$ has type I(a, b) for some nonzero pair (a, b). Such a nonzero pair (a, b) must be unique up to a scalar factor, by the nondegeneracy of $[x_0, x_1, x_2, x_3]$. Then by Lemma 2.14, for every $0 \leq k \leq n-2$, either $ax_k + bx_{k+1} - ax_{k+2} = 0$ or $[x_k, x_{k+1}, x_{k+2}]$ has the form $[2a\lambda, b\lambda, -2a\lambda]$.

If $a = 0$, then $b \neq 0$, and $x_1 = x_2 = 0$ from the linear equations. Then $[x_0, x_1, x_2]$ is degenerate, a contradiction. So $a \neq 0$ and we can normalize it and assume $a = 1$. Now we show that the form $[2\lambda, b\lambda, -2\lambda]$ cannot appear among $[x_k, x_{k+1}, x_{k+2}]$ $(0 \leq k \leq n-2)$. This will conclude that f has type I(a, b).

Suppose the form $[2\lambda, b\lambda, -2\lambda]$ does occur. Such a λ certainly is nonzero, otherwise it is degenerate. Also $b \neq \pm 2\mathrm{i}$, since $[2, \pm 2\mathrm{i}, -2]$ is degenerate. If it occurs as $[x_k, x_{k+1}, x_{k+2}]$ for $k = 0$ or 1, then together with $ax_0 + bx_1 - ax_2 = 0$ and $ax_1 + bx_2 - ax_3 = 0$, and $a = 1$, we get $b = \pm 2\mathrm{i}$, a contradiction. Thus let k be the minimum index such that $[x_k, x_{k+1}, x_{k+2}]$ has the form $\lambda[2, b, -2]$; then $k \geq 2$, and there is a subsignature $[x_{k-2}, x_{k-1}, x_k, x_{k+1}, x_{k+2}] = \lambda[x, y, 2, b, -2]$ with arity 4, where $ax + by - 2a = 0$ and $ay + 2b - ab = 0$. So the subsignature is $\lambda[b^2 + 2, -b, 2, b, -2]$, of which there is a subsignature $\lambda[-b, 2, b, -2]$. Applying Lemma 2.15 to the nondegenerate signature pair $[-b, 2, b, -2]$ and $[b^2 + 2, -b, 2]$, we have $b^2 + 2 = -2$ or $b^2 + 2 = 2$. The first case would imply that $b = \pm 2\mathrm{i}$. So we have $b^2 + 2 = 2$ and thus $b = 0$. In this case, the

https://doi.org/10.1017/9781107477063.003 Published online by Cambridge University Press

arity 4 subsignature is $[2, 0, 2, 0, -2]$. Using a unary signature $[1, 1]$, we can get $[2, 2, 2, -2]$. By Lemma 2.13, Holant*$([2, 2, 2, -2])$ is #P-hard.

(ii) If $[x_0, x_1, x_2, x_3]$ is of the form $[x, y, -x, -y]$, then by Lemma 2.15, for every $0 \le k \le n-2$, we have $x_k + x_{k+2} = 0$ or $[x_k, x_{k+1}, x_{k+2}]$ is of the form $[\lambda, 0, \lambda]$. We prove that $[\lambda, 0, \lambda]$ can not appear. This will conclude that f has type II.

Suppose the form $[\lambda, 0, \lambda]$ does appear among $[x_k, x_{k+1}, x_{k+2}]$. $\lambda \neq 0$ by nondegeneracy. It is easy to see that if it occurred at $k = 0$ or 1, then $[x_0, x_1, x_2, x_3] = [x, y, -x, -y]$ is degenerate. Let k be the minimum index such that $[x_k, x_{k+1}, x_{k+2}] = [\lambda, 0, \lambda]$, then $k \ge 2$. To its left we have $x_\ell + x_{\ell+2} = 0$, for $0 \le \ell < k$. It follows that there must be a subsignature (after a scaling) of the form $[1, 0, -1, 0, -1]$. Again we use a unary signature $[1, 1]$ and get $[1, -1, -1, -1]$. By Lemma 2.13 Holant*$([1, -1, -1, -1])$ is #P-hard.

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

If f is of the form $[x_0, 0, \ldots, 0, x_n]$, then f has type I$(0, 1)$ and Lemma 2.16 holds. In the following we assume f is not of this form.

By Proposition 2.8, there exists $0 \le s \le n-3$, such that the arity 3 subsignature $f' = [x_s, x_{s+1}, x_{s+2}, x_{s+3}]$ is nondegenerate. We want to find a nondegenerate subsignature of arity 3 that contains a *degenerate* subsignature of arity 2. Starting from a degenerate subsignature $g = [x_k, x_{k+1}, x_{k+2}]$ of arity 2, if $s \le k \le s+1$, then g is already a subsignature of f'. Otherwise, either $k < s$ or $k > s+1$. Suppose $k < s$. Consider the subsignature $[x_{k+1}, x_{k+2}, x_{k+3}]$. If it is degenerate, we can substitute it for g, and continue. If it is nondegenerate, then we can substitute f' by $[x_k, x_{k+1}, x_{k+2}, x_{k+3}]$, and it will also be nondegenerate. The proof for $k > s+1$ is similar.

Thus we can find f' and g as specified. This f' must be of the form $[s^2, sr, r^2, x]$ or $[y, s^2, sr, r^2]$. By symmetry, we consider only $f' = [s^2, sr, r^2, x]$. By Lemma 2.13, we have two cases. For the first case, f' has type I. Then $\det \begin{bmatrix} s^2 - r^2 & sr \\ sr - x & r^2 \end{bmatrix} = 0$, by Proposition 2.10. This implies that $srx = r^4$. If $sr \neq 0$, $srx = r^4$ implies that $[s^2, sr, r^2, x]$ is degenerate, a contradiction. If $sr = 0$, then $r = 0$ by $srx = r^4$. Since $[s^2, sr, r^2, x] = [s^2, 0, 0, x]$ is nondegenerate, we must have $s^2 \neq 0$ and $x \neq 0$. Because $n > 3$, there must be a subsignature of f of the form $[s^2, 0, 0, x, z]$ or $[z', s^2, 0, 0, x]$. Suppose we have $[s^2, 0, 0, x, z]$; the other case is similar. Then consider the pair $[s^2, 0, 0, x]$ and $[0, x, z]$. If $z \neq 0$, then Holant*$(\{[s^2, 0, 0, x], [0, x, z]\})$ is #P-hard by Lemma 2.14. If $z = 0$, then we have a subsignature $[0, 0, x, 0]$. Holant*$([0, 0, x, 0])$ is also #P-hard by Lemma 2.13. Finally, for the second case, f' has type II: $s^2 + r^2 = 0$ and

https://doi.org/10.1017/9781107477063.003 Published online by Cambridge University Press

$sr + x = 0$. Then the signature must be $[s^2, s^2\mathrm{i}, -s^2, -s^2\mathrm{i}]$ or $[s^2, -s^2\mathrm{i}, -s^2, s^2\mathrm{i}]$. Both are degenerate, a contradiction. □

2.3.4 Putting Things Together

Finally we extend these lemmas to handle a set of signatures and finish the proof of Theorem 2.12. The main idea is to find a nondegenerate signature of arity 3, and anchor the proof on this arity 3 signature, namely to show that everything else must conform to it, or else the problem is #P-hard.

Proof. of Theorem 2.12: The tractability part has already been proved. We prove the hardness part. We assume every signature in $\mathcal{F}$ is nondegenerate and has arity ≥ 2.

Suppose Holant*($\mathcal{F}$) is *not* #P-hard. We want to show that $\mathcal{F}$ must fall into one of the three classes.

If no signature of $\mathcal{F}$ has arity ≥ 3, then this is in class $\mathfrak{A}$ and we are done. Suppose there are some signatures of arity ≥ 3. Each such signature is of type I or II by Lemma 2.13 and Lemma 2.16. First suppose all signatures in $\mathcal{F}$ of arity ≥ 3 are of the same type, namely either for some (a, b) they are all of type I(a, b) or they are all of type II. If they are all of type I(a, b), then we may obtain a nondegenerate signature of arity 3 having the same type by either taking a nondegenerate subsignature (Proposition 2.8), or connecting it with the unary $[1, 1]$ if it is of type I$(0, 1)$. Lemma 2.14 finishes the proof in this case and we are in class $\mathfrak{B}$. If all signatures in $\mathcal{F}$ of arity ≥ 3 are of type II, again we can obtain a nondegenerate subsignature of arity 3. Lemma 2.15 then finishes the proof and we are in class $\mathfrak{C}$.

In the following we assume there is more than one type of signatures in $\mathcal{F}$ having arity ≥ 3. Each such signature declares a type I or II.

- Case 1. Suppose there is a nondegenerate type I$(0, 1)$ signature in $\mathcal{F}$ of arity ≥ 3. By connecting it with the unary $[1, 1]$ we may assume we have a nondegenerate type I$(0, 1)$ signature $f \in \mathcal{F}$ of arity 3. By assumption, there exists another nondegenerate signature g of arity ≥ 3 and of a different type. By Proposition 2.8, there exists a nondegenerate subsignature $[x, y, z]$ of arity two. By being *not* #P-hard, we apply Lemma 2.14 to the pair f and $[x, y, z]$ and conclude that there are two cases: $[x, y, z] = [x, 0, z]$, or $[0, y, 0]$. Being nondegenerate, $xz \neq 0$ in the first case, and $y \neq 0$ in the second case. Within g there is a subsignature containing $[x, y, z]$. In the first case, it has the form $[w, x, 0, z]$ (or $[x, 0, z, w]$); then $[w, x, 0]$ (or $[0, z, w]$) is nondegenerate. We apply Lemma 2.14 once again to f and $[w, x, 0]$ (or $[0, z, w]$), and get $w = 0$. In the second case, it has the form $[u, 0, y, 0]$

https://doi.org/10.1017/9781107477063.003 Published online by Cambridge University Press

(or $[0, y, 0, u]$). If $u = 0$, Holant*$([0, 0, y, 0])$ (or Holant*$([0, y, 0, 0])$) is #P-hard by Lemma 2.13. Hence in all cases we get a signature of the form $[u, 0, v, 0]$ (or its reversal $[0, v, 0, u]$), where $uv \neq 0$. By symmetry we consider $[u, 0, v, 0]$, as $[0, v, 0, u]$ is changed to $[u, 0, v, 0]$ by a 0-1 exchange, and the type I(0, 1) signature f retains its type. We connect a unary $[s, t]$ to $[u, 0, v, 0]$, and get $[su, tv, sv]$. As long as $st \neq 0$ and $s^2/t^2 \neq v/u$ this is nondegenerate and Lemma 2.14 on the pair f and $[su, tv, sv]$ leads to #P-hardness.

- Case 2. Now we assume case 1 does not apply, but there exists a nondegenerate type I(a, b) signature in $\mathcal{F}$ of arity ≥ 3, where $a \neq 0$ and $b \neq \pm 2a\mathrm{i}$. By Proposition 2.8, we may assume there is a nondegenerate type I(a, b) signature f having arity 3. We may normalize (a, b) to $a = 1$, and $b \neq \pm 2\mathrm{i}$.

  Then the characteristic equation $X^2 - bX - 1$ has two distinct roots $\lambda \neq \mu$. It follows that $\lambda\mu = -1$, and $\lambda, \mu \neq \pm\mathrm{i}$. Define $T = \begin{bmatrix} \frac{1}{\sqrt{1+\lambda^2}} & \frac{1}{\sqrt{1+\mu^2}} \\ \frac{\lambda}{\sqrt{1+\lambda^2}} & \frac{\mu}{\sqrt{1+\mu^2}} \end{bmatrix}$, then T is orthogonal. It is easy to verify that $f = T^{\otimes 3} f'$ for some $f' = [x_0, 0, 0, x_3]$, a nondegenerate signature of type I(0, 1). Performing a holographic transformation by T on all signatures in $\mathcal{F}$, we have reduced case 2 to case 1.
- Case 3. We assume cases 1 and 2 do not apply, but there exists a nondegenerate type I$(1, b)$ signature in $\mathcal{F}$ of arity ≥ 3, where $b = \pm 2\mathrm{i}$. By Proposition 2.8, we may assume such a signature f of arity 3. We still suppose there exists a nondegenerate signature g of arity ≥ 3 and of a different type; for otherwise we are done. g could be of type I$(1, b')$ or II, where $b' = \mp 2\mathrm{i}$ (but $b' \neq b$, and hence $b' = -b$). By Proposition 2.8, there exists a nondegenerate subsignature $[x, y, z]$ of g.

If g has type I$(1, b')$, then $x + b'y - z = 0$. Note that for $b = \pm 2\mathrm{i}$, the binary signature $[2, b, -2]$ is degenerate. Since $[x, y, z]$ is nondegenerate, it cannot be $\lambda[2, b, -2]$. Hence, by applying Lemma 2.14 to f and $[x, y, z]$, we get $x + by - z = 0$. This implies that $y = 0$ since $b' \neq b$, and $x = z \neq 0$ by the nondegeneracy of $[x, y, z]$. But within g there is a subsignature $[w, x, 0, x]$ or $[x, 0, x, w]$. As $x \neq 0$, $[w, x, 0]$ and $[0, x, w]$ are nondegenerate. Applying Lemma 2.14 once again we get a contradiction. If g is of type II, then we still have $x + by - z = 0$ as well as $z = -x$. Since $b = \pm 2\mathrm{i}$, this implies that $[x, y, z]$ is degenerate, also a contradiction.

If cases 1, 2 and 3 all do not apply, then all nondegenerate signatures in $\mathcal{F}$ of arity ≥ 3 are of type II. This case has been proved before; we are in class $\mathfrak{C}$. Theorem 2.12 is proved. □

https://doi.org/10.1017/9781107477063.003 Published online by Cambridge University Press

2.4 The Road Ahead

Theorem 2.12 is our first major dichotomy theorem. Although it covers a fairly broad class of problems, we should also recognize its limitations. There are at least three aspects in which Theorem 2.12 is restricted.

The first is the restriction to a symmetric set of constraint functions $\mathcal{F}$. One can ask what happens when we remove this restriction of symmetry. The answer will be given in Chapter 10, where we present a dichotomy theorem, Theorem 10.14, for Holant* problems applicable to a set of constraint functions $\mathcal{F}$ that are not necessarily symmetric. In so doing we will further reveal an alternative perspective on the classification achieved here. To arrive at Theorem 10.14, we will have to develop the theory much further; in particular, the proof of Theorem 10.14 will use Theorem 2.12, the dichotomy for the symmetric case.

The second is the assumption of the presence of all unary functions $\mathcal{U}$ for free in the definition of Holant* problems. One can ask what happens if this assumption is removed. In Chapter 7 we will present a dichotomy for Holant problems on symmetric signatures without assuming any freely available auxiliary functions, in particular unary functions. The proof will utilize a substantial part of the theory to be developed in the next few chapters. We develop this theory starting with free unary functions. The theory will also consider other freely available auxiliary functions, for example, when EQUALITY functions of all arities are assumed to be free, namely Boolean #CSP problems, which we turn to next in Chapter 3. We will also consider the case where only two special unary functions, Δ_0 and Δ_1, are free. This is called Holant^c problems. It is an interesting open problem when we remove both the symmetry restriction and the presence of unary functions. It is open even for Holant^c problems when the symmetry restriction is removed. However, the study of Holant* problems by assuming unary functions $\mathcal{U}$ are free is more than an arbitrary choice. As we develop the theory, aside from explicitly using Theorem 2.12, the *idea* of simulating the free presence of unary functions is a major theme.

Finally, the third limitation is that we restrict our study to the Boolean domain in Volume I of this book. In Volume II we will develop the theory on a general finite domain. However, the theory on general domains is not as well developed as on the Boolean domain. In particular, the analogue of Theorem 2.12 is still open for the general domain.

https://doi.org/10.1017/9781107477063.003 Published online by Cambridge University Press