# 6

# Holant Problems and #CSP

We have by now encountered three representative frameworks for counting problems: Holant problems, #CSP, and spin systems. #CSP can be viewed as the special case of Holant problems where the local constraint function set is assumed to contain EQUALITY of all arities. If we allow other auxiliary functions to be free, such as all unary functions, we get other restricted Holant problems, in this case Holant$^*$ problems. Spin systems can be seen as a further specialization of #CSP. At the same time, Holant problems can be viewed as #CSP where each variable appears twice, syntactically.

Their relationship goes deeper than this. We have already seen in Chapters 3 and 5 how results from Holant$^*$ problems can be used to derive dichotomy theorems for Boolean #CSP and 2-spin systems. In this chapter we will go the other way. We use results from Boolean #CSP and 2-spin systems to derive several results on Holant problems. We introduce, and prove a dichotomy for, a class of Holant problems called Holant$^c$ problems, an intermediate class between Holant and Holant$^*$ problems. We also prove a dichotomy for a version of Boolean #CSP, called #CSP$^d$ where each variable appears a multiple of $d$ times, for some integer $d \geq 1$. Finally we discuss a concrete problem called the EULERIAN ORIENTATION problem, which has a very simple expression as a Holant problem. Although interesting in their own right, the primary purpose of these results is to pave the way for the dichotomy theorem for Holant problems without auxiliary functions in Chapter 7.

In addition to the new theorems, this chapter also introduces a new transformational perspective not only as a proof technique but also how a dichotomy is conceptually understood and even stated. Some group theoretic ideas are naturally introduced in this setting.



https://doi.org/10.1017/9781107477063.007 Published online by Cambridge University Press

## 6.1 A Dichotomy for a Single Ternary Signature

In this section, we consider the complexity of Holant($f$), where $f = [f_0, f_1, f_2, f_3]$ is a symmetric signature of arity 3. We also denote by Pl-Holant($f$) the restriction of Holant($f$) to planar graphs. If $f$ is degenerate, then Holant($f$) is trivially tractable, since a degenerate signature factors as a tensor product of unary signatures, and the signature grid simply decomposes into isolated edges. In the following we assume that $f$ is nondegenerate. In Chapter 2 (Section 2.3.1), we have shown that any nondegenerate signature $f = [f_0, f_1, f_2, f_3]$ can be expressed in one of the following three forms of parameterization (with the convention that $\alpha^0 = 1$, and $k\alpha^{k-1} = 0$ if $k = 0$, even when $\alpha = 0$):

1. $f_k = \alpha^{3-k}\beta^k + \gamma^{3-k}\delta^k$, where $\det \begin{bmatrix} \alpha & \gamma \\ \beta & \delta \end{bmatrix} \neq 0$
2. $f_k = ck\lambda^{k-1} + d\lambda^k$, where $c \neq 0$
3. $f_k = c(3-k)\lambda^{2-k} + d\lambda^{3-k}$, where $c \neq 0$

The first form corresponds to the case when the characteristic equation of a second-order linear recurrence has two distinct roots and we call it the *generic* case. The second form, and its reversal in the third form, correspond to the *double-root* case. We can omit the formal proof for the third form since it is symmetric to the second form.

For the *generic* case, we can apply a holographic transformation using $T = \begin{bmatrix} \alpha & \gamma \\ \beta & \delta \end{bmatrix}$. We have the following reduction:

$$\begin{aligned} \text{Holant}(f) &\equiv_{\text{T}} \text{Holant}(=_2 \mid f) \\ &\equiv_{\text{T}} \text{Holant}([1,0,1]T^{\otimes 2} \mid (T^{-1})^{\otimes 3}[f_0, f_1, f_2, f_3]) \\ &\equiv_{\text{T}} \text{Holant}(g \mid [1,0,0,1]), \end{aligned}$$

where $g = [1,0,1]T^{\otimes 2}$. The crucial step is of course $[f_0, f_1, f_2, f_3] = T^{\otimes 3}[1,0,0,1]$.

Therefore for the *generic* case, the complexity of Holant($f$) is equivalent to the spin system on 3-regular graphs Holant($g \mid =_3$), where $g$ is a symmetric binary signature, for which Theorem 5.1 is a dichotomy (the cases $k = 3$ and 4 are proved in Chapter 5). Theorem 5.3 is an equivalent statement of Theorem 5.1. Theorem 5.61 is a more conceptual form of Theorems 5.1 and 5.3. For easy reference we state these together below for $k = 3$.

**Theorem 6.1.** *The problem* Holant($[x, y, z] \mid =_3$) *is #P-hard for all* $x, y, z \in \mathbb{C}$ *except in the following cases, where the problem is solvable in polynomial time:*

https://doi.org/10.1017/9781107477063.007 Published online by Cambridge University Press

*(1)* $xz = y^2$*; (2)* $y = 0$*; (3)* $x = z = 0$*; or (4)* $x^{12} = y^{12}$ *and* $xz = -y^2$*. This is also valid for* Holant$([x, y, z], \Delta_0, \Delta_1 | =_3)$*, and when it is #P-hard a reduction exists that does not use Lemma 3.13.*

*Furthermore, in addition to these cases* Pl-Holant$(f)$ *is solvable in polynomial time if* $x^3 = z^3$*, and remains* #P-hard *in all other cases.*

*Equivalently,* Holant$([x, y, z] | =_3)$ *is in P if there exists* $T \in \mathbf{GL}_2(\mathbb{C})$ *such that*

1. $[x, y, z]T^{\otimes 2} \in \mathscr{P}$ *and* $(=_3) \in T\mathscr{P}$*, or*
2. $[x, y, z]T^{\otimes 2} \in \mathscr{A}$ *and* $(=_3) \in T\mathscr{A}$*.*

*Otherwise it is* #P-hard*. In addition,* Pl-Holant$(f)$ *is in P if there exists* $T \in \mathbf{GL}_2(\mathbb{C})$ *such that* $[x, y, z]T^{\otimes 2} \in \mathscr{M}$ *and* $(=_3) \in T\mathscr{M}$*. It remains* #P-hard *in all other cases.*

**Remark:** The signature $[x, y, z]$ in case (1) is degenerate, which is a special case of product type. Cases (2) and (3) are of product type, being binary generalized Equality and Disequality. These are the only cases for a binary symmetric signature to be of product type. It is easy to show that assuming $y \neq 0$, condition (4) can be equivalently stated as (4′) $x^6 = z^6$ and $xz = -y^2$. Case (4) includes what is transformable to an affine signature by $T = \begin{bmatrix} 1 & 0 \\ 0 & \omega \end{bmatrix}$, where $\omega^3 = 1$ and $T$ keeps $(=_3)$ invariant. One can easily verify that the only matrices $T$ that keep $(=_3)$ invariant are of the form $\begin{bmatrix} \omega & 0 \\ 0 & \omega' \end{bmatrix}$ or $\begin{bmatrix} 0 & \omega \\ \omega' & 0 \end{bmatrix}$, where $\omega^3 = \omega'^3 = 1$, and such matrices $T$ preserve $\mathscr{P}$. The tractable cases of Theorem 6.1 can also be equivalently stated as follows: There exists some $T$ such that $[x, y, z]T^{\otimes 2} \in \mathscr{A} \cup \mathscr{P}$ (additionally in $\mathscr{M}$ for Pl-Holant$(f)$) and $(T^{-1})^{\otimes 3}(=_3) = (=_3)$.

To get a dichotomy for Holant$([f_0, f_1, f_2, f_3])$, we next deal with the *double-root* case. In this part of the proof we will use nondiagonal holographic transformations. We have the following lemma:

**Lemma 6.2.** *Let* $f_k = ck\lambda^{k-1} + d\lambda^k$*, where* $c \neq 0$ *and* $0 \leq k \leq 3$*. If* $\lambda = \pm \mathrm{i}$*, then* Holant$([f_0, f_1, f_2, f_3])$ *is in P; in fact, the Holant value is always 0 in this case. If* $\lambda \neq \pm \mathrm{i}$*, then* Holant$([f_0, f_1, f_2, f_3])$ *is #P-hard. Additionally,* Pl-Holant$(f)$ *is in P if there exists orthogonal* $T \in \mathbf{O}_2(\mathbb{C})$ *such that* $(=_2)T^{\otimes 2} = (=_2) \in \mathscr{M}$ *and* $T^{\otimes 3}[f_0, f_1, f_2, f_3] = [0, 1, 0, 0] \in \mathscr{M}$*, and it remains #P-hard otherwise.*

*Proof.* If $\lambda = \pm \mathrm{i}$, the signature $f = [f_0, f_1, f_2, f_3]$ satisfies the recurrence relation $f_{k+2} = 2\lambda f_{k+1} + f_k$, where $k = 0, 1$. This is a generalized Fibonacci

https://doi.org/10.1017/9781107477063.007 Published online by Cambridge University Press

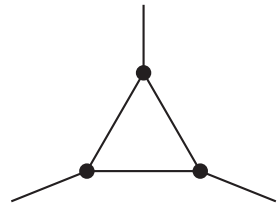

Figure 6.1. All three vertices are assigned the signature $[z, 1, 0, 0]$.

signature. Thus Holant$(f)$ is solvable in P. Note that for $\lambda = \pm\mathrm{i}$, $f$ can also be written as $Z[u, v, 0, 0]$ or $Z[0, 0, v, u]$ where $Z = \begin{bmatrix} 1 & 1 \\ \mathrm{i} & -\mathrm{i} \end{bmatrix}$ is defined in (2.11). In fact, for $\lambda = \mathrm{i}$ we may take $u = -(f_0 + 3\mathrm{i}f_1)/2$ and $v = (f_0 + \mathrm{i}f_1)/2$. Similarly for $\lambda = -\mathrm{i}$. Under a holographic transformation by $Z$, Holant$(f) \equiv_{\mathrm{T}}$ Holant$(\neq_2 |\ (Z^{-1})^{\otimes 3} f)$. As $(Z^{-1})^{\otimes 3} f = [u, v, 0, 0]$ or $[0, 0, v, u]$, in any bipartite signature grid for Holant$(\neq_2 |\ (Z^{-1})^{\otimes 3} f)$, the left-hand side (LHS) $(\neq_2)$ requires exactly the same number of 0's and 1's, while the right-hand side (RHS) requires strictly more, or strictly less, 0's than 1's. Thus the Holant value is always 0.

In the following we assume $\lambda \neq \pm\mathrm{i}$. We can view Holant$(f)$ as the bipartite Holant$(=_2 |\, f)$. Recall from (2.8) that an orthogonal transformation $T \in \mathbf{O}_2(\mathbb{C})$ keeps the binary EQUALITY $(=_2)$ invariant: $(=_2)T^{\otimes 2} = (=_2)$.

Let $M = \begin{bmatrix} 1 & \frac{d-1}{3} \\ \lambda & c + \frac{d-1}{3}\lambda \end{bmatrix}$, then the signature $[f_0, f_1, f_2, f_3]$ can be expressed as

$$(f_0, f_1, f_1, f_2, f_1, f_2, f_2, f_3)^{\mathrm{T}} = M^{\otimes 3}(1, 1, 1, 0, 1, 0, 0, 0)^{\mathrm{T}}.$$

Let $Q = \frac{1}{\sqrt{1+\lambda^2}} \begin{bmatrix} 1 & \lambda \\ \lambda & -1 \end{bmatrix}$, then $Q$ is orthogonal. In exactly the same way as in (2.18) from Chapter 2 (replacing $\lambda$ for $\alpha$), the orthogonal transformation $Q$ transforms $f$ to $[z, 1, 0, 0]$ for some $z \in \mathbb{C}$, up to a scalar. Then the complexity of Holant$(f)$ is the same as that of Holant$([z, 1, 0, 0])$.

For $z = 0$ in the signature $[z, 1, 0, 0]$, it is the problem of counting perfect matchings on 3-regular graphs. Because $[0, 1, 0, 0] \in \mathscr{M}$, Pl-Holant$(f)$ is in P in this case. Over general 3-regular graphs Holant$([0, 1, 0, 0])$ is #P-hard [DL92].

In the following we assume $z \neq 0$. We show that Pl-Holant$([z, 1, 0, 0])$ is #P-hard. We can realize a new ternary signature $f' = [z^3 + 3z, z^2 + 1, z, 1]$ by connecting $[z, 1, 0, 0]$'s in a triangle gadget as illustrated in Figure 6.1. So if Pl-Holant$(f')$ is #P-hard then Pl-Holant$([z, 1, 0, 0])$ is also #P-hard. To see that the gadget in Figure 6.1 has the signature $[z^3 + 3z, z^2 + 1, z, 1]$ note that all

https://doi.org/10.1017/9781107477063.007 Published online by Cambridge University Press

three vertices are assigned the weighted matching signature $[z, 1, 0, 0]$ and can have at most one incident edge assigned 1. Thus, for example, on external input 000, there are three ways to assign exactly one internal edge the value 1 giving the value $3z$, and one way to assign all internal edges the value 0 giving the value $z^3$. The other entries of $f'$ can be computed similarly.

In tensor product notation this signature is

$$f' = [z^3 + 3z, z^2 + 1, z, 1]^{\mathrm{T}} = \frac{1}{2}\left(\begin{bmatrix} z+1 \\ 1 \end{bmatrix}^{\otimes 3} + \begin{bmatrix} z-1 \\ 1 \end{bmatrix}^{\otimes 3}\right).$$

Then the following reduction chain holds:

$$\begin{aligned} \text{Pl-Holant}(f') &\equiv_{\mathrm{T}} \text{Pl-Holant}(=_2 \mid f') \\ &\equiv_{\mathrm{T}} \text{Pl-Holant}([z^2 + 2z + 2, z^2, z^2 - 2z + 2] \mid =_3), \end{aligned} \tag{6.1}$$

where the second step is a holographic reduction using $T = \begin{bmatrix} z+1 & z-1 \\ 1 & 1 \end{bmatrix} \in \mathbf{GL}_2(\mathbb{C})$, and $T$ transforms $(=_2)$ to $(=_2)T^{\otimes 2}$, namely

$$T^{\mathrm{T}} I_2 T = \begin{bmatrix} z^2 + 2z + 2 & z^2 \\ z^2 & z^2 - 2z + 2 \end{bmatrix}.$$

Now we can apply Theorem 6.1 to the Holant problem in (6.1), and check against the tractable cases.

1. $[z^2 + 2z + 2, z^2, z^2 - 2z + 2]$ is nondegenerate. This is because it is the transformed binary signature from the nondegenerate $(=_2)$ under the nonsingular matrix $T \in \mathbf{GL}_2(\mathbb{C})$.
2. $z^2 \neq 0$, since by assumption $z \neq 0$.
3. It cannot be the case that $z^2 + 2z + 2 = z^2 - 2z + 2 = 0$, again by $z \neq 0$.
4. We verify the equivalent condition (4′) stated in the Remark after Theorem 6.1, that there is no solution to $(z^2 + 2z + 2)^6 = (z^2 - 2z + 2)^6$ and $(z^2 + 2z + 2)(z^2 - 2z + 2) + z^4 = 0$. This is by a direct computation.

Therefore Holant($f'$) is #P-hard, and so is Holant($[z, 1, 0, 0]$) for all $z \in \mathbb{C}$.

Next we consider the planar case, with $z \neq 0$. We still have the equivalence Pl-Holant($f$) $\equiv_{\mathrm{T}}$ Pl-Holant($[z, 1, 0, 0]$) by the orthogonal transformation $Q$.

To show that Pl-Holant($[z, 1, 0, 0]$) is also #P-hard for $z \neq 0$, we need to rule out the additional planar tractable case $(z^2 + 2z + 2)^3 = (z^2 - 2z + 2)^3$, if we still use the gadget in Figure 6.1. This condition is

$$3z^4 + 16z^2 + 12 = 0. \tag{6.2}$$

https://doi.org/10.1017/9781107477063.007 Published online by Cambridge University Press



Figure 6.2. Both vertices are assigned the signature $[z, 1, 0, 0]$.

If $3z^4 + 16z^2 + 12 \neq 0$ then Pl-Holant($[z, 1, 0, 0]$) is #P-hard by the gadget in Figure 6.1. There are four exceptional values of $z$, which we prove separately as follows. Note that for these values, $z^2$ is real and $z^2 < 0$.

In addition to the gadget in Figure 6.1, we can construct a gadget in Pl-Holant($[z, 1, 0, 0]$) with a binary signature $[z^2 + 2, z, 1]$ in Figure 6.2. Thus the bipartite planar Holant problem Pl-Holant($[z^2 + 2, z, 1] | f'$) is reducible to Pl-Holant($[z, 1, 0, 0]$). Under the same transformation $T = \begin{bmatrix} z+1 & z-1 \\ 1 & 1 \end{bmatrix} \in \mathbf{GL}_2(\mathbb{C})$, we get an equivalent problem Pl-Holant($[X, Y, Z] \mid [1, 0, 0, 1]$), where $\begin{bmatrix} X & Y \\ Y & Z \end{bmatrix} = T^{\mathrm{T}} \begin{bmatrix} z^2+2 & z \\ z & 1 \end{bmatrix} T$; thus

$$X = z^4 + 2z^3 + 5z^2 + 6z + 3,$$
$$Y = z^4 + 3z^2 - 1,$$
$$Z = z^4 - 2z^3 + 5z^2 - 6z + 3.$$

Now again we can apply Theorem 6.1 to this problem, and verify that no tractable condition is satisfied for any root of (6.2). For the reader's convenience we include this verification as follows.

1. Nondegeneracy of $[X, Y, Z]$ follows from the fact that it is transformed from a nondegenerate binary signature by $T$.
2. $Y = 0$ is incompatible with (6.2).
3. It cannot be the case that $X = Z = 0$ assuming (6.2). In fact $X = Z$, which is equivalent to $z^3 + 3z = 0$, is already incompatible with (6.2).

4′. There is no solution to $XZ = -Y^2$ and $X^6 = Z^6$. Luckily we don't have to check $X^6 = Z^6$; $XZ = -Y^2$ is already incompatible with (6.2). We can write $X = A + zB$ and $Z = A - zB$, where $A = z^4 + 5z^2 + 3 \in \mathbb{R}$ and $B = 2z^2 + 6 \in \mathbb{R}$ by (6.2). Then $XZ + Y^2 = 0$ is equivalent to $A^2 + Y^2 = z^2B^2$, yet $A^2 + Y^2 > 0$ and $z^2B^2 < 0$ by $z^2 < 0$, which follows from (6.2).

5. Finally, for the planar condition $X^3 = Z^3$, we already know that $X = Z$ is incompatible with (6.2). We can verify that $X^2 + XZ + Z^2 = 0$ is also incompatible. In fact $X^2 + XZ + Z^2 = 3A^2 + z^2B^2$. By (6.2), $3A = -(z^2 + 3)$ and $B = 2(z^2 + 3)$, and therefore $(3A)^2 + 3z^2B^2 = (z^2 + 3)^2(12z^2 + 1)$, which is not zero, again by (6.2).

https://doi.org/10.1017/9781107477063.007 Published online by Cambridge University Press

This completes the proof of Lemma 6.2. □

By Theorem 6.1 and Lemma 6.2, we have a dichotomy theorem for Holant($[f_0, f_1, f_2, f_3]$). From this, we can get a further dichotomy for all bipartite Holant problems on 2-3-regular graphs Holant($[g_0, g_1, g_2]|[f_0, f_1, f_2, f_3]$). The reduction is standard. For any nondegenerate $[g_0, g_1, g_2]$, we can find a transformation $M$, such that $[g_0, g_1, g_2] = [1, 0, 1]M^{\otimes 2}$, as shown in (2.19) in Chapter 2. Then the bipartite problem Holant($[g_0, g_1, g_2]|[f_0, f_1, f_2, f_3]$) is transformed to the equivalent problem Holant($M^{\otimes 3}[f_0, f_1, f_2, f_3]$), for which we can apply our dichotomy.

We state Theorems 6.3 and 6.4 in a more conceptual form, in terms of what is transformable to various tractable classes $\mathscr{A}$, $\mathscr{P}$ and $\mathscr{M}$.

**Theorem 6.3.** *Let $f = [f_0, f_1, f_2, f_3]$ be a ternary symmetric signature.* Holant($f$) *is #P-hard unless $f$ satisfies one of the following conditions, in which case the problem is in P:*

1. *$f$ is degenerate.*
2. *There exists $T \in \mathbf{GL}_2(\mathbb{C})$ such that $f = T^{\otimes 3}(=_3)$, and the signature $(=_2)T^{\otimes 2} \in \mathscr{A} \cup \mathscr{P}$.*
3. *For $\lambda \in \{2\mathrm{i}, -2\mathrm{i}\}$, $f_2 - \lambda f_1 - f_0 = 0$ and $f_3 - \lambda f_2 - f_1 = 0$.*

*Additionally,* Pl-Holant($f$) *is in P if there exists $T \in \mathbf{GL}_2(\mathbb{C})$ such that $(=_2)T^{\otimes 2} \in \mathscr{M}$ and $f \in T\mathscr{M}$.* Pl-Holant($f$) *is #P-hard otherwise.*

As shown in Section 2.2, $f = [f_0, f_1, f_2, f_3]$ satisfies the condition in item (3) in Theorem 6.3 iff $f$ is a generalized Fibonacci gate with parameter $\lambda = \pm 2\mathrm{i}$. This happens iff $f$ has the form $f_k = ck\mu^{k-1} + d\mu^k$ ($0 \le k \le 3$ and $\mu = \pm\mathrm{i}$). Furthermore, this happens iff $(Z^{-1})^{\otimes 3} f$ has the form $[x, y, 0, 0]$ or its reversal $[0, 0, y, x]$, where $Z = \begin{bmatrix} 1 & 1 \\ \mathrm{i} & -\mathrm{i} \end{bmatrix}$. (See also the remark after Theorem 3.5.) Thus, item (3) in Theorem 6.3 can also be stated as $(Z^{-1})^{\otimes 3} f$ has the form $[x, y, 0, 0]$ or its reversal $[0, 0, y, x]$, while $(=_2)Z^{\otimes 2} = 2(\neq_2)$, a constant multiple of the binary Disequality function. Signatures of the form $Z[x, y, 0, 0]$ or $Z[0, 0, y, x]$ are special cases of the so-called *vanishing* signatures; they will be discussed in Chapter 7.

Thus we can state condition (3) of Theorem 6.3 equivalently as (3′): $(Z^{-1})^{\otimes 3} f$ has the form $[*, *, 0, 0]$ or its reversal $[0, 0, *, *]$.

The next theorem is concerned with 2-3-regular graphs.

**Theorem 6.4.** *Let $f = [f_0, f_1, f_2, f_3]$ and $g = [g_0, g_1, g_2]$ be symmetric signatures of arity 3 and 2 respectively.* Holant($g|f$) *is #P-hard unless $f$ and $g$ satisfy one of the following conditions, in which case the problem is in P:*

https://doi.org/10.1017/9781107477063.007 Published online by Cambridge University Press

1. *$f$ or $g$ is degenerate.*
2. *There exists $T \in \mathbf{GL}_2(\mathbb{C})$ such that $f = T^{\otimes 3}(=_3)$, and the signature $gT^{\otimes 2} \in \mathscr{A} \cup \mathscr{P}$.*
3. *There exists $T \in \mathbf{GL}_2(\mathbb{C})$ such that $f = T^{\otimes 3}[1, 1, 0, 0]$, and $gT^{\otimes 2}$ is of form $[0, *, *]$.*

*Additionally,* Pl-Holant$(g \mid f)$ *is in P if there exists $T \in \mathbf{GL}_2(\mathbb{C})$ such that $gT^{\otimes 2} \in \mathscr{M}$ and $f \in T\mathscr{M}$.* Pl-Holant$(g \mid f)$ *is #P-hard otherwise.*

Theorem 6.3 is a special case of Theorem 6.4, with $g = (=_2)$. We show that the respective conditions in item (3) are equivalent when $g = (=_2)$. Suppose $f$ is nondegenerate and satisfies the condition (3) of Theorem 6.3. We will write it in the form (3′), namely $(Z^{-1})^{\otimes 3} f = [x, y, 0, 0]$ or its reversal $[0, 0, y, x]$, for some $x, y \in \mathbb{C}$. Suppose $f = Z^{\otimes 3}[x, y, 0, 0]$; the other case is similar. By being nondegenerate, $y \neq 0$. Using this we can find an upper triangular matrix $R = \begin{bmatrix} 1 & \frac{x-1}{3} \\ 0 & y \end{bmatrix} \in \mathbf{GL}_2(\mathbb{C})$ such that $R^{\otimes 3}[1, 1, 0, 0] = [x, y, 0, 0]$. Then we may take $T = ZR$ for the required matrix in item (3) of Theorem 6.4. Note that for an upper triangular matrix $R$, we have $[0, 1, 0]R^{\otimes 2} = [0, *, *]$. In the other direction, suppose $f$ is nondegenerate and satisfies the condition in item (3) of Theorem 6.4 with $T = \begin{bmatrix} a & c \\ b & d \end{bmatrix}$. The fact that $(=_2)T^{\otimes 2}$ has the form $[0, *, *]$ implies that $a^2 + b^2 = 0$, and so $b = \epsilon a\mathrm{i}$, where $\epsilon = \pm 1$. Since $f$ is nondegenerate, we have $\det T \neq 0$, thus $a \neq 0$ and $d \neq \epsilon c\mathrm{i}$. Using this we can find an upper triangular matrix $R = \begin{bmatrix} a & \frac{c - \epsilon d\mathrm{i}}{2} \\ 0 & \frac{c + \epsilon d\mathrm{i}}{2} \end{bmatrix}$ such that $TR^{-1} = Z = \begin{bmatrix} 1 & 1 \\ \mathrm{i} & -\mathrm{i} \end{bmatrix}$ (for $\epsilon = +1$), or $TR^{-1} = Z' = \begin{bmatrix} 1 & 1 \\ -\mathrm{i} & \mathrm{i} \end{bmatrix}$ (for $\epsilon = -1$). Then $(Z^{-1})^{\otimes 3} f$ has the form $[*, *, 0, 0]$ or $[0, 0, *, *]$.

We can also derive Theorem 6.4 from Theorem 6.3. Suppose $f$ and $g$ are nondegenerate. Write $g = (=_2)M$ as in (2.19) from Chapter 2. Then Holant$(g|f) \equiv_{\mathrm{T}}$ Holant$(=_2| Mf)$. The condition in item (2) of Theorem 6.3 is the existence of some $T$ such that $Mf = T^{\otimes 3}(=_3)$, and $(=_2)T^{\otimes 2}$ is in $\mathscr{A} \cup \mathscr{P}$. It translates to the condition in item (2) of Theorem 6.4 by taking $M^{-1}T$ as the transformation. For item (3), the condition that $Z^{-1}Mf$ has the form $[c, d, 0, 0]$ (where $d \neq 0$ by nondegeneracy) is equivalent to the condition that for some upper triangular $R$, we have $R^{-1}Z^{-1}Mf = [1, 1, 0, 0]$. If so we can let $T = M^{-1}ZR$, then $T^{-1}f = [1, 1, 0, 0]$ and $gT = (=_2)MT = (=_2)ZR = 2(\neq_2)R$ has the form $[0, *, *]$. In the other direction, suppose for some $T$, we have $f = T[1, 1, 0, 0]$ and $gT = [0, a, b]$. By nondegeneracy, $\det T \neq 0$ and $a \neq 0$. Then we may take an upper triangular

https://doi.org/10.1017/9781107477063.007 Published online by Cambridge University Press

$R = \begin{bmatrix} 1 & -b/(2a) \\ 0 & 1 \end{bmatrix}$. Then $[0, a, b]R = a(\neq_2)$, which we can verify by its matrix product form $R^{\mathrm{T}} \begin{bmatrix} 0 & a \\ a & b \end{bmatrix} R = \begin{bmatrix} 0 & a \\ a & 0 \end{bmatrix}$. Then $g = (=_2)M$, with $M$ taken to be a scalar multiple of $ZR^{-1}T^{-1}$, and $\text{Holant}(g|f) \equiv \text{Holant}(=_2|\, Mf)$. With $T^{-1}f = [1, 1, 0, 0]$ and $R^{-1}[1, 1, 0, 0] = [*, *, 0, 0]$, it follows that $Mf$ has the form $Z[*, *, 0, 0]$. We have proved that the condition in item (3) of Theorem 6.4 corresponds precisely to the condition in item (3) of Theorem 6.3 after transforming $g$ to $gM^{-1} = (=_2)$. The same equivalence of Theorem 6.3 and Theorem 6.4 for the planar case can be shown.

The third cases in both Theorems 6.3 and 6.4 are a manifestation of a phenomenon called *vanishing*, which we will discuss in detail in Chapter 7. There is also a symmetric statement to case (3) in Theorem 6.4: $\text{Holant}(g|f)$ is in P, if there is a $2 \times 2$ matrix $T$ such that $f = T^{\otimes 3}[0, 0, 1, 1]$ and $gT^{\otimes 2}$ is of form $[*, *, 0]$. This is equivalent to case (3) in Theorem 6.4 if we use $T \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ as $T$.

## 6.2 Reductions Between Holant and #CSP

In this section, we consider $\text{Holant}(\{g\} \cup \mathcal{G} \mid \{f\} \cup \mathcal{F})$, where $f = [f_0, f_1, f_2, f_3]$ and $g = [g_0, g_1, g_2]$ are nondegenerate, and $f$ is in the generic case of having distinct eigenvalues. We establish an equivalence between this problem and a #CSP problem, under a mild condition, and thus giving a dichotomy via the #CSP dichotomy of Chapter 3. This can be viewed as extending Theorems 6.3 and 6.4 in Section 6.1 for a single ternary signature to a set of signatures containing a nondegenerate ternary signature of the generic case. We will consider the double-root case in Section 6.3.

For $f$ in the generic case, we can apply a holographic reduction to transform $f$ to $(=_3)$. Therefore we need to consider only Holant problems of the form $\text{Holant}(\{g\} \cup \mathcal{G}|\{=_3\} \cup \mathcal{F})$, where $g$ is a nondegenerate symmetric binary signature. Furthermore, we observed after Theorem 6.1 that the ternary EQUALITY signature $(=_3)$ is invariant under any transformation from the following set:

$$\mathscr{T}_3 = \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & \omega \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & \omega^2 \end{bmatrix} \right\}, \tag{6.3}$$

where $\omega = \omega_3 = e^{2\pi \mathrm{i}/3}$. Thus, for any $T \in \mathscr{T}_3$,

$$\text{Holant}(\{g\} \cup \mathcal{G}|\{=_3\} \cup \mathcal{F}) \equiv_{\mathrm{T}} \text{Holant}(\{gT^{\otimes 2}\} \cup \mathcal{G}T|\{=_3\} \cup T^{-1}\mathcal{F}).$$

As a result, we can normalize $g$ by a holographic reduction $T \in \mathscr{T}_3$. We define a symmetric binary signature $g = [g_0, g_1, g_2]$ to be *normalized* if either $g_0 = 0$ or

https://doi.org/10.1017/9781107477063.007 Published online by Cambridge University Press

it is not the case that $g_2$ is $g_0$ times a $t$th primitive root of unity, for some integer $t = 3t'$ where $\gcd(t', 3) = 1$. If $g$ is not normalized, then $g_0 \neq 0$ and $g_2 = g_0\omega_t^s$, where $\omega_t = e^{2\pi \mathrm{i}/t}$, $t = 3t'$, $\gcd(t', 3) = 1$ and $\gcd(s, t) = 1$. Write $1 = 3u + t'v$ for some integers $u$ and $v$, then $\omega_t = \omega^v \omega_{t'}^u$ and $\omega_t^s = \omega^k \omega_{t'}^l$, for some integers $k$ and $l$, where $k \equiv sv \bmod 3$, and $\gcd(k, 3) = 1$. Hence we may take $k = 1$ or 2. After applying the transformation $\begin{bmatrix} 1 & 1 \\ 0 & \omega^k \end{bmatrix} \in \mathscr{T}_3$, the signature $[g_0, g_1, g_2]$ becomes $[g_0, g_1\omega^k, g_0\omega_{t'}^l]$, which is normalized. So in the following, we only deal with normalized $g$. In one case, we also need to normalize a unary signature $[u_0, u_1]$. We say $[u_0, u_1]$ is *normalized* if either $u_0 = 0$ or $u_1$ is not a multiple of $u_0$ by a $t$th primitive root of unity, for some integer $t = 3t'$ where $\gcd(t', 3) = 1$. Again we can normalize the unary signature by a suitable $T \in \mathscr{T}_3$. We note that a normalized signature is still normalized after a nonzero scalar multiple.

**Theorem 6.5.** *Let $\mathcal{F}$ and $\mathcal{G}$ be arbitrary sets of complex-valued signatures. Let $g = [g_0, g_1, g_2]$ be a normalized and nondegenerate symmetric signature. And in the case of $g_0 = g_2 = 0$, we further assume that $\mathcal{G}$ contains a unary signature $[u_0, u_1]$, which is normalized and $u_0u_1 \neq 0$. Then*

$$\text{Holant}(\{g\} \cup \mathcal{G} | \{=_3\} \cup \mathcal{F}) \equiv_{\mathrm{T}} \#\text{CSP}(\{g\} \cup \mathcal{F} \cup \mathcal{G}).$$

*Consequently, by Theorem 3.7,* $\text{Holant}(\{g\} \cup \mathcal{G} | \{=_3\} \cup \mathcal{F})$ *is #P-hard unless $\{g\} \cup \mathcal{F} \cup \mathcal{G} \subseteq \mathscr{P}$ or $\{g\} \cup \mathcal{F} \cup \mathcal{G} \subseteq \mathscr{A}$, in which cases the problem is in P. Furthermore, by Theorem 3.21, if $\mathcal{G}$ contains $\Delta_0$ and $\Delta_1$, then a #P-hardness reduction exists for* $\text{Holant}(\{g\} \cup \mathcal{G} | \{=_3\} \cup \mathcal{F})$ *that does not use Lemma 3.13.*

We prove Theorem 6.5 in the rest of this section. We note that we do not require the signatures in $\mathcal{F}$ and $\mathcal{G}$ to be symmetric. This dichotomy establishes an interesting link between Holant and #CSP. The dichotomy on Holant problems established in Theorem 6.5 uses results from #CSP. The applicability of Theorem 6.5 is not limited by the assumption on signature normalization of $[g_0, g_1, g_2]$; for a nonnormalized binary signature, we can first normalize it and then apply the theorem. The additional assumption of the existence of a nonzero unary signature circumvents a technical difficulty.

One direction of the reduction in Theorem 6.5, from Holant to #CSP, is straightforward by definition. Thus the main claim in Theorem 6.5 is a reduction from #CSP to these bipartite Holant problems. The approach is to construct the binary EQUALITY function $(=_2)$ on the LHS in $\text{Holant}(\{g\} \cup \mathcal{G} | \{=_3\} \cup \mathcal{F})$. As soon as we have $(=_2)$ on LHS, together with $(=_3)$ on the RHS, we get EQUALITY gates of all arities $(=_k)$ on RHS. Also with the help of $(=_2)$ on LHS we can transfer $\mathcal{F}$ to LHS. Then we can simulate all of $\#\text{CSP}(\{g\} \cup \mathcal{F} \cup \mathcal{G})$.

https://doi.org/10.1017/9781107477063.007 Published online by Cambridge University Press

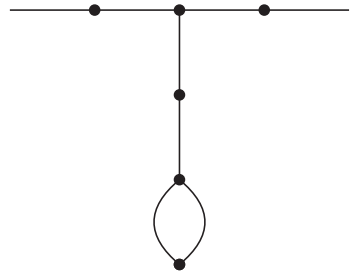

Figure 6.3. A gadget construction for the binary DISEQUALITY function ($\neq_2$) on the LHS. Degree-3 vertices have signature ($=_3$), degree-2 vertices have signature $[1, \pm\mathrm{i}, 1]$.

If the problem Holant($\{g\}|\ =_3$) is already #P-hard, then for any $\mathcal{F}$ and $\mathcal{G}$, the more general problem Holant($\{g\} \cup \mathcal{G}|\{=_3\} \cup \mathcal{F}$) is also #P-hard, and #CSP($\{g\} \cup \mathcal{F} \cup \mathcal{G}$) is reducible to any #P-hard problem. So we only need to consider the case Holant($\{g\}|\ =_3$) is not #P-hard. For this, we again apply Theorem 6.1 to $g = [g_0, g_1, g_2]$. The first tractable case $g_1^2 = g_0 g_2$ is degenerate, which does not apply here. The following three lemmas deal with the remaining three tractable cases respectively. (Note that Holant($\{g\}|\ =_3$) is #P-hard (resp. tractable) iff Holant($\{g\}, \Delta_0, \Delta_1|\ =_3$) is #P-hard (resp. tractable), by Theorem 5.1. But for the latter problem, when it is #P-hard, there is a reduction that does not use Lemma 3.13.)

Let us consider the case $g_0^{12} = g_1^{12}$ and $g_0 g_2 = -g_1^2$. Assume $g_1 \neq 0$, we can scale it to $[a, 1, b]$, where $a^{12} = 1$ and $ab = -1$. Since $[a, 1, b]$ is assumed to be normalized, it follows that $a^4 = 1$.

**Lemma 6.6.** *Let $\mathcal{F}$ and $\mathcal{G}$ be two sets of signatures. For all pairs of $a$ and $b$ satisfying $a^4 = 1$ and $ab = -1$,* Holant($[a, 1, b] \cup \mathcal{G}|\{=_3\} \cup \mathcal{F}$) *is #P-hard unless $\mathcal{F} \cup \mathcal{G} \subseteq \mathscr{A}$, in which case it is in P. Furthermore, if $\mathcal{G}$ contains $\Delta_0$ and $\Delta_1$, then a #P-hardness reduction exists for* Holant($[a, 1, b] \cup \mathcal{G}|\{=_3\} \cup \mathcal{F}$) *that does not use Lemma 3.13.*

*Proof.* We first prove that when $a^4 = 1$ and $ab = -1$,

$$\text{Holant}([a, 1, b] \cup \mathcal{G}|\{=_3\} \cup \mathcal{F}) \equiv_{\mathrm{T}} \#\text{CSP}([a, 1, b] \cup \mathcal{F} \cup \mathcal{G}).$$

To get this, it is sufficient to construct $(=_2)$ on the LHS.

*Case 1*: $a = \pm\mathrm{i}$.
It is equivalent to consider Holant($[1, \pm\mathrm{i}, 1] \cup \mathcal{G}|\{=_3\} \cup \mathcal{F}$) because they differ only by a constant factor. We can construct $[1, 1]$ on the RHS by connecting the two edges of $[1, \pm\mathrm{i}, 1]$ on LHS with two edges of a $(=_3)$ on the RHS. With the gadget in Figure 6.3, we can construct the binary DISEQUALITY function $(\neq_2)$

https://doi.org/10.1017/9781107477063.007 Published online by Cambridge University Press

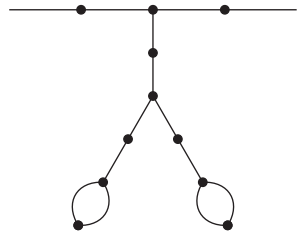

Figure 6.4. Another gadget construction for the binary DISEQUALITY function ($\neq_2$) on the LHS. Degree-3 vertices have signature ($=_3$), degree-2 vertices have signature $[1, \pm 1, -1]$.

on the LHS up to a nonzero scalar. Together with the ($=_3$) on RHS, we can have ($=_3$) on the LHS. Connecting this LHS ($=_3$) with $[1, 1]$ on the RHS, we obtain the binary EQUALITY function ($=_2$) on the LHS.

*Case 2*: $a = \pm 1$.
It is equivalent to consider Holant$([1, \pm 1, -1] \cup \mathcal{G} | \{=_3\} \cup \mathcal{F})$. With the gadget in Figure 6.4, we can construct ($\neq_2$) on the LHS, and thus ($=_3$) on the LHS, using the given ($=_3$) on the RHS. Furthermore, we can construct $[1, -1]$ and $[1, 0, -1]$ on both sides. For example, $[1, -1]$ is constructed as a fragment of the gadget in Figure 6.4; we first obtain $[1, -1]$ on the RHS by connecting two edges of ($=_3$) with a copy of $[1, \pm 1, -1]$, and later obtain $[1, -1]$ on the LHS as well. Using $[1, -1]$ and ($=_3$) on both sides we get $[1, 0, -1]$ on both sides. By connecting $[1, -1]$ with $[1, 0, -1]$, we can realize $[1, 1]$ on both sides, and consequently ($=_2$) on both sides.

By Theorem 3.7, #CSP$([a, 1, b] \cup \mathcal{F} \cup \mathcal{G})$ is #P-hard unless $[a, 1, b] \cup \mathcal{F} \cup \mathcal{G} \subseteq \mathscr{P}$ or $[a, 1, b] \cup \mathcal{F} \cup \mathcal{G} \subseteq \mathscr{A}$. Since $[a, 1, b] \in \mathscr{A} - \mathscr{P}$, we conclude that the only possible case that is not #P-hard is $\mathcal{F} \cup \mathcal{G} \subseteq \mathscr{A}$. This is also sufficient for tractability. The proof is complete. □

Back to Holant$(\{g\} | =_3)$ with $g = [g_0, g_1, g_2]$, for the tractable case $g_1 = 0$ in Theorem 6.1, by nondegeneracy, we can scale it to be $[1, 0, a]$, where $a \neq 0$. Then we have the following lemma:

**Lemma 6.7.** *Let $\mathcal{F}$ and $\mathcal{G}$ be two sets of signatures, and let $a \neq 0$ be a complex number. We assume $[1, 0, a]$ is normalized. Then we have the following dichotomy:*

- *If $a^4 = 1$, then* Holant$([1, 0, a] \cup \mathcal{G} | \{=_3\} \cup \mathcal{F})$ *is #P-hard unless $\mathcal{F} \cup \mathcal{G} \subseteq \mathscr{P}$ or $\mathcal{F} \cup \mathcal{G} \subseteq \mathscr{A}$, in which cases it is in P.*
- *If $a^4 \neq 1$, then* Holant$([1, 0, a] \cup \mathcal{G} | \{=_3\} \cup \mathcal{F})$ *is #P-hard unless $\mathcal{F} \cup \mathcal{G} \subseteq \mathscr{P}$, in which case it is in P.*

https://doi.org/10.1017/9781107477063.007 Published online by Cambridge University Press

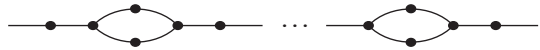

Figure 6.5. A recursive gadget to realize $[1, 0, a^{3k+1}]$ on the LHS. All ternary signatures are $(=_3)$, and all binary signatures are $[1, 0, a]$.

*Furthermore, if $\mathcal{G}$ contains $\Delta_0$ and $\Delta_1$, then a #P-hardness reduction exists that does not use Lemma 3.13.*

*Proof.* As above, it is sufficient to construct $(=_2)$ on the LHS, and establish the equivalence $\text{Holant}([1, 0, a] \cup \mathcal{G}|\{=_3\} \cup \mathcal{F}) \equiv_T \#\text{CSP}([1, 0, a] \cup \mathcal{F} \cup \mathcal{G})$. Note that for $a^4 \neq 1$, the signature $[1, 0, a] \notin \mathscr{A}$ and thus $\#\text{CSP}([1, 0, a] \cup \mathcal{F} \cup \mathcal{G})$ being not #P-hard implies that $\mathcal{F} \cup \mathcal{G} \subseteq \mathscr{P}$.

We will use the gadget in Figure 6.5 in our proof. We can use it to realize $[1, 0, a^{3k+1}]$ for any $k \in \mathbb{N}$ on LHS.

If $a$ is not a root of unity, then we will be able to interpolate all signatures of the form $[1, 0, x]$, where $x \in \mathbb{C}$ on LHS. This uses a Vandermonde system as we did in Chapter 3. In particular, we will be able to interpolate $(=_2) = [1, 0, 1]$ on the LHS.

**Exercise:** If $a$ is not a root of unity, then prove that we can interpolate all signatures of the form $[1, 0, x]$, where $x \in \mathbb{C}$ on the LHS, then complete the proof of Lemma 6.7.

Next we can assume that $a$ is a $t$th primitive root of unity, that is, $a = \omega_t^b$ for some $b$ relatively prime to $t$, where $\omega_t = e^{2\pi \mathrm{i}/t}$. If $t$ is not a multiple of 3, then we can find an integer $k$, such that $3k + 1 \equiv 0 \pmod{t}$. Therefore, we can realize $(=_2) = [1, 0, 1]$ on the LHS.

Now we consider the case of $t = 3^l t'$, where $l \geq 1$ and $\gcd(t', 3) = 1$. Since $[1, 0, a]$ is normalized, we have a further condition that $l > 1$. For this case, we cannot construct $(=_2)$ on the LHS directly. Instead we will further apply a diagonal holographic reduction. Also in this case, we have $a^4 \neq 1$, so we want to prove that $\text{Holant}([1, 0, a] \cup \mathcal{F}|\{=_3\} \cup \mathcal{G})$ is #P-hard unless $\mathcal{F} \cup \mathcal{G} \subseteq \mathscr{P}$. The fact that the problem is in P when $\mathcal{F} \cup \mathcal{G} \subseteq \mathscr{P}$ is obvious by Theorem 3.7, since $[1, 0, a] \in \mathscr{P}$.

Since $2t'$ is not a multiple of 3, there exist some integers $k$ and $s$, such that $3k + 1 = 2st'$. Since

$$a^{3k+1} = a^{2st'} = \omega_{3^l t'}^{2bst'} = \omega_{3^l}^{2bs}$$

we can realize $[1, 0, \omega_{3^l}^{2bs}]$ on the LHS. So

$$\text{Holant}([1, 0, \omega_{3^l}^{2bs}] \cup \mathcal{G}|\{=_3\} \cup \mathcal{F}) \leq_T \text{Holant}([1, 0, a] \cup \mathcal{G}|\{=_3\} \cup \mathcal{F}).$$

https://doi.org/10.1017/9781107477063.007 Published online by Cambridge University Press

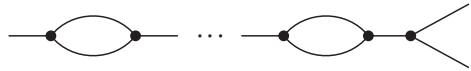

Figure 6.6. A recursive gadget to realize $[1, 0, 0, \omega_{3^{(l-1)}}^{mbs}]$ where $m$ is any odd integer. All ternary signatures are $[1, 0, 0, \omega_{3^{(l-1)}}^{bs}]$. Edges have $(=_2)$.

Therefore it suffices to prove that Holant$([1, 0, \omega_{3^l}^{2bs}] \cup \mathcal{G}|\{=_3\} \cup \mathcal{F})$ is #P-hard if $\mathcal{F} \cup \mathcal{G} \not\subseteq \mathscr{P}$.

We apply a diagonal holographic reduction $T = \begin{bmatrix} 1 & 0 \\ 0 & \omega_{3^l}^{-bs} \end{bmatrix}$, and get

$$\begin{aligned} &\text{Holant}\big(\big[1, 0, \omega_{3^l}^{2bs}\big] \cup \mathcal{G}|\{=_3\} \cup \mathcal{F}\big) \\ &\quad \equiv_{\text{T}} \text{Holant}\big(\{=_2\} \cup \mathcal{G}T|\big[1, 0, 0, \omega_{3^{(l-1)}}^{bs}\big] \cup T^{-1}\mathcal{F}\big). \end{aligned}$$

We then use the gadget in Figure 6.6 to realize $[1, 0, 0, \omega_{3^{(l-1)}}^{3^{l-1}bs}] = [1, 0, 0, 1] = (=_3)$ on the RHS. Together with $(=_2)$ on the LHS this gives all EQUALITIES on both sides. As a result,

$$\begin{aligned} &\#\text{CSP}\big(\big[1, 0, 0, \omega_{3^{(l-1)}}^{bs}\big] \cup \mathcal{G}T \cup T^{-1}\mathcal{F}\big) \\ &\quad \leq_{\text{T}} \text{Holant}\big(\{=_2\} \cup \mathcal{G}T|\big[1, 0, 0, \omega_{3^{(l-1)}}^{bs}\big] \cup T^{-1}\mathcal{F}\big). \end{aligned}$$

Since $l > 1$, $[1, 0, 0, \omega_{3^{(l-1)}}^{bs}] \notin \mathscr{A}$, as both $b$ and $s$ are relatively prime to 3. So the problem is #P-hard unless $\mathcal{G}T \cup T^{-1}\mathcal{F} \subseteq \mathscr{P}$. Since $T$ and $T^{-1}$ are diagonal matrices, it is equivalent to the statement $\mathcal{F} \cup \mathcal{G} \subseteq \mathscr{P}$. Note that diagonal transformations preserve $\{\Delta_0, \Delta_1\}$. If $\mathcal{G}$ contains $\Delta_0$ and $\Delta_1$, then a #P-hardness reduction exists that does not use Lemma 3.13. This completes the proof of Lemma 6.7. □

For the last tractable case $g_0 = g_2 = 0$ in Theorem 6.1, we can scale it to $(\neq_2) = [0, 1, 0]$.

**Lemma 6.8.** *Let $\mathcal{F}$ and $\mathcal{G}$ be two sets of signatures, and $a \neq 0$ be a complex number. We assume $[1, a]$ is normalized. Then we have the following dichotomy:*

- *If $a^4 = 1$, then* Holant$(\{[0, 1, 0], [1, a]\} \cup \mathcal{G}|\{=_3\} \cup \mathcal{F})$ *is #P-hard unless $\mathcal{F} \cup \mathcal{G} \subseteq \mathscr{P}$ or $\mathcal{F} \cup \mathcal{G} \subseteq \mathscr{A}$, in which cases it is in P.*
- *If $a^4 \neq 1$, then* Holant$(\{[0, 1, 0], [1, a]\} \cup \mathcal{G}|\{=_3\} \cup \mathcal{F})$ *is #P-hard unless $\mathcal{F} \cup \mathcal{G} \subseteq \mathscr{P}$, in which case it is in P.*

*Furthermore, if $\mathcal{G}$ contains $\Delta_0$ and $\Delta_1$, then a #P-hardness reduction exists for* Holant$([a, 1, b] \cup \mathcal{G}|\{=_3\} \cup \mathcal{F})$ *that does not use Lemma 3.13.*

https://doi.org/10.1017/9781107477063.007 Published online by Cambridge University Press

*Proof.* By connecting one copy of $[1, a]$ and two copies of $[0, 1, 0]$ to a $(=_3)$ on the RHS, we can realize $[a, 0, 1]$, or equivalently $[1, 0, 1/a]$, on the LHS. Now Lemma 6.8 follows from Lemma 6.7. □

This completes the proof of Theorem 6.5.

## 6.3 Holant$^c$ Problems

In Holant$^*$ problems, all unary functions are assumed to be freely available. Two special unary functions, Is-Zero $\Delta_0 = [1, 0]$ and Is-One $\Delta_1 = [0, 1]$, are particularly meaningful. They correspond to setting a variable to the constant 0 or 1. This is called pinning, and by Lemma 3.13 these are available for all #CSP problems, namely $\Delta_0$ and $\Delta_1$ are implied by the presence of Equalities of all arities. Note that if we have $\Delta_0$ and $\Delta_1$, then we can construct all subsignatures of any given signature.

We define a class of restricted Holant problems where only the two special unary functions $\Delta_0$ and $\Delta_1$ are assumed to be freely available. This class is called the Holant$^c$ problems.

**Definition 6.9.** Given a set of signatures $\mathcal{F}$, we use Holant$^c(\mathcal{F})$ to denote Holant$(\mathcal{F} \cup \{\Delta_0, \Delta_1\})$.

In this section, we prove a dichotomy theorem for Holant$^c$ problems with complex-valued symmetric signatures, Theorem 6.12. The proof uses the dichotomies proved in the previous two sections. To use them, we first prove in Lemma 6.10 that we can realize a nondegenerate ternary symmetric signature except in some trivial cases. From this ternary signature we immediately have #P-hardness if it is not in one of the tractable cases in Theorem 6.3. For a tractable ternary signature, we use Theorem 6.5 to extend the dichotomy theorem to a set of signatures. In Theorem 6.5, we considered only the *generic* case of the ternary function. The double-root case is handled here in Lemma 6.11. After that we will be ready to prove Theorem 6.12.

**Lemma 6.10.** *Given any set of symmetric signatures $\mathcal{F}$ which contains $\Delta_0$ and $\Delta_1$, we can construct a nondegenerate ternary symmetric signature $f$, except in the following two cases:*

1. *All nondegenerate signatures in $\mathcal{F}$ have arity $\leq 2$.*
2. *In $\mathcal{F}$, all unary signatures are of the form $[x, 0]$ or $[0, x]$; all binary signatures are of the form $[x, 0, y]$ or $[0, x, 0]$; and all signatures of arity $> 2$ are of the form $[x, 0, \ldots, 0, y]$.*

https://doi.org/10.1017/9781107477063.007 Published online by Cambridge University Press

*Proof.* Suppose case 1 does not hold, and let $f = [f_0, f_1, \ldots, f_n] \in \mathcal{F}$ be a nondegenerate signature of arity $\geq 3$. Since we have $\Delta_0$ and $\Delta_1$, we can construct all subsignatures of any signature in $\mathcal{F}$. (A signature is also considered a subsignature of itself.) If there exists a ternary nondegenerate subsignature, we are done. Now suppose all ternary subsignatures of $f$ are degenerate, in particular $n > 3$. By Proposition 2.8, $f$ must be of the form $[f_0, 0, \ldots, 0, f_n]$, where $f_0 f_n \neq 0$. Then, if we have a unary signature or a unary subsignature of the form $[a, b]$ from $\mathcal{F}$ with $ab \neq 0$, we can connect $n-3$ copies of $[a, b]$ to $n-3$ dangling edges of $f$ to get a nondegenerate ternary signature of the form $[x, 0, 0, y]$, and we are done. Otherwise, all unary signatures are of the form $[x, 0]$ or $[0, x]$ and all binary signatures are of the form $[x, 0, y]$ or $[0, x, 0]$. Any nondegenerate signature of arity $> 2$ must be of the form $[x, 0, \ldots, 0, y]$. If a degenerate signature $[a, b]^{\otimes n}$ has $ab \neq 0$, then we can get a unary subsignature of the form $[a', b']$ with $a'b' \neq 0$. Thus the degenerate signature is also a Gen-Eq of the form $[x, 0, \ldots, 0, y]$, where $x = 0$ or $y = 0$. □

**Remark:** All the exceptional cases listed in case 2 are special cases of functions of product type.

We next consider Holant($\{f, g\}$), where $f$ and $g$ are nondegenerate symmetric signatures of arity 3 and 2 respectively, and the characteristic equation for the second-order linear recurrence of $f = [f_0, f_1, f_2, f_3]$ has a double root. By Lemma 6.2, Holant($f$) is already #P-hard unless the double eigenvalue is $\mathrm{i}$ or $-\mathrm{i}$. Therefore we may assume the double root is $\lambda = \pm\mathrm{i}$, and $f_{k+2} - \lambda f_{k+1} - f_k = 0$ for $k = 0, 1$.

**Lemma 6.11.** *Let $f = [f_0, f_1, f_2, f_3]$ be a nondegenerate complex signature satisfying $f_{k+2} - \lambda f_{k+1} - f_k = 0$ for $k = 0, 1$, where $\lambda = \pm 2\mathrm{i}$. Let $g = [g_0, g_1, g_2]$ be a nondegenerate binary signature. Then Holant($g \mid f$) is #P-hard, and a fortiori Holant($\{f, g\}$) is also #P-hard, unless $g_2 - \lambda g_1 - g_0 = 0$; if so, the problem Holant($\{f, g\}$) is in P by generalized Fibonacci gates.*

*Proof.* We prove this result for $\lambda = 2\mathrm{i}$. The other case is similar.

The sequence $\{f_k\}$ can be written as follows: $f_k = ck\mathrm{i}^{k-1} + d\mathrm{i}^k$, where $c \neq 0$. Thus, we have

$$f = T^{\otimes 3}[1, 1, 0, 0]^{\mathrm{T}}, \quad \text{where} \quad T = \begin{bmatrix} 1 & \frac{d-1}{3} \\ \mathrm{i} & c + \frac{d-1}{3}\mathrm{i} \end{bmatrix}.$$

https://doi.org/10.1017/9781107477063.007 Published online by Cambridge University Press

We can express $\begin{bmatrix} g_0 & g_1 \\ g_1 & g_2 \end{bmatrix} = M^{\mathrm{T}}M$, as shown in (2.19) of Section 2.3.2, for some $M = \begin{bmatrix} \alpha & \gamma \\ \beta & \delta \end{bmatrix}$. $M$ is nonsingular since $g$ is nondegenerate. Then we have

$$g_0 = \alpha^2 + \beta^2, \quad g_1 = \alpha\gamma + \beta\delta, \quad \text{and} \quad g_2 = \gamma^2 + \delta^2.$$

Thus we have the following reduction:

$$\begin{aligned} \text{Holant}(g \mid f) &= \text{Holant}((=_2)M^{\otimes 2} \mid T^{\otimes 3}[1, 1, 0, 0]) \\ &\equiv_{\mathrm{T}} \text{Holant}(=_2 \mid (MT)^{\otimes 3}[1, 1, 0, 0]), \end{aligned}$$

where

$$MT = \begin{bmatrix} \alpha + \gamma\mathrm{i} & * \\ \beta + \delta\mathrm{i} & * \end{bmatrix}, \quad \text{and we will call it} \quad \begin{bmatrix} p & r \\ q & s \end{bmatrix}.$$

Our next step is to use an orthogonal matrix to transform $MT$ to be upper-triangular, if we can. If $p^2 + q^2 \neq 0$, then we can find a (complex) orthogonal matrix $Q = \frac{1}{\sqrt{p^2+q^2}} \begin{bmatrix} p & q \\ q & -p \end{bmatrix}$ such that $QMT$ is upper-triangular. As shown in Section 6.1, for a nonsingular upper-triangular $QMT$, $(QMT)^{\otimes 3}[1, 1, 0, 0]$ has the form $[x, y, 0, 0]$ for some $y \neq 0$. We can normalize it to $[z, 1, 0, 0]$.

For any orthogonal matrix $Q$, the LHS $(=_2)$ is unchanged under the holographic transformation: $(=_2)(Q^{-1})^{\otimes 2} = (=_2)$. This gives us

$$\begin{aligned} \text{Holant}((=_2) \mid (MT)^{\otimes 3}[1, 1, 0, 0]) &\equiv_{\mathrm{T}} \text{Holant}((=_2) \mid [z, 1, 0, 0]) \\ &\equiv_{\mathrm{T}} \text{Holant}([z, 1, 0, 0]), \end{aligned}$$

for some $z$. This shows the equivalence of $\text{Holant}(g \mid f)$ with $\text{Holant}([z, 1, 0, 0])$. By Lemma 6.2, $\text{Holant}([z, 1, 0, 0])$ is #P-hard. Hence $\text{Holant}(g \mid f)$ is also #P-hard.

Finally, $p^2 + q^2 = 0$ is $(\alpha + \gamma\mathrm{i})^2 + (\beta + \delta\mathrm{i})^2 = (\alpha^2 + \beta^2) + 2\mathrm{i}(\alpha\gamma + \beta\delta) - (\gamma^2 + \delta^2) = 0$. This is exactly $g_2 - \lambda g_1 - g_0 = 0$. □

**Exercise:** Show that in any instance $\Omega$ of $\text{Holant}(\{f, g\})$ if $f$ appears in the signature grid, then the value $\text{Holant}_\Omega = 0$.

A perceptive reader will have noticed that the very explicit dichotomy statement in Lemma 6.11 should be consistent with, and derivable from, the more conceptual dichotomy statement of Theorem 6.4. We now give such a derivation.

As the discussion after Theorems 6.3 and 6.4 shows, the assumption on $f$ here is equivalent to $f$ having the form $Z[c, d, 0, 0]$, or $Z[0, 0, d, c]$, with $d \neq 0$

https://doi.org/10.1017/9781107477063.007 Published online by Cambridge University Press

by nondegeneracy. We will assume the first case. The second case is similar. We have

$$\text{Holant}(g \mid f) = \text{Holant}((=_2)M \mid Z[c, d, 0, 0]) \equiv_{\text{T}} \text{Holant}(=_2 \mid MZ[c, d, 0, 0]),$$

where $g = (=_2)M$ as shown in (2.19) for some $M = \begin{bmatrix} \alpha & \gamma \\ \beta & \delta \end{bmatrix}$. The property that a ternary symmetric signature is nondegenerate, or is in the generic case (having distinct eigenvalues) is invariant under a nonsingular holographic transformation. This is easily seen in terms of the tensor product expression. Hence the property that it has a double root is also invariant. It follows that, by Theorem 6.4 the only possibility for Holant$(=_2 \mid MZ[c, d, 0, 0])$ to be tractable is when $MZ[c, d, 0, 0]$ has the form $Z[*, *, 0, 0, ]$ or $Z[0, 0, *, *]$.

The following fact can be directly verified: For any nonsingular matrix $T$ and $d \neq 0$, $T[c, d, 0, 0] = [*, *, 0, 0, ]$ iff $T$ is an upper-triangular matrix. Hence the tractability criterion from Theorem 6.4 is $Z^{-1}MZ$ or $XZ^{-1}MZ$ is upper-triangular, where $X = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$. Since

$$\begin{bmatrix} 1 & -\mathrm{i} \\ 1 & \mathrm{i} \end{bmatrix} \begin{bmatrix} \alpha & \gamma \\ \beta & \delta \end{bmatrix} \begin{bmatrix} 1 & 1 \\ \mathrm{i} & -\mathrm{i} \end{bmatrix} = \begin{bmatrix} 1 & -\mathrm{i} \\ 1 & \mathrm{i} \end{bmatrix} \begin{bmatrix} \alpha + \mathrm{i}\gamma & \alpha - \mathrm{i}\gamma \\ \beta + \mathrm{i}\delta & \beta - \mathrm{i}\delta \end{bmatrix}$$
$$= \begin{bmatrix} (\alpha + \mathrm{i}\gamma) - \mathrm{i}(\beta + \mathrm{i}\delta) & * \\ (\alpha + \mathrm{i}\gamma) + \mathrm{i}(\beta + \mathrm{i}\delta) & * \end{bmatrix},$$

the two tractable cases can be expressed as

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

This is just $(\alpha + \mathrm{i}\gamma)^2 + (\beta + \mathrm{i}\delta)^2 = 0$, and it matches exactly with $g_2 - \lambda g_1 - g_0 = 0$.

The next theorem is the main result of this section. It gives a dichotomy for Holant$^c$ problems. Define

$$\mathscr{T} = \{T \in \mathbf{GL}_2(\mathbb{C}) \mid (=_2)T^{\otimes 2}, \Delta_0 T, \Delta_1 T \in \mathscr{A}\}. \tag{6.4}$$

**Theorem 6.12.** *Let $\mathcal{F}$ be a set of complex symmetric signatures.* Holant$^c(\mathcal{F})$ *is #P-hard unless $\mathcal{F}$ satisfies one of the following conditions, in which case it is tractable:*

1. Holant$^*(\mathcal{F})$ *is tractable; or*
2. *There exists a $T \in \mathscr{T}$ such that $T^{-1}\mathcal{F} \subseteq \mathscr{A}$.*

**Remark:** As we have $\Delta_0$ and $\Delta_1$, any degenerate $f = [a, b]^{\otimes n} \in \mathcal{F}$ can be replaced by $[a, b]$, as follows: If $a = b = 0$ then $f$ is identically 0 and applying

https://doi.org/10.1017/9781107477063.007 Published online by Cambridge University Press

$\Delta_0$ to the signature $n-1$ times gives the unary $[0,0]$. If $a, b$ are not both 0, then using $\Delta_0$ or $\Delta_1$ we can obtain a nonzero multiple of $[a,b]$. Thus we may assume all signatures of arity $\geq 2$ in $\mathcal{F}$ are nondegenerate. Also, since we deal with symmetric signatures, by Theorem 3.8, the affine class $\mathscr{A}$ can be replaced by $\mathscr{F}_1 \cup \mathscr{F}_2 \cup \mathscr{F}_3$ in the statement of Theorem 6.12.

The dichotomy criterion in Theorem 6.12 can be made explicit. For case 1 we have the very explicit dichotomy for Holant$^*$ problems – Theorem 2.12. This also has a more conceptual but less explicit form in Theorem 3.5. For case 2 we will give a finite list of $T \in \mathbf{GL}_2(\mathbb{C})$ in place of $\mathscr{T}$. We will discuss this in more detail after the proof.

*Proof.* First of all, if $\mathcal{F}$ is in one of the two listed cases of Lemma 6.10, then Holant$^*(\mathcal{F})$ is tractable and we are done. Now we can assume that we have a nondegenerate symmetric ternary signature $f = [f_0, f_1, f_2, f_3]$ and the problem is Holant$^c(\mathcal{F} \cup \{f\})$.

As discussed in Section 6.1, there are three categories for $f$ and we only need to consider the first two:

1. $f_k = \alpha^{3-k}\beta^k + \gamma^{3-k}\delta^k$, where $\det \begin{bmatrix} \alpha & \gamma \\ \beta & \delta \end{bmatrix} \neq 0$;
2. $f_k = ck\lambda^{k-1} + d\lambda^k$, where $c \neq 0$.

*Case 1:* $f_k = \alpha^{3-k}\beta^k + \gamma^{3-k}\delta^k$.

In this case, $f = T^{\otimes 3}(=_3)$, where $T = \begin{bmatrix} \alpha & \gamma \\ \beta & \delta \end{bmatrix}$. Furthermore, we will choose such a $T$ so that $(=_2)T^{\otimes 2}$ is normalized. We may accomplish this by replacing $T$ with $TT'$ for some $T' \in \mathscr{T}_3$ defined in (6.3), as explained in Section 6.2. Then we have the following reduction chain:

$$\begin{aligned}
\text{Holant}^c(\mathcal{F}) &\equiv_{\mathrm{T}} \text{Holant}^c(\mathcal{F} \cup \{f\}) \\
&\equiv_{\mathrm{T}} \text{Holant}(\mathcal{F} \cup \{f, \Delta_0, \Delta_1\}) \\
&\equiv_{\mathrm{T}} \text{Holant}((=_2) \mid \mathcal{F} \cup \{f, \Delta_0, \Delta_1\}) \\
&\equiv_{\mathrm{T}} \text{Holant}(\{(=_2), \Delta_0, \Delta_1\} \mid \mathcal{F} \cup \{f\}) \\
&\equiv_{\mathrm{T}} \text{Holant}(\{(=_2)T^{\otimes 2}, \Delta_0 T, \Delta_1 T\} \mid \{=_3\} \cup T^{-1}\mathcal{F}).
\end{aligned}$$

Note that $(=_2)$ on the LHS connected with $\Delta_c$ on the RHS gives $\Delta_c$ on the LHS, for $c = 0$ and 1. Since $(=_2)T^{\otimes 2}$ is a normalized nondegenerate binary signature, we can apply Theorem 6.5. The only thing we need to verify is that in case $(=_2)T^{\otimes 2} = [\alpha^2 + \beta^2, \alpha\gamma + \beta\delta, \gamma^2 + \delta^2] = [0, \alpha\gamma + \beta\delta, 0]$, at least one of $\Delta_0 T = [1,0]T = [\alpha, \gamma]$ or $\Delta_1 T = [0,1]T = [\beta, \delta]$ has both

https://doi.org/10.1017/9781107477063.007 Published online by Cambridge University Press

entries nonzero. If not, we would have $\alpha\gamma = 0$ and $\beta\delta = 0$, which implies that $(=_2)T^{\otimes 2} = [0, \alpha\gamma + \beta\delta, 0] = [0, 0, 0]$, a contradiction. In the case of $(=_2)T^{\otimes 2} = [0, *, 0]$, we may further normalize the unary signature $\Delta_0 T$ or $\Delta_1 T$ without affecting $(=_2)T^{\otimes 2}$ being normalized. Therefore, by Theorem 6.5, the problem Holant$^c(\mathcal{F})$ is #P-hard unless

$$\{(=_2)T^{\otimes 2}\} \cup T^{-1}\mathcal{F} \subseteq \mathscr{P}$$

(note that unary $\Delta_0 T$, $\Delta_1 T$ are automatically in $\mathscr{P}$), or

$$\{(=_2)T^{\otimes 2}, \Delta_0 T, \Delta_1 T\} \cup T^{-1}\mathcal{F} \subseteq \mathscr{A}.$$

In the first case, Holant$^*(\mathcal{F})$ is tractable, by the dichotomy theorem for Holant$^*$ problems, Theorem 2.12 (see also Theorem 3.5). In the second case, this is equivalent to having $T \in \mathscr{T}$ satisfying $T^{-1}\mathcal{F} \subseteq \mathscr{A}$.

*Case 2: $f_k = ck\lambda^{k-1} + d\lambda^k$, where $c \neq 0$.*

In this case, if $\lambda \neq \pm \mathrm{i}$, the problem is #P-hard by Lemma 6.2 and we are done. Now we consider the case $\lambda = \mathrm{i}$ (the case $\lambda = -\mathrm{i}$ is similar). Consider the following equation:

$$z_{k+2} - 2\mathrm{i}z_{k+1} - z_k = 0. \tag{6.5}$$

We note that $f = [f_0, f_1, f_2, f_3]$ satisfies this equation for $k = 0, 1$. If for all nondegenerate signatures $\Xi = [z_0, z_1, \ldots, z_n]$ in $\mathcal{F}$ with arity $n \geq 2$ we have the following

**Condition**: $\Xi$ satisfies Equation (6.5) for $0 \leq k \leq n-2$, then, by Theorem 2.12 tractable Class $\mathfrak{B}$ type I$(1, 2\mathrm{i})$ (equivalently case 3 of Theorem 3.5), Holant$^*(\mathcal{F})$ is tractable, and we are done. So suppose this is not the case, and there is a nondegenerate signature $\Xi = [z_0, z_1, \ldots, z_n] \in \mathcal{F}$, for some $n \geq 2$, that does not satisfy the displayed condition. By Lemma 6.11, if any nondegenerate subsignature $[z_k, z_{k+1}, z_{k+2}]$ does not satisfy (6.5), then, together with $f$ which does satisfy (6.5), we know that the problem is #P-hard and we are done. So we assume every nondegenerate subsignature $[z_k, z_{k+1}, z_{k+2}]$ of $\Xi$ satisfies (6.5). As $\Xi$ is nondegenerate and does not satisfy the displayed condition, we have $n \geq 3$. Observe that $\Xi$ does not satisfy the displayed condition is equivalent to some binary subsignature of $\Xi$ does not satisfy (6.5). Hence there exists some binary subsignature of $\Xi$ that is *degenerate* and does not satisfy (6.5).

*Subcase 1:* All binary subsignatures of $\Xi$ are degenerate (but $\Xi$ itself is nondegenerate).

By Proposition 2.8, $\Xi$ must be of the form $[z_0, 0, \ldots, 0, z_n]$, where $z_0 z_n \neq 0$.

https://doi.org/10.1017/9781107477063.007 Published online by Cambridge University Press

We claim that there exists a unary subsignature $[f_k, f_{k+1}]$ of $f$ with both entries nonzero. If $f_0 f_1 \neq 0$, then we set $k = 0$ and the claim is proved; if $f_0 = 0$, $f_1 \neq 0$, then we have $f_2 = 2\mathrm{i} f_1 + f_0 \neq 0$; if $f_0 \neq 0$, $f_1 = 0$, then we have $f_2 = 2\mathrm{i} f_1 + f_0 \neq 0$ and $f_3 = 2\mathrm{i} f_2 + f_1 \neq 0$. Note that $f_0 = f_1 = 0$ is impossible because $c \neq 0$.

Connect this unary signature to $n - 3$ dangling edges of $[z_0, 0, \ldots, 0, z_n]$, we have a ternary signature $[a, 0, 0, b]$, where $ab \neq 0$. We can use this as the nondegenerate ternary signature $f$ and we have reduced this case to case 1.

*Subcase 2:* Some binary subsignatures of $\Xi$ are nondegenerate (some others are degenerate).

Then we can find a ternary subsignature $[z_k, z_{k+1}, z_{k+2}, z_{k+3}]$ (or its reversal) where $[z_k, z_{k+1}, z_{k+2}]$ is degenerate and $[z_{k+1}, z_{k+2}, z_{k+3}]$ is nondegenerate and thus satisfies $z_{k+3} - 2\mathrm{i} z_{k+2} - z_{k+1} = 0$. We have

$$z_{k+1}^2 = z_k z_{k+2} \quad \text{and} \quad z_{k+2}^2 \neq z_{k+1} z_{k+3}. \tag{6.6}$$

Thus $z_{k+1}$ and $z_{k+2}$ cannot both be zero. First suppose $z_k \neq 0$. Then $z_{k+1} \neq 0$, for otherwise $z_{k+1} = z_{k+2} = 0$ by $z_{k+1}^2 = z_k z_{k+2}$. Then $(z_k, z_{k+1}, z_{k+2})$ is a geometric progression with a nonzero ratio $\rho = z_{k+1}/z_k$, and we can assume that $z_k = 1$ after a scaling. Then $[z_k, z_{k+1}, z_{k+2}, z_{k+3}] = [1, \rho, \rho^2, 2\mathrm{i}\rho^2 + \rho]$, and we have

$$\begin{bmatrix} 1 & 0 \\ \rho & (\rho + 2\mathrm{i}\rho^2 - \rho^3)^{1/3} \end{bmatrix}^{\otimes 3} [1, 0, 0, 1] = \begin{bmatrix} 1 \\ \rho \end{bmatrix}^{\otimes 3} + (\rho + 2\mathrm{i}\rho^2 - \rho^3) \begin{bmatrix} 0 \\ 1 \end{bmatrix}^{\otimes 3}$$
$$= [1, \rho, \rho^2, 2\mathrm{i}\rho^2 + \rho].$$

Note that $\rho + 2\mathrm{i}\rho^2 - \rho^3 \neq 0$ so that the $2 \times 2$ matrix is nonsingular. We already know that $\rho \neq 0$. If $1 + 2\mathrm{i}\rho - \rho^2 = 0$ then $\rho = \mathrm{i}$, which would imply that $[z_k, z_{k+1}, z_{k+2}, z_{k+3}] = [1, \mathrm{i}, -1, -\mathrm{i}]$ is degenerate, a contradiction. Therefore, the ternary subsignature $[z_k, z_{k+1}, z_{k+2}, z_{k+3}]$ is in the first category and we have reduced the problem to case 1.

Now suppose $z_k = 0$. Then it must be that $z_{k+1} = 0$ and $z_{k+2} \neq 0$ by (6.6). The signature $[z_k, z_{k+1}, z_{k+2}, z_{k+3}]$ becomes $[0, 0, z_{k+2}, z_{k+3}]$, which is equivalent to $[0, 0, 1, z]$ for some $z \in \mathbb{C}$, and as we proved in Lemma 6.2, it is #P-hard. □

There is a different character to this Holant$^c$ dichotomy, as stated, from previous dichotomies we have proved so far. For example, Theorem 2.12 for Holant$^*$ problems, or Theorem 3.7 for Boolean #CSP are very *explicit*, so that for any given finite signature set $\mathcal{F}$, it is clear one can decide whether $\mathcal{F}$ satisfies the tractability criteria, in time polynomial in the size of $\mathcal{F}$. In fact, to

https://doi.org/10.1017/9781107477063.007 Published online by Cambridge University Press

say that the tractability condition is decidable in polynomial time does not do full justice to the explicitness of these dichotomy theorems; their tractability conditions are expressed essentially in closed form. This level of explicitness is useful, beyond the important issue of decidability in formal logic, or even polynomial time decidability, of the dichotomy criteria. It gives us a most concrete understanding of the dichotomy, as well as it can be used effectively to prove other dichotomy theorems, as shown, for example, when we applied Theorem 2.12 and Theorem 3.7.

Theorem 6.12 for Holant$^c$ problems has a more conceptual form. It has an *existential* statement referring to the set $\mathscr{T}$. The conceptual form provides a better understanding of the reason why the dichotomy works. But it brings up the question of how one may decide whether a given $\mathcal{F}$ satisfies the tractability criteria of Theorem 6.12. One should note that for Holant$^*$ problems, Theorem 2.12 has an equivalent form Theorem 3.5, which also has such an existential statement. Our next goal is to show that the Holant$^c$ dichotomy Theorem 6.12 in fact also has a very explicit form. Denote by $\mathscr{F}_{123} = \mathscr{F}_1 \cup \mathscr{F}_2 \cup \mathscr{F}_3$. We will show that the following set

$$\left\{ I, \begin{bmatrix} 1 & 0 \\ 0 & \sqrt{\mathrm{i}} \end{bmatrix}, \begin{bmatrix} \sqrt{2} & 0 \\ \mathrm{i} & 1 \end{bmatrix}, \begin{bmatrix} \sqrt{2} & 0 \\ -\mathrm{i} & 1 \end{bmatrix} \right\} \tag{6.7}$$

can be used in place of the infinite set $\mathscr{T}$ in criterion 2 of Theorem 6.12, namely whether for one of these four matrices $T$, we have $T^{-1}\mathcal{F} \subseteq \mathscr{F}_{123}$. For convenience, we denote these four matrices by $I, A, B, C$ respectively. We verify that they all belong to $\mathscr{T}$. Trivially $I \in \mathscr{T}$. For $A$, we have $(=_2)A^{\otimes 2} = [1, 0, \mathrm{i}] \in \mathscr{F}_1$. Also we have $\Delta_0 A = [1, 0]$ and $\Delta_1 A = \sqrt{\mathrm{i}}[0, 1]$; thus $A \in \mathscr{T}$. For $B$, we have $(=_2)B^{\otimes 2} = [1, \mathrm{i}, 1] \in \mathscr{F}_2$, which has matrix form

$$\begin{bmatrix} \sqrt{2} & \mathrm{i} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \sqrt{2} & 0 \\ \mathrm{i} & 1 \end{bmatrix} = \begin{bmatrix} 1 & \mathrm{i} \\ \mathrm{i} & 1 \end{bmatrix}.$$

Also we have $\Delta_0 B = \sqrt{2}[1, 0]$ and $\Delta_1 B = \mathrm{i}[1, -\mathrm{i}] \in \mathscr{F}_2$; thus $B \in \mathscr{T}$. For $C$, we can take conjugation in the above calculations for $B$.

The key to this simplification of replacing (6.7) for the infinite set $\mathscr{T}$ is a normalizing process afforded by the following group action by $\mathbf{G} = \mathbf{GL}_2(\mathbb{C})$ and a certain stabilizer subgroup. For signatures $f$ written as covaraint tensors, we consider the group action $(f, T) \mapsto fT$ by $\mathbf{G}$, where $T \in \mathbf{G}$. There is a corresponding group action $(T, f) \mapsto T^{-1}f$ by $\mathbf{G}$ on signatures written as contravariant tensors. We wish to investigate those $T \in \mathbf{G}$ such that

$$(=_2)T,\ \Delta_0 T,\ \Delta_1 T \in \mathscr{F}_{123}, \quad \text{and} \quad T^{-1}\mathcal{F} \subseteq \mathscr{F}_{123}. \tag{6.8}$$

https://doi.org/10.1017/9781107477063.007 Published online by Cambridge University Press

Let

$$\mathbf{S} = \{S \in \mathbf{G} \mid \mathscr{F}_{123} S \subseteq \mathscr{F}_{123}\} \tag{6.9}$$

be the stabilizer of $\mathscr{F}_{123}$. Clearly $\mathbf{S}$ is a subgroup of $\mathbf{G}$. In fact, by definitions (6.9) and (6.4), it is clear that $\mathbf{S} \subseteq \mathscr{T}$, since $(=_2), \Delta_0, \Delta_1 \in \mathscr{F}_{123}$. The purpose of defining this subgroup $\mathbf{S}$ is that we will use multiplication on the right from $\mathbf{S}$ to normalize the set of transformations.

Let $D = \begin{bmatrix} 1 & 0 \\ 0 & \mathrm{i} \end{bmatrix}$ and $H = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$. Also let $X = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ and $Z = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ \mathrm{i} & -\mathrm{i} \end{bmatrix}$. Note that $Z = DH$. Also $D^2 Z = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ -i & i \end{bmatrix} = ZX$; hence $X = Z^{-1} D^2 Z$. It is easy to verify that $D, H \in \mathbf{S}$, and hence also $X, Z \in \mathbf{S}$, as well as all nonzero scalar multiples of these. We will show that $\mathbf{S}$ consists of exactly the nonzero scalar multiples of members of the group $\langle D, H \rangle$ generated by $D$ and $H$,

$$\mathbf{S} = \mathbb{C}^* \cdot \langle D, H \rangle. \tag{6.10}$$

Note that $X, Z$ are also in the group $\langle D, H \rangle$ generated by $D$ and $H$.

To show that $\mathbf{S}$ is contained in the RHS of (6.10), let $T \in \mathbf{S}$. Take $(=_3) \in \mathscr{F}_{123}$, then $(=_3) T^{\otimes 3} \in \mathscr{F}_{123}$. Then by the form of $\mathscr{F}_{123}$, for some $M \in \langle D, H \rangle$, chosen to be either $I$, or $H^{\mathrm{T}} = H$, or $Z^{\mathrm{T}} = HD$, we have $(=_3)(TM^{-1})^{\otimes 3} \in \mathscr{F}_1$, which is a Gen-Eq. Then either $T' = TM^{-1}$ or $TM^{-1}X$ is a diagonal matrix $\lambda \begin{bmatrix} 1 & 0 \\ 0 & d \end{bmatrix}$. Furthermore, by applying it to $(=_4) \in \mathscr{F}_{123}$ we conclude that $(=_4) T'^{\otimes 4} \in \mathscr{F}_1$, since it is in $\mathscr{F}_{123}$ but not in $\mathscr{F}_2 \cup \mathscr{F}_3$. It follows that $d$ is a power of i, and hence $\begin{bmatrix} 1 & 0 \\ 0 & d \end{bmatrix}$ is a power of $D$. Thus $T \in \mathbb{C}^* \cdot \langle D, H \rangle$.

Since both generators $D$ and $H$ are symmetric, $\langle D, H \rangle$ is closed under transpose. Hence we also have $\mathbf{S}$ is the stabilizer of $\mathscr{F}_{123}$ by left action.

$$\mathbf{S} = \{S \in \mathbf{G} \mid S\mathscr{F}_{123} \subseteq \mathscr{F}_{123}\} = \{S \in \mathbf{G} \mid S^{-1}\mathscr{F}_{123} \subseteq \mathscr{F}_{123}\},$$

where equivalently we can replace $\subseteq$ by $=$.

Define the left and right stabilizer groups of $\mathscr{A}$:

$$\mathrm{LStab}(\mathscr{A}) = \{T \in \mathbf{GL}_2(\mathbb{C}) | T\mathscr{A} \subseteq \mathscr{A}\};$$

$$\mathrm{RStab}(\mathscr{A}) = \{T \in \mathbf{GL}_2(\mathbb{C}) | \mathscr{A}T \subseteq \mathscr{A}\}.$$

Then it is easy to verify that both $D, H \in \mathrm{LStab}(\mathscr{A}) \cap \mathrm{RStab}(\mathscr{A})$. It is also clear that both $\mathrm{LStab}(\mathscr{A}) \subseteq \mathbf{S}$ and $\mathrm{RStab}(\mathscr{A}) \subseteq \mathbf{S}$. Hence $\mathbf{S} = \mathbb{C}^* \cdot \langle D, H \rangle = \mathrm{LStab}(\mathscr{A}) = \mathrm{RStab}(\mathscr{A})$. From now on we will write $\mathrm{Stab}(\mathscr{A})$, or $\mathbf{S}$, for this group.

https://doi.org/10.1017/9781107477063.007 Published online by Cambridge University Press

For future reference, we record this as a lemma.

**Lemma 6.13.** *The left and right stabilizer groups of $\mathscr{A}$ and the left and right stabilizer groups of $\mathscr{F}_{123}$ all coincide, and it is the group $\mathbb{C}^* \cdot \langle D, H \rangle$.*

We remark that $\mathbf{S}$ is not normalized by $\mathscr{T}$. We take $H \in \mathbf{S}$, and calculate that

$$AHA^{-1} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 0 \\ 0 & \sqrt{\mathrm{i}} \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1/\sqrt{\mathrm{i}} \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1/\sqrt{\mathrm{i}} \\ \sqrt{\mathrm{i}} & -1 \end{bmatrix},$$

which maps $[1, 1] \in \mathscr{A}$ to

$$[1, 1] \begin{bmatrix} 1 & 1/\sqrt{\mathrm{i}} \\ \sqrt{\mathrm{i}} & -1 \end{bmatrix} = [1 + \sqrt{\mathrm{i}}, \frac{1}{\sqrt{\mathrm{i}}} - 1]$$

(ignoring the scalar $\frac{1}{\sqrt{2}}$), having two nonzero entries of unequal norm, and hence not in $\mathscr{A}$. Therefore $AHA^{-1} \notin \mathbf{S}$.

The following observation is a key point: If $T$ and $T'$ are two elements of $\mathscr{T}$ from the same left coset of $\mathbf{S}$, i.e., $T\mathbf{S} = T'\mathbf{S}$, then $T$ satisfies (6.8) iff $T'$ satisfies (6.8). Thus we may use right multiplication from $\mathbf{S}$ to normalize $\mathscr{T}$ in Theorem 6.12. Formally, we only need to check for $T \in \mathscr{T}$ in the second condition in Theorem 6.12 for a complete set of left coset representatives of $\mathbf{S}$ in $\mathscr{T}$. We claim that the four matrices in (6.7) form such a list.

We show that for every $T \in \mathscr{T}$, $T$ belongs to one of the cosets $\mathbf{S}$, $A\mathbf{S}$, $B\mathbf{S}$ and $C\mathbf{S}$. To do that we will first narrow down the list of possible signatures for $[1, 0, 1]T^{\otimes 2}$ and $[1, 0]T$, by right multiplication from $\mathbf{S}$. By assumption that $T \in \mathscr{T}$, both $[1, 0, 1]T^{\otimes 2}$ and $[1, 0]T$ belong to $\mathscr{F}_{123}$. Binary signatures of $\mathscr{F}_{123}$ are nonzero scalar multiples of

$$[1, 0, \pm 1], [1, 0, \pm \mathrm{i}], [1, \pm 1, -1], [1, \pm \mathrm{i}, 1], [0, 1, 0].$$

Unary signatures of $\mathscr{F}_{123}$ are nonzero scalar multiples of

$$[1, \pm 1], [1, \pm \mathrm{i}], [1, 0], [0, 1].$$

Using powers of $D = \begin{bmatrix} 1 & 0 \\ 0 & \mathrm{i} \end{bmatrix}$, we may already reduce $[1, 0, -1]$ to $[1, 0, 1]$, reduce $[1, 0, -\mathrm{i}]$ to $[1, 0, \mathrm{i}]$, and reduce $[1, 1, -1]$ to $[1, -1, -1]$, then to $[1, -\mathrm{i}, 1]$, and finally to $[1, \mathrm{i}, 1]$. We may use $Z^{-1}$ to reduce $[0, 1, 0]$ to $[1, 0, 1]$, as $[1, 0, 1]Z^{\otimes 2} = [0, 1, 0]$. We may further reduce $[1, 0, \mathrm{i}]$ to $[1, \mathrm{i}, 1]$ as follows. $[1, 0, \mathrm{i}]H^{\otimes 2} = \frac{1}{2} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & \mathrm{i} \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} = \frac{1+\mathrm{i}}{2}[1, -\mathrm{i}, 1]$, and $[1, -\mathrm{i}, 1](D^2)^{\otimes 2} = [1, \mathrm{i}, 1]$.

https://doi.org/10.1017/9781107477063.007 Published online by Cambridge University Press

So we may assume $[1,0,1]T^{\otimes 2}$ is a nonzero scalar multiple of $[1,0,1]$ or $[1,\mathrm{i},1]$. Note that these two signatures are invariant under reversal. Now we further normalize $[1,0]T$ by the reversal operation $X = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$, which does not affect the normalization already done for $[1,0,1]T^{\otimes 2}$. Then it is clear that we may assume $[1,0]T$ is a nonzero scalar multiple of $\{[1,0],[1,1],[1,-1],[1,\mathrm{i}]\}$.

In the following, we denote by $\alpha = \frac{1+\mathrm{i}}{\sqrt{2}} = \sqrt{\mathrm{i}}$. We also omit a nonzero scalar factor if it does not cause any confusion.

Suppose $[1,0,1]T^{\otimes 2} = \lambda[1,0,1]$. If $[1,0]T = \mu[1,0]$, then $T = \begin{bmatrix} 1 & 0 \\ 0 & \pm 1 \end{bmatrix} \in \mathbf{S}$, with coset representative $I$. If $[1,0]T = \mu[1,1]$, then $T = \begin{bmatrix} 1 & 1 \\ \pm 1 & \mp 1 \end{bmatrix} \in \mathbf{S}$, with coset representative $I$. Note that $\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} = \sqrt{2}H$ and $\begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} X = \sqrt{2}H$. If $[1,0]T = \mu[1,-1]$, then $T = \begin{bmatrix} 1 & -1 \\ \pm 1 & \pm 1 \end{bmatrix} \in \mathbf{S}$, with coset representative $I$. Note that $\begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} D^2 = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \in \mathbf{S}$ and $\begin{bmatrix} 1 & -1 \\ -1 & -1 \end{bmatrix} D^2 X \in \mathbf{S}$. Finally $[1,0]T = \mu[1,\mathrm{i}]$ with $[1,0,1]T^{\otimes 2} = \lambda[1,0,1]$ is impossible, since the latter implies that $T$ is orthogonal up to a nonzero factor, but the dot product $[1,\mathrm{i}] \cdot [1,\mathrm{i}] = 0$.

Next suppose $[1,0,1]T^{\otimes 2} = \lambda[1,\mathrm{i},1]$. If $[1,0]T = \mu[1,0]$, then $T = \begin{bmatrix} \sqrt{2} & 0 \\ \pm \mathrm{i} & \pm 1 \end{bmatrix}$. Clearly with the $+$ sign, $T = B \in B\mathbf{S}$, and with the $-$ sign, $T = CD^2 \in C\mathbf{S}$. If $[1,0]T = \mu[1,1]$, then $T = \begin{bmatrix} 1 & 1 \\ \pm\alpha^3 & \mp\alpha^3 \end{bmatrix}$. Note that $A^{-1}\begin{bmatrix} 1 & 1 \\ \alpha^3 & -\alpha^3 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ \mathrm{i} & -\mathrm{i} \end{bmatrix} \in \mathbf{S}$, being a scalar multiple of $Z$, and also $A^{-1}\begin{bmatrix} 1 & 1 \\ -\alpha^3 & \alpha^3 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ -\mathrm{i} & \mathrm{i} \end{bmatrix} \in \mathbf{S}$, being a scalar multiple of $ZX$. If $[1,0]T = \mu[1,-1]$, then $T = \begin{bmatrix} 1 & -1 \\ \pm\alpha & \pm\alpha \end{bmatrix}$. Note that $A^{-1}\begin{bmatrix} 1 & -1 \\ \alpha & \alpha \end{bmatrix} = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \in \mathbf{S}$, being a multiple of $HD^2$, and also $A^{-1}\begin{bmatrix} 1 & -1 \\ -\alpha & -\alpha \end{bmatrix} = \begin{bmatrix} 1 & -1 \\ -1 & -1 \end{bmatrix} \in \mathbf{S}$, being a multiple of $HXD^2$. Finally, if $[1,0]T = \mu[1,\mathrm{i}]$, then $T = \begin{bmatrix} 1 & \mathrm{i} \\ 0 & \pm\sqrt{2} \end{bmatrix}$. Note that $C^{-1}\begin{bmatrix} 1 & \mathrm{i} \\ 0 & \sqrt{2} \end{bmatrix} = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 0 \\ \mathrm{i} & \sqrt{2} \end{bmatrix}\begin{bmatrix} 1 & \mathrm{i} \\ 0 & \sqrt{2} \end{bmatrix} = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & \mathrm{i} \\ \mathrm{i} & 1 \end{bmatrix} = ZD \in \mathbf{S}$,

https://doi.org/10.1017/9781107477063.007 Published online by Cambridge University Press

and

$$B^{-1}\begin{bmatrix}1 & \mathrm{i}\\ 0 & -\sqrt{2}\end{bmatrix} = \frac{1}{\sqrt{2}}\begin{bmatrix}1 & 0\\ -\mathrm{i} & \sqrt{2}\end{bmatrix}\begin{bmatrix}1 & \mathrm{i}\\ 0 & -\sqrt{2}\end{bmatrix} = \frac{1}{\sqrt{2}}\begin{bmatrix}1 & \mathrm{i}\\ -\mathrm{i} & -1\end{bmatrix} = ZXD \in \mathbf{S}.$$

We have proved that for every $T \in \mathscr{T}$, there exists some $M \in \{I, A, B, C\}$, such that $T \in M\mathbf{S}$. We remark that the four coset representatives $\{I, A, B, C\}$ represent distinct left cosets. This amounts to verifying that $A, B, C, B^{-1}A, C^{-1}A, C^{-1}B$ all do not belong to $\mathbf{S}$. We have $[1, 1]A = [1, \sqrt{\mathrm{i}}]$, $[1, 1]B = [\sqrt{2} + \mathrm{i}, 1]$, and $[1, 1]C = [\sqrt{2} - \mathrm{i}, 1]$, having two nonzero entries with unequal norm or a ratio not a power of i, hence not in $\mathscr{A}$. Similarly $[0, 1]B^{-1}A = \frac{1}{\sqrt{2}}[0, 1]\begin{bmatrix}1 & 0\\ -\mathrm{i} & \sqrt{2}\end{bmatrix}\begin{bmatrix}1 & 0\\ 0 & \sqrt{\mathrm{i}}\end{bmatrix} = \frac{1}{\sqrt{2}}[-\mathrm{i}, \sqrt{2\mathrm{i}}]$, $[0, 1]C^{-1}A = \frac{1}{\sqrt{2}}[0, 1]\begin{bmatrix}1 & 0\\ \mathrm{i} & \sqrt{2}\end{bmatrix}\begin{bmatrix}1 & 0\\ 0 & \sqrt{\mathrm{i}}\end{bmatrix} = \frac{1}{\sqrt{2}}[\mathrm{i}, \sqrt{2\mathrm{i}}]$, and $[0, 1]C^{-1}B = \frac{1}{\sqrt{2}}[0, 1]\begin{bmatrix}1 & 0\\ \mathrm{i} & \sqrt{2}\end{bmatrix}\begin{bmatrix}\sqrt{2} & 0\\ \mathrm{i} & 1\end{bmatrix} = [2\mathrm{i}, 1]$, all have two nonzero entries of unequal norm. Thus these four matrices are distinct coset representatives.

We conclude this discussion by restating Theorem 6.12 in the following effective form:

**Theorem 6.14.** *Let $\mathcal{F}$ be a set of complex symmetric signatures.* Holant$^c(\mathcal{F})$ *is #P-hard unless $\mathcal{F}$ satisfies one of the following conditions, in which case it is tractable:*

1. Holant$^*(\mathcal{F})$ *is tractable, for which we have an effective dichotomy – Theorem 2.12; or*
2. *There exists some $T$ among the four matrices defined in (6.7), such that $T^{-1}\mathcal{F} \subseteq \mathscr{A}$.*

## 6.4 A Dichotomy for #CSP$^d$

In this section we consider a variant of the standard Boolean #CSP problem, where every variable occurs a multiple of $d$ times, for some integer $d \geq 1$. We will prove a complexity dichotomy, due to Huang and Lu [HL16], for this class of problems for any set of symmetric signatures on Boolean variables. This variant is interesting in its own right, but more importantly it is a key step in the proof of the general dichotomy in Chapter 7.

**Definition 6.15.** Let $d \geq 1$ be an integer. Given a set of signatures $\mathcal{F}$, we use #CSP$^d(\mathcal{F})$ to denote the restriction of Boolean #CSP($\mathcal{F}$) where every variable occurs a multiple of $d$ times.

https://doi.org/10.1017/9781107477063.007 Published online by Cambridge University Press

For example, #CSP$^2$ refers to #CSP problems where every variable occurs an even number of times. Define

$$\mathcal{T}_d = \left\{ \begin{bmatrix} 1 & 0 \\ 0 & \omega_d^k \end{bmatrix} \mid 0 \le k < d \right\},$$

where $\omega_d = e^{2\pi \mathrm{i}/d}$. We may apply a holographic transformation by $T \in \mathcal{T}_d$. Since each variable occurs a multiple of $d$ times, this transforms #CSP$^d(\mathcal{F})$ to the equivalent #CSP$^d(T^{-1}\mathcal{F})$. Consider the problem #CSP$^3([1, \omega_3, \omega_6])$, where $\omega_3 = e^{2\pi \mathrm{i}/3} = \frac{-1+\sqrt{3}\mathrm{i}}{2}$ and $\omega_6 = e^{2\pi \mathrm{i}/6} = \frac{1+\sqrt{3}\mathrm{i}}{2}$. Taking $T = \begin{bmatrix} 1 & 0 \\ 0 & \omega_3 \end{bmatrix}$, the problem #CSP$^3([1, \omega_3, \omega_6])$ is transformed to the equivalent problem #CSP$^3([1, 1, -1])$ (where $-1 = \omega_6/\omega_3^2$), which is computable in polynomial time, by Theorem 3.7. Note that the standard #CSP$([1, \omega_3, \omega_6])$ is #P-hard, by the same theorem.

As another example, #CSP$^2([1, \mathrm{i}^{3/2}, \mathrm{i}])$ is computable in polynomial time. We can first write #CSP$^2([1, \mathrm{i}^{3/2}, \mathrm{i}])$ as an equivalent bipartite Holant problem Holant$(\{=_2, =_4, \ldots\} \mid [1, \mathrm{i}^{3/2}, \mathrm{i}])$. Then we can perform a holographic transformation by $T = \begin{bmatrix} 1 & 0 \\ 0 & \sqrt{\mathrm{i}} \end{bmatrix} \in \mathcal{T}_8$. The LHS $(=_{2k})$ is transformed to $(=_{2k})T^{\otimes 2k} = [1, 0]^{\otimes 2k} + \mathrm{i}^k[0, 1]^{\otimes 2k} \in \mathscr{A}$. On the RHS, the transformed signature is $T^{-1}[1, \mathrm{i}^{3/2}, \mathrm{i}] = [1, \mathrm{i}, 1] \in \mathscr{A}$ as well. It follows that the problem is tractable in P. Note that the standard #CSP$([1, \mathrm{i}^{3/2}, \mathrm{i}])$ is #P-hard by Theorem 3.7.

The dichotomy theorem for #CSP$^d$ problems says that these are essentially the only additional tractable cases; all others are #P-hard.

**Theorem 6.16** (Huang-Lu)**.** *Let $d \ge 1$ be an integer, and $\mathcal{F}$ be a set of symmetric signatures from Boolean variables to $\mathbb{C}$. Then* #CSP$^d(\mathcal{F})$ *is #P-hard, unless there exists some $T \in \mathcal{T}_{4d}$ such that $\mathcal{F} \subseteq T\mathscr{P}$ or $\mathcal{F} \subseteq T\mathscr{A}$, in which case the problem* #CSP$^d(\mathcal{F})$ *is in P.*

We remark that since $T$ in the theorem is diagonal, the statement $\mathcal{F} \subseteq T\mathscr{P}$ can be replaced by simply $\mathcal{F} \subseteq \mathscr{P}$.

Before we prove this theorem, we recall Theorem 6.5. The condition that the needed binary or unary signatures being normalized can always be achieved by a holographic transformation $T \in \mathcal{T}_3$ defined in (6.3). Thus we can restate Theorem 6.5 as follows:

**Theorem 6.17.** *Suppose $\mathcal{F}$ and $\mathcal{G}$ are two sets of signatures, where $\mathcal{G}$ contains a nondegenerate binary signature $[g_0, g_1, g_2]$. Furthermore, if $g_0 = g_2 = 0$*

https://doi.org/10.1017/9781107477063.007 Published online by Cambridge University Press

*then we assume $\mathcal{G}$ also contains a unary signature $[a, b]$ with $ab \neq 0$. Then*

$$\text{Holant}(\mathcal{G} \mid [1, 0, 0, 1] \cup \mathcal{F})$$

*is #P-hard unless there exists some $T \in \mathcal{T}_3$ such that $\mathcal{G}T \cup T^{-1}\mathcal{F} \subseteq \mathscr{P}$ or $\mathcal{G}T \cup T^{-1}\mathcal{F} \subseteq \mathscr{A}$, in which case the problem is in P. Furthermore, if $\mathcal{G}$ contains $\Delta_0$ and $\Delta_1$, then a #P-hardness reduction exists that does not use Lemma 3.13.*

The tractability proof of Theorem 6.16 is straightforward. The hardness proof of Theorem 6.16 will use Theorem 6.17. For general #CSP problems, there is a pinning lemma (Lemma 3.13) that allows the use of the unary $\Delta_0$ and $\Delta_1$ functions. It turns out that pinning with $\Delta_0$ and $\Delta_1$ is still available in the #CSP$^d$ setting; however, the proof is a bit more subtle. We will show in Lemma 6.22 that a weaker version of pinning holds that lets us pin $d$ occurrences of variables at a time to the same value in the #CSP$^d$ setting. Then we will show that this is sufficient for Theorem 6.16. This weaker version of pinning can be proved constructively. After the proof of Theorem 6.16 we will come back to this issue of pinning in general. The main part of the proof of Theorem 6.16 is to show that the theorem holds when $\Delta_0$ and $\Delta_1$ are present, which becomes Holant$^c$ in essence. The proof uses Theorem 6.17 to repeatedly *regularize* a given signature set, lest the problem be #P-hard. Finally, the signature set is sufficiently regularized that it must satisfy the tractability criteria if it is not #P-hard.

We will postpone the proof of Lemma 6.22 for now, and focus first on the following Theorem 6.18. Note that $T^{-1}\Delta_c$ is a scalar multiple of $\Delta_c$ for an invertible diagonal $T$; thus $T^{-1}\Delta_c \in \mathscr{P} \cap \mathscr{A}$. Hence Theorem 6.18 is also a special case of Theorem 6.16.

**Theorem 6.18.** *Let $d \geq 1$ be an integer, and $\mathcal{F}$ be a set of symmetric signatures from Boolean variables to $\mathbb{C}$. Then #CSP$^d(\mathcal{F} \cup \{\Delta_0, \Delta_1\})$ is #P-hard, unless there exists some $T \in \mathcal{T}_{4d}$ such that $\mathcal{F} \subseteq T\mathscr{P}$ or $\mathcal{F} \subseteq T\mathscr{A}$, in which case the problem #CSP$^d(\mathcal{F} \cup \{\Delta_0, \Delta_1\})$ is in P. When it is #P-hard, a reduction exists that does not use Lemma 3.13.*

Again we remark that since $T$ in the theorem is diagonal, $\mathcal{F} \subseteq T\mathscr{P}$ is equivalent to $\mathcal{F} \subseteq \mathscr{P}$. We may also remove identically 0 signatures from $\mathcal{F}$ since it does not affect both the complexity of the problem and the tractability criteria.

Before we prove Theorem 6.18 we make some simple observations: Suppose $M = \begin{bmatrix} 1 & 0 \\ 0 & z \end{bmatrix} \in \mathbf{GL}_2(\mathbb{C})$.

https://doi.org/10.1017/9781107477063.007 Published online by Cambridge University Press

1. Signatures of the form $[f_0, 0, f_2, \ldots]$ or $[0, f_1, 0, \ldots]$ maintain their respective form under a holographic transformation by $M$. In particular, this is true for $\Delta_0$ and $\Delta_1$.
2. $f \in \mathscr{P}$ iff $fM \in \mathscr{P}$. This can be seen by the effect of the transformation $M$ on $(=_2)$, $(\neq_2)$ and the unary functions.
3. For all $d \geq 1$, $\mathcal{T}_d$ is a cyclic group. $M \in \mathcal{T}_d$ iff $M^{-1} \in \mathcal{T}_d$.
4. For all $d \geq 1$, $(=_d)M^{\otimes d}$ is a scalar multiple of $(=_d)$ iff $M \in \mathcal{T}_d$.
5. For all $d \geq 1$, $(=_d)M^{\otimes d} \in \mathscr{A}$ iff $M \in \mathcal{T}_{4d}$.

*Proof of Theorem 6.18.* We start with the following equivalence:

$$\#\mathrm{CSP}^d(\mathcal{F} \cup \{\Delta_0, \Delta_1\}) \equiv_{\mathrm{T}} \mathrm{Holant}(\mathscr{E}_d \mid \mathcal{F} \cup \{\Delta_0, \Delta_1\}),$$

where we denote $\mathscr{E}_d = \{=_d, =_{2d}, \ldots\}$. We first prove tractability. Suppose there exists some $T \in \mathcal{T}_{4d}$ such that $T^{-1}\mathcal{F} \subseteq \mathscr{P}$, or $T^{-1}\mathcal{F} \subseteq \mathscr{A}$; then we transform the bipartite Holant problem by $T$:

$$\mathrm{Holant}(\mathscr{E}_d \mid \mathcal{F} \cup \{\Delta_0, \Delta_1\}) \equiv_{\mathrm{T}} \mathrm{Holant}(\mathscr{E}_d T \mid T^{-1}\mathcal{F} \cup \{\Delta_0, \Delta_1\}).$$

Note that a diagonal $T$ preserves $\{\Delta_0, \Delta_1\}$. We have $\mathscr{E}_d T \subset \mathscr{P} \cap \mathscr{A}$, and either all $T^{-1}\mathcal{F} \cup \{\Delta_0, \Delta_1\} \subseteq \mathscr{P}$ or all $T^{-1}\mathcal{F} \cup \{\Delta_0, \Delta_1\} \subseteq \mathscr{A}$. Thus the problem is computable in polynomial time.

We turn to the proof of hardness next. We first prove a general reduction about $\#\mathrm{CSP}^d$ problems. For a function $f$, let $f^d$ be the function such that $f^d(\mathbf{x}) = (f(\mathbf{x}))^d$.

**Lemma 6.19.** *For all $d \geq 1$, $\#\mathrm{CSP}(f^d) \leq_{\mathrm{T}} \#\mathrm{CSP}^d(f)$. Similarly, $\#\mathrm{CSP}(f^d, \Delta_0, \Delta_1) \leq_{\mathrm{T}} \#\mathrm{CSP}^d(f, \Delta_0, \Delta_1)$.*

*Proof.* Given an instance of $\#\mathrm{CSP}(f^d)$, we may replace each constraint of $f^d$ by $d$ many copies of $f$, each applied to the same sequence of occurrences of variables. This gives the same answer to the given instance, and the new instance is a valid instance of $\#\mathrm{CSP}^d(f)$ since the number of occurrences of every variable is multiplied by $d$. The lemma is proved. □

By Theorem 3.7, $\#\mathrm{CSP}([0, 1, 0, 0])$ is #P-hard, and by Theorem 3.21 a #P-hardness reduction exists for $\#\mathrm{CSP}([0, 1, 0, 0], \Delta_0, \Delta_1)$ that does not use Lemma 3.13. The same is true for $\#\mathrm{CSP}([0, 0, 1, 0])$ and $\#\mathrm{CSP}([0, 0, 1, 0], \Delta_0, \Delta_1)$. Then by Lemma 6.19, we obtain the next lemma.

**Lemma 6.20.** *For all $d \geq 1$, $\#\mathrm{CSP}^d([0, 1, 0, 0])$ and $\#\mathrm{CSP}^d([0, 0, 1, 0])$ are #P-hard.*

*Reductions for #P-hardness of $\#\mathrm{CSP}^d([0, 1, 0, 0], \Delta_0, \Delta_1)$ and $\#\mathrm{CSP}^d([0, 0, 1, 0], \Delta_0, \Delta_1)$ exist that do not use Lemma 3.13.*

https://doi.org/10.1017/9781107477063.007 Published online by Cambridge University Press

We continue with the proof of Theorem 6.18. Suppose Holant($\mathscr{E}_d \mid \mathcal{F} \cup \{\Delta_0, \Delta_1\}$) is not #P-hard. Connecting $(=_d)$ on the LHS and $d-1$ copies of $\Delta_c$ on the RHS we get $\Delta_c$ on the LHS, for $c = 0, 1$, with which we can obtain all subsignatures of $\mathcal{F}$ on the RHS. Since we can obtain all subsignatures of $\mathcal{F}$, we can and will replace every degenerate (but not identically 0) $f = [a, b]^{\otimes n} \in \mathcal{F}$ by (a nonzero multiple of) the unary $[a, b]$, without changing the complexity of the problem. If $\mathcal{F}$ consists of only unary functions then $\mathcal{F} \subseteq \mathcal{U} \subset \mathscr{P}$. If all binary subsignatures of $\mathcal{F}$ are degenerate, then by Proposition 2.8, we also have $\mathcal{F} \subseteq \mathscr{P}$. Therefore we may assume $\mathcal{F}$ contains a nondegenerate binary subsignature.

The next claim is crucial: We may assume for any $f = [f_0, f_1, \ldots, f_n] \in \mathcal{F}$ $(n \geq 1)$ for any two consecutive terms at least one term must be 0,

$$f_i f_{i+1} = 0 \quad (0 \leq i < n). \tag{6.11}$$

Suppose for some $f$ and some $0 \leq i < n$, $f_i f_{i+1} \neq 0$. Then by pinning we can realize $[f_i, f_{i+1}]$, which is $[1, a]$ up to a scale, for some $a \neq 0$. By connecting $3d-3$ copies of $[1, a]$ to $=_{3d}$ on the LHS we get $[1, 0, 0, a^{3d-3}]$ on the LHS. Now we apply the holographic transformation by $M = \begin{bmatrix} 1 & 0 \\ 0 & a^{-(d-1)} \end{bmatrix}$; the LHS signature $[1, 0, 0, a^{3d-3}]$ then is transformed to $[1, 0, 0, a^{3d-3}]M^{\otimes 3} = [1, 0, 0, 1]$ and the RHS has a nondegenerate binary signature, as well as a unary signature $M^{-1}[1, a] = [1, a^d]$ with both entries nonzero. By Theorem 6.17 Holant($\mathscr{E}_d \mid \mathcal{F} \cup \{\Delta_0, \Delta_1\}$) is #P-hard unless there exists some $M' = MT$ where $T \in \mathcal{T}_3$, such that $\mathscr{E}_d M' \cup M'^{-1}\mathcal{F}$ is a subset of $\mathscr{P}$ or a subset of $\mathscr{A}$. Since $M'$ is diagonal, if $M'^{-1}\mathcal{F} \subseteq \mathscr{P}$ then $\mathcal{F} \subseteq \mathscr{P}$. For the second alternative, $(=_d)M' \in \mathscr{A}$ implies that $M' \in \mathcal{T}_{4d}$. Thus we have established the claim in (6.11), or else we have finished the proof of Theorem 6.18.

Thus, in particular, the only unary signatures in $\mathcal{F}$ are (nonzero multiples of) $\Delta_0$ or $\Delta_1$; the only binary signatures in $\mathcal{F}$ have the form $[0, *, 0]$ (Disequality) or $[*, 0, *]$ (Gen-Eq). If all signatures in $\mathcal{F}$ have arity $\leq 2$, then $\mathcal{F} \subseteq \mathscr{P}$, and Theorem 6.18 is proved.

Therefore we may assume there are signatures in $\mathcal{F}$ of arity $\geq 3$. If they are all Gen-Eq, then again $\mathcal{F} \subseteq \mathscr{P}$. So we may assume there is $f = [f_0, f_1, \ldots, f_n] \in \mathcal{F}$ of arity $n \geq 3$ and is not a Gen-Eq. We claim that for any such $f$, there are no two consecutive zero entries: $f_i = f_{i+1} = 0$. Suppose otherwise. Since it is not a Gen-Eq, there exists some $0 < j < n$ such that $f_j \neq 0$. Either $j < i$, in which case we pick the maximum such $j < i$; or $j > i+1$, in which case we pick the minimum such $j > i+1$. By (6.11), we get a subsignature $[0, f_j, 0, 0]$ or $[0, 0, f_j, 0]$. By Lemma 6.20, we conclude that #CSP$^d(\mathcal{F} \cup \{\Delta_0, \Delta_1\})$ is #P-hard.

https://doi.org/10.1017/9781107477063.007 Published online by Cambridge University Press

We can therefore strengthen the claim in (6.11) to say that every signature in $\mathcal{F}$ other than Gen-Eq must have the following strictly alternating form:

$$\forall i, (0 \le i < n), \quad \text{exactly one of } f_i \text{ or } f_{i+1} = 0. \tag{6.12}$$

Thus other than Gen-Eq, every $f \in \mathcal{F}$ has the form $[f_0, 0, f_2, \ldots]$ or $[0, f_1, 0, \ldots]$, where all symbols $f_i \neq 0$, and $\mathcal{F}$ contains such a signature of arity $\ge 3$. The strengthened claim (6.12) applies to all signatures in $\mathcal{F}$ of any arity $n \ge 1$, other than Gen-Eq.

In particular there exists some subsignature $[1, 0, a]$ from $\mathcal{F}$ with $a \neq 0$. Suppose there is a nondegenerate Gen-Eq in $\mathcal{F}$ with an odd arity $r \ge 3$. We connect this Gen-Eq on the RHS with $=_{2rd}$ on the LHS having an even arity, to create a nondegenerate Gen-Eq on the LHS of an odd arity $2rd - r \ge 3$. Then we connect zero or more copies of $[1, 0, a]$ on the RHS to it to realize a nondegenerate Gen-Eq on the LHS of arity 3. Now we apply a diagonal holographic transformation $M$ to map it to $[1, 0, 0, 1]$ on the LHS, while $M^{-1}[1, 0, a]$ on the RHS is nondegenerate. It does not have the form $[0, *, 0]$ since $M$ is diagonal. This allows us to apply Theorem 6.17. This completes the proof of Theorem 6.18 in this case.

We summarize what we may assume for $\mathcal{F}$:

> Every unary signature in $\mathcal{F}$ is $\Delta_0$ or $\Delta_1$. Every signature in $\mathcal{F}$ of arity $> 1$ is nondegenerate, and is either (1) a Gen-Eq of an even arity, or (2) has a strictly alternating form. And there is at least one signature in $\mathcal{F}$ of type (2), with arity $\ge 3$.

Without loss of generality, we may assume $\mathcal{F}$ contains a subsignature of the form $[0, 1, 0, a]$, for $a \neq 0$. The other form $[a, 0, 1, 0]$ is symmetric.

By Lemma 6.19, we see that

$$\#\mathrm{CSP}([0, 1, 0, a^d]) \le_{\mathrm{T}} \#\mathrm{CSP}^d([0, 1, 0, a]).$$

If $\#\mathrm{CSP}^d([0, 1, 0, a])$ is #P-hard, then Theorem 6.18 is proved. Suppose $\#\mathrm{CSP}^d([0, 1, 0, a])$ is not #P-hard, then $\#\mathrm{CSP}([0, 1, 0, a^d])$ is also not #P-hard, and by Theorem 3.7, $[0, 1, 0, a^d] \in \mathscr{P} \cup \mathscr{A}$. However, $[0, 1, 0, a^d] \notin \mathscr{P}$, so we have $[0, 1, 0, a^d] \in \mathscr{A}$. This forces $a^d = \pm 1$ (see the table for $\mathscr{F}_1 \cup \mathscr{F}_2 \cup \mathscr{F}_3$ in Section 3.2.) Now we can apply a holographic transformation with $T = \begin{bmatrix} 1 & 0 \\ 0 & a^{1/2} \end{bmatrix}$ to $\mathrm{Holant}(\mathscr{E}_d \mid \mathcal{F} \cup \{\Delta_0, \Delta_1, [0, 1, 0, a]\})$. Note that $T \in \mathcal{T}_{4d}$, as $a^{2d} = 1$. $(T^{-1})^{\otimes 3}$ maps $[0, 1, 0, a]$ to $[0, 1, 0, 1]$ in the RHS up to a scalar $a^{-1/2}$. Being diagonal $T^{-1}$ maintains $\Delta_0$ and $\Delta_1$. All Equality functions $(=_{4kd})$ in $\mathscr{E}_{4d} \subset \mathscr{E}_d$ on the LHS remain unchanged. Thus we still can obtain $\Delta_0$ and $\Delta_1$ on the LHS and thus obtain all subsignatures of the RHS of $\mathrm{Holant}(\mathscr{E}_d T \mid T^{-1}\mathcal{F} \cup \{\Delta_0, \Delta_1, [0, 1, 0, 1]\})$. In particular, we can realize

https://doi.org/10.1017/9781107477063.007 Published online by Cambridge University Press

$(=_2) = [1, 0, 1]$ on the RHS. Then we can realize all $\mathscr{E}_2 = \{=_2, =_4, \ldots\}$ on the LHS from $\mathscr{E}_{4d}$ on the LHS and $(=_2)$ on the RHS. Since we have $(=_2)$ on both sides now, we can reduce the following nonbipartite Holant problem to the original problem #CSP$^d(\mathcal{F} \cup \{\Delta_0, \Delta_1\})$:

$$\begin{aligned} &\text{Holant}(\mathscr{E}_2 \cup T^{-1}\mathcal{F} \cup \{\Delta_0, \Delta_1, [0, 1, 0, 1]\}) \qquad (6.13)\\ &\leq_{\text{T}} \text{Holant}(\mathscr{E}_d T \mid T^{-1}\mathcal{F} \cup \{\Delta_0, \Delta_1, [0, 1, 0, 1]\})\\ &\equiv_{\text{T}} \text{Holant}(\mathscr{E}_d \mid \mathcal{F} \cup \{\Delta_0, \Delta_1, [0, 1, 0, a]\})\\ &\equiv_{\text{T}} \#\text{CSP}^d(\mathcal{F} \cup \{\Delta_0, \Delta_1, [0, 1, 0, a]\})\\ &\equiv_{\text{T}} \#\text{CSP}^d(\mathcal{F} \cup \{\Delta_0, \Delta_1\}). \end{aligned}$$

Our final step is to prove that the problem in (6.13) is #P-hard, unless $T^{-1}\mathcal{F} \subseteq \mathscr{A}$, which will complete the proof of Theorem 6.18.

Since $T$ is diagonal, the summary on what we may assume for $\mathcal{F}$ is also valid for $T^{-1}\mathcal{F}$. Next we show that if there is a nondegenerate Gen-Eq in $T^{-1}\mathcal{F}$ not in $\mathscr{A}$, then the problem in (6.13) is #P-hard. After a scalar multiplication we may assume it is $[1, 0, \ldots, 0, b] \in T^{-1}\mathcal{F}$ of even arity $2k \geq 2$, with $b \neq 0$. If $2k = 2$, then it is $[1, 0, b]$. If $2k \geq 4$, then by connecting it to $(=_{2k-2}) \in \mathscr{E}_2$ we can realize $[1, 0, b]$. Now we reduce #CSP$([1, 0, b], [0, 1, 0, 1])$ to #CSP$^2([1, 0, b], [1, 0, 1], [0, 1, 0, 1])$, which in turn is reducible to the problem in (6.13): Each variable $x$ that occurs in $\ell$ constraints in an instance of #CSP$([1, 0, b], [0, 1, 0, 1])$ will be made to occur in $2\ell$ constraints. Each constraint $[1, 0, b]$ is replaced by a pair of constraints $[1, 0, b]$ and $(=_2) = [1, 0, 1]$ on the same two occurrences of variables. Each constraint $[0, 1, 0, 1]$ is replaced by two copies of the same constraint $[0, 1, 0, 1]$ on the same three occurrences of variables. This reduction shows that the problem in (6.13) is #P-hard unless $\{[1, 0, b], [0, 1, 0, 1]\} \subset \mathscr{P}$ or $\{[1, 0, b], [0, 1, 0, 1]\} \subset \mathscr{A}$. As $[0, 1, 0, 1] \notin \mathscr{P}$, we get $[1, 0, b] \in \mathscr{A}$, and thus $b^4 = 1$. This shows that all nondegenerate Gen-Eq in $T^{-1}\mathcal{F}$ are in $\mathscr{A}$. In particular we have shown that all signatures in $T^{-1}\mathcal{F}$ of arity $\leq 2$ are in $\mathscr{A}$. Note that binary signatures in $T^{-1}\mathcal{F}$ are either Gen-Eq or of the form $[0, *, 0] \in \mathscr{A}$. Also for the problem in (6.13), since $\Delta_0$ and $\Delta_1$ are present, if some nondegenerate Gen-Eq in $T^{-1}\mathcal{F}$ does not belong to $\mathscr{A}$, then there is a #P-hardness reduction that does not use Lemma 3.13.

Now we prove that all signatures in $T^{-1}\mathcal{F}$ of the strictly alternating form are also in $\mathscr{A}$. We may assume it has arity $\geq 3$. Suppose $T^{-1}\mathcal{F}$ has a ternary subsignature of the form $[0, 1, 0, b]$ $(b \neq 0)$, after a scalar multiple. We can reduce #CSP$([0, 1, 0, b])$ to the problem #CSP$^2([0, 1, 0, b], [0, 1, 0, 1])$, which is in turn reducible to the problem in (6.13). (Again due to the

https://doi.org/10.1017/9781107477063.007 Published online by Cambridge University Press

presence of $\Delta_0$ and $\Delta_1$ in (6.13), a #P-hardness proof of this problem exists that does not use Lemma 3.13.) The reduction from #CSP($[0, 1, 0, b]$) to #CSP$^2$($[0, 1, 0, b]$, $[0, 1, 0, 1]$) is as follows: We replace each constraint $[0, 1, 0, b]$ with a copy of $[0, 1, 0, b]$ and $[0, 1, 0, 1]$. Each variable now occurs an even number of times. Thus the problem #CSP($[0, 1, 0, b]$) is #P-hard unless $[0, 1, 0, b] \in \mathscr{A}$, which forces $b = \pm 1$. This argument can be extended to ternary subsignatures of the form $[b, 0, 1, 0]$ ($b \neq 0$). Indeed, within the problem in (6.13) we can construct $(\neq_2) = [0, 1, 0]$ as a subsignature of $[0, 1, 0, 1]$, and then connecting $(\neq_2)$ to all three inputs of $[b, 0, 1, 0]$ converts it to $[0, 1, 0, b]$, and we conclude that $b = \pm 1$. Now extending this pattern to any signature in $T^{-1}\mathcal{F}$ of the strictly alternating form, we see that either it is in $\mathscr{A}$ or there is a subsignature of arity 4 of the form

$$[1, 0, 1, 0, -1] \quad \text{or} \quad [1, 0, -1, 0, -1].$$

By Theorem 6.14, both Holant$^c$($[1, 0, 1, 0, -1]$) and Holant$^c$($[1, 0, -1, 0, -1]$) are #P-hard. This involves a check with Theorem 2.12 (or more directly Lemma 2.16; see Definition 2.9), and a check with the four matrices in (6.7) with reference to the table for $\mathscr{F}_1 \cup \mathscr{F}_2 \cup \mathscr{F}_3$ in Section 3.2. Hence the problem in (6.13) is also #P-hard.

This essentially completes our proof of Theorem 6.18, except for a technical point of nondependence of Lemma 3.13. Suppose we use the Holant$^c$ dichotomy Theorem 6.14. In the proof of Theorem 6.14 we used nondiagonal holographic transformations, which do not preserve $\{\Delta_0, \Delta_1\}$, and we used, for example, results on Holant($f$) for a trnary signature $f = [f_0, f_1, f_2, f_3]$) in the *double-root* case, Lemma 6.2. These proofs ultimately use the pinning reduction (Lemma 3.13). In the following we give an alternative proof that Holant$^c$($[1, 0, 1, 0, -1]$) and Holant$^c$($[1, 0, -1, 0, -1]$) are #P-hard without using Lemma 3.13.

**Lemma 6.21.** *If $a \in \mathbb{R}$ and $a \neq \pm 1$, then* Holant$^c$($[0, 1, 0, a]$) *is #P-hard. By symmetry,* Holant$^c$($[a, 0, 1, 0]$) *is also #P-hard. Furthermore, #P-hardness reductions exist that do not use Lemma 3.13.*

*Proof.* We start with Holant($[0, 1, 0, 0]$), which is #P-hard. This is the problem of counting perfect matchings in 3-regular graphs [DL92]. So the problem is #P-hard if $a = 0$.

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

The recursive construction is depicted in Figure 6.7. We claim that the signature of the $j$th gadget $N_j$ has the form $[0, x_j, 0, y_j]$. By an easy induction,

https://doi.org/10.1017/9781107477063.007 Published online by Cambridge University Press

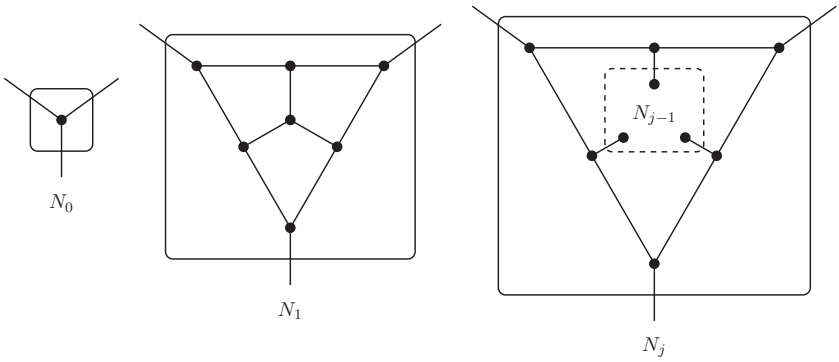


Figure 6.7. The recursive construction. The signature of every vertex in the gadget is $[0, 1, 0, a]$.

there are exactly $6j + 1$ vertices in $N_j$, which is odd. Let $\sigma \in \{0, 1\}^3$ be any assignment for the external edges of $N_j$, and let $\hat{\sigma}$ be an extension of $\sigma$ to all internal edges of $N_j$. Suppose $\hat{\sigma}$ has a nonzero evaluation. Then every node of $N_j$ must have an input of odd Hamming weight. Let $\mathrm{wt}(\sigma)$ be the Hamming weight of $\sigma$. Each internal edge assigned a value 1 by $\hat{\sigma}$ contributes 2 to the sum, over all nodes, of the Hamming weight of inputs, while each external edge assigned a value 1 by $\sigma$ contributes 1 to the sum. Hence $\mathrm{wt}(\sigma)$ must be odd.

A calculation shows that the sequence $(x_j, y_j)$ satisfies the following recurrence relation:

$$\begin{bmatrix} x_{j+1} \\ y_{j+1} \end{bmatrix} = \begin{bmatrix} 3(a^2+1) & (a^3+a) \\ 3(a^3+a) & a^6+1 \end{bmatrix} \begin{bmatrix} x_j \\ y_j \end{bmatrix},$$

with the initial vector $s = \begin{bmatrix} x_0 \\ y_0 \end{bmatrix} = \begin{bmatrix} 1 \\ a \end{bmatrix}$. Let $A = \begin{bmatrix} 3(a^2+1) & (a^3+a) \\ 3(a^3+a) & a^6+1 \end{bmatrix}$ be the transition matrix. Then $\det A = 3(a^4-1)^2 \neq 0$. The characteristic polynomial is $X^2 - (a^6 + 3a^2 + 4)X + 3(a^4-1)^2$ with discriminant $(a^6 - 3a^2 - 2)^2 + 12(a^3+a)^2 > 0$. Therefore $A$ has two distinct real roots, with sum $a^6 + 3a^2 + 4 > 0$. In particular, they are not opposite to each other. Therefore, the ratio of the two eigenvalues is not a root of unity. By Lemma 5.11, $\{A^j \mid j \geq 0\}$ is pairwise linearly independent. By Lemma 5.10, $A$ has infinite projective order.

We claim that $\det[s\ As] \neq 0$. We can calculate that $As = (a^2+1)\begin{bmatrix} a^2+3 \\ a(a^4-a^2+4) \end{bmatrix}$, and therefore $\det[s\ As] = a(a^2+1)(a^2-1)^2 \neq 0$. It follows that for any row eigenvector $v$ of $A$, the dot product $v \cdot s \neq 0$, for otherwise $v \cdot [s\ As] = 0$ and $v \neq 0$ would imply that $\det[s\ As] = 0$. It follows

https://doi.org/10.1017/9781107477063.007 Published online by Cambridge University Press

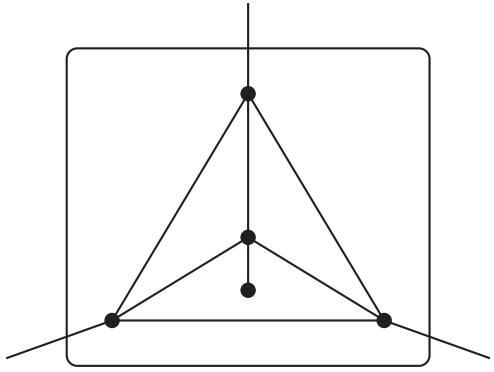

Figure 6.8. The signature of every degree-4 vertex in the gadget is $[1, 0, 1, 0, -1]$. And the signature of the degree-1 vertex in the gadget is $[1, 0]$.

from Lemma 5.15 that $\{A^j s\}$ is pairwise linearly independent. Therefore in Holant$^c([0, 1, 0, a])$ we can interpolate all signatures of the form $[0, 1, 0, x]$, and thus Holant$^c([0, 1, 0, a])$ is #P-hard. □

**Exercise:** Prove that with $A$ and $s$ we can interpolate all signatures of the form $[0, 1, 0, x]$ in Holant$^c([0, 1, 0, a])$.

Now we complete the proof that Holant$^c([1, 0, 1, 0, -1])$ is #P-hard. By a reversal $[1, 0, -1, 0, -1]$ becomes $-[1, 0, 1, 0, -1]$ and so Holant$^c([1, 0, -1, 0, -1])$ is also #P-hard.

Consider the gadget in Figure 6.8, where every vertex of degree 4 is assigned $[1, 0, 1, 0, -1]$, and the degree one vertex is assigned $[1, 0]$. This gadget has $[8, 0, 4, 0]$. So by Lemma 6.21, we know that Holant$^c([1, 0, 1, 0, -1])$ is #P-hard and the proof of Theorem 6.18 is complete.

Now we give the (weak) pinning lemma for #CSP$^d$, to remove the extra assumption that the signature set contains $\Delta_0$ and $\Delta_1$. This will improve Theorem 6.18 to Theorem 6.16.

**Lemma 6.22.** *For any integer $d \geq 1$ and any $\mathcal{F}$,* #CSP$^d(\mathcal{F}) \equiv_T$ #CSP$^d(\mathcal{F} \cup \{\Delta_0^{\otimes d}, \Delta_1^{\otimes d}\})$.

*Proof.* The proof of this lemma is an adaptation of the proof of Lemma 3.13. Given an instance $I$ of #CSP$^d(\mathcal{F} \cup \{\Delta_0^{\otimes d}, \Delta_1^{\otimes d}\})$, let $V_0$ denote the set of variables to which some $\Delta_0^{\otimes d}$ is applied among its $d$ occurrences of variables. Similarly define the set of variables $V_1$ in terms of $\Delta_1^{\otimes d}$. We may assume $V_0 \cap V_1 = \emptyset$; otherwise the partition function $Z(I) = 0$ trivially. Let $V_2 = V - (V_0 \cup V_1)$ be the set of other variables.

https://doi.org/10.1017/9781107477063.007 Published online by Cambridge University Press

Since $I$ is a #CSP$^d$ instance, each variable occurs a multiple of $d$ times, and therefore the total occurrences of variables in $V_0$ and in $V_1$ in all constraints are also multiples of $d$. If we remove all constraints $\Delta_0^{\otimes d}$ and $\Delta_1^{\otimes d}$, the total occurrences of variables in $V_0$ and in $V_1$ in all remaining constraints are still multiples of $d$.

Now we define an instance $I'$ of #CSP$^d(\mathcal{F})$ that has the variable set $V_2 \cup \{y_0, y_1\}$, where $y_0, y_1$ are two new variables. The constraints of $I'$ are exactly the same as those of $I$, except (1) every variable in $V_0$ (respectively $V_1$) is replaced by $y_0$ (respectively $y_1$), and (2) the constraints $\Delta_0^{\otimes d}$ and $\Delta_1^{\otimes d}$ are not applied. It is clear that $I'$ is indeed a valid instance of #CSP$^d(\mathcal{F})$. The partition function value $Z(I')$ can be expressed as the sum of four quantities,

$$Z(I') = \sum_{(b_0,b_1)\in\{0,1\}^2} Z(I' \mid b_0 b_1),$$

where $Z(I' \mid b_0 b_1)$ is the sum in $Z(I')$ restricted to assigning $(y_0, y_1)$ to $(b_0, b_1) \in \{0, 1\}^2$.

We also define a second instance $I''$ of #CSP$^d(\mathcal{F})$ that has the variable set $V_2 \cup \{y\}$, where $y$ is a new variable. It is the same as $I'$ except we identify both $y_0$ and $y_1$ by $y$. Then clearly

$$Z(I'') = \sum_{b_0=b_1} Z(I' \mid b_0 b_1) = Z(I' \mid 00) + Z(I' \mid 11).$$

Hence

$$Z(I') - Z(I'') = Z(I' \mid 01) + Z(I' \mid 10). \tag{6.14}$$

Obviously $Z(I) = Z(I' \mid 01)$. If for every possible #CSP$^d(\mathcal{F})$ instance $I^*$ with two distinguished variables $z_0$ and $z_1$, it holds that $Z(I^* \mid 01) = Z(I^* \mid 10)$, where $Z(I^* \mid b_0 b_1)$ is the sum in $Z(I^*)$ restricted to assigning $(z_0, z_1)$ to $(b_0, b_1) \in \{0, 1\}^2$, then clearly $Z(I') - Z(I'') = 2Z(I)$, and the lemma is proved.

Suppose there exists some instance $I^*$ of #CSP$^d(\mathcal{F})$ with two distinguished variables $z_0$ and $z_1$, such that $Z(I^* \mid 01) \neq Z(I^* \mid 10)$. Then we will define two more instances of #CSP$^d(\mathcal{F})$. The first is $J'$, which is the same as $I'$, except it adds a copy of $I^*$ and identifies the variables $z_0$ and $z_1$ with $y_0$ and $y_1$, respectively. Then

$$Z(J') = \sum_{(b_0,b_1)\in\{0,1\}^2} Z(I' \mid b_0 b_1) Z(I^* \mid b_0 b_1).$$

https://doi.org/10.1017/9781107477063.007 Published online by Cambridge University Press

The second is $J''$, which is the same as $J'$, except it further identifies the variables $y_0$ and $y_1$. Then

$$Z(J'') = \sum_{b_0=b_1} Z(I' \mid b_0 b_1) Z(I^* \mid b_0 b_1).$$

Clearly

$$Z(J') - Z(J'') = Z(I' \mid 01) Z(I^* \mid 01) + Z(I' \mid 10) Z(I^* \mid 10). \quad (6.15)$$

Equations (6.14) and (6.15) form a nonsingular linear equation system, from which we can solve $Z(I' \mid 01) = Z(I)$. □

There is a subtlety in the proof of Lemma 6.22. It proves the existence of a polynomial-time Turing reduction from the RHS to the LHS in the statement of the lemma. But the proof given above does not constructively produce such a reduction. Given a finite constraint set $\mathcal{F}$, the proof is nonconstructive in deciding the existence of an instance $I^*$ with $Z(I^* \mid 01) \neq Z(I^* \mid 10)$, because there are an infinite number of potential instances $I^*$ of #CSP$^d(\mathcal{F})$ with two distinguished variables.

Artem Govorov [Gov15] has found an alternative proof that is constructive. With his permission we present it below.

Suppose such an instance $I^*$ exists. It has two distinguished variables, $z_0$ and $z_1$, and $n$ other "internal" variables $x_1, x_2, \ldots, x_n$ ($n \geq 0$). The values $Z(I^* \mid 01)$ and $Z(I^* \mid 10)$ are each defined by a sum over $2^n$ assignments on $x_1, x_2, \ldots, x_n$ over $\{0, 1\}^n$. We define a 1-1 correspondence between $(z_0, z_1, x_1, \ldots, x_n) \in \{0, 1\}^{n+2}$ and $(\overline{z_0}, \overline{z_1}, \overline{x_1}, \ldots, \overline{x_n}) \in \{0, 1\}^{n+2}$, by flipping each bit. Because $Z(I^* \mid 01) \neq Z(I^* \mid 10)$, in the two sums defining $Z(I^* \mid 01)$ and $Z(I^* \mid 10)$, there must be one corresponding pair of terms that are unequal. Let $\sigma \in \{0, 1\}^n$ define such a term in the sum for $Z(I^* \mid 01)$. Let $U_b = \{x_i \mid \sigma(x_i) = b\}$, for $b = 0, 1$. Note that each variable appears a multiple of $d$ times in $I^*$, and therefore the total number of occurrences in $I^*$ of variables in $U_b$ is also a multiple of $d$.

Now we modify the instance $I^*$ to $I^{**}$ as follows: We identify every variable $x_i \in U_b$ with $z_b$, for $b = 0, 1$. This creates a valid instance $I^{**}$ of #CSP$^d(\mathcal{F})$. It has two distinguished variables $z_0$ and $z_1$, and no "internal" variables. Clearly $Z(I^{**} \mid 01) \neq Z(I^{**} \mid 10)$. Each application of $f \in \mathcal{F}$ in $I^{**}$ can be only in a finite number of ways. Indeed if $f$ has arity $n$, then there can be at most $2^n$ many ways it can be applied to $z_0$ and $z_1$ as an ordered sequence of variables. Since $\mathcal{F}$ is finite, there can be at most a finite number of ways all constraints in $\mathcal{F}$ can be applied in $I^{**}$. Define a type $\tau$ as a pair $(f, s)$, where $f \in \mathcal{F}$ has arity $n$, and $s \in \{0, 1\}^n$ indicating the application $f(z_{s_1}, \ldots, z_{s_n})$. Now let $N_\tau$ be

https://doi.org/10.1017/9781107477063.007 Published online by Cambridge University Press

the number of times the type $\tau$ occurs in $I^{**}$, and we write $N_\tau = q_\tau d + r_\tau$, with $0 \le r_\tau < d$. Since the total number of occurrences of $z_0$ and of $z_1$ are multiples of $d$, if we remove all $q_\tau d$ applications of constraints of each type $\tau$, we obtain from $I^{**}$ a modified instance of #CSP$^d(\mathcal{F})$ with two distinguished variables, where each type of constraint can occur less than $d$ times each. There is only a finite number of such instances of #CSP$^d(\mathcal{F})$. For each type $\tau$ there is also exactly one instance of #CSP$^d(\mathcal{F})$ with two distinguished variables where $\tau$ is applied exactly $d$ times. Of these finitely many instances of #CSP$^d(\mathcal{F})$, at least one instance $I_*$ must satisfy

$$Z(I_* \mid 01) \neq Z(I_* \mid 10), \tag{6.16}$$

because $Z(I^{**} \mid 01) \neq Z(I^{**} \mid 10)$, and they are respectively products of the terms from the LHS and RHS in (6.16). This gives us a decidable procedure; consequently the reduction from the RHS to the LHS in Lemma 6.22 is made constructive.

Using Lemma 6.22, we can show the equivalence of #CSP$^d(\mathcal{F})$ and #CSP$^d(\mathcal{F} \cup \{\Delta_0, \Delta_1\})$. *Given* the fact that we have already proved a dichotomy theorem for the latter, this establishes a dichotomy theorem for the former, Theorem 6.16.

**Lemma 6.23.** #CSP$^d(\mathcal{F})$ *is in P (or is #P-hard, respectively) iff* #CSP$^d(\mathcal{F} \cup \{\Delta_0, \Delta_1\})$ *is in P (or is #P-hard, respectively).*

*Proof.* Clearly #CSP$^d(\mathcal{F}) \le_{\mathrm{T}}$ #CSP$^d(\mathcal{F} \cup \{\Delta_0, \Delta_1\})$. We have already proved a dichotomy theorem for the latter problem #CSP$^d(\mathcal{F} \cup \{\Delta_0, \Delta_1\})$. If it is computable in polynomial time, then so is #CSP$^d(\mathcal{F})$. If #CSP$^d(\mathcal{F} \cup \{\Delta_0, \Delta_1\})$ is #P-hard, we show #CSP$^d(\mathcal{F})$ is also #P-hard.

We observe that in all the proofs for #P-hardness so far, which ultimately lead to the #P-hardness of #CSP$^d(\mathcal{F} \cup \{\Delta_0, \Delta_1\})$, we reduce from one of the following three problems to it by a chain of reductions: (a) Holant($=_3 \mid [0, 1, 1]$) (Vertex Cover for 3-regular graphs), (b) Holant([1, 1, 0, 0]) (Matching), or (c) Holant([0, 1, 0, 0]) (Perfect Matching). There are only three reduction methods in this reduction chain, namely, direct gadget construction, polynomial interpolation, and holographic reduction. Importantly, by Theorem 6.18, when #CSP$^d(\mathcal{F} \cup \{\Delta_0, \Delta_1\})$ is #P-hard, a reduction exists that does not use Lemma 3.13.

For any signature grid $\Omega$ as an input to one of the three Holant problems listed earlier, the answer Holant$_\Omega$ to the Holant problem is a nonnegative integer. Let $d\Omega$ denote the signature grid that is a disjoint union of $d$ copies of $\Omega$. Then the Holant value Holant$_{d\Omega} = (\text{Holant}_\Omega)^d$, the $d$th power of the Holant

https://doi.org/10.1017/9781107477063.007 Published online by Cambridge University Press

value on $\Omega$. Being a nonnegative integer, the value $\text{Holant}_\Omega$ can be recovered uniquely from $\text{Holant}_{d\Omega}$, and therefore we may start our reduction from $d\Omega$ instead of $\Omega$.

Now we apply the chain of reductions to $d\Omega$. Each step is a reduction $f$ from a problem $\Pi$ to a problem $\Pi'$. When the reduction is applied to an instance $I$ of $\Pi$, $f$ defines a sequence $\{I'_i\}$ of instances of $\Pi'$, and outputs $I'_1, I'_2, \ldots, I'_m$, where both $m$ and the sizes of $I'_i$, $(1 \le i \le m)$, are polynomially bounded in the size of $I$. (Typically for polynomial interpolations there will be a polynomial number of output instances.) The three reduction methods all satisfy the following *locality* property: If the reduction is applied to an instance of $\Pi$ that consists of $d$ identical copies $dI$, it will output a polynomial number of instances of $\Pi'$, where each instance consists of $d$ identical copies of $I'_i$, namely $dI'_1, dI'_2, \ldots, dI'_{m'}$, ($m'$ is polynomially bounded in the size of $d\Omega$, possibly a larger polynomial than $m$.) Clearly this property is closed under composition. Note that the number of steps in the reduction chain is a constant depending only on the problem, but not the size of the problem instance. (Crucially the reduction method in Lemma 3.13 *does not* satisfy this locality property. This is why we made annotations all along to indicate when a reduction exists without using Lemma 3.13.)

Finally we reach a sequence of polynomially many instances of the form $d\Omega_i$ of polynomially bounded size for the problem $\#\text{CSP}^d(\mathcal{F} \cup \{\Delta_0, \Delta_1\})$, following the #P-hardness reduction of this problem. By the syntactic form of $d\Omega_i$, each $\Delta_0$ and $\Delta_1$ is applied a multiple of $d$ times. Thus we can rewrite this instance as an instance in $\#\text{CSP}^d(\mathcal{F} \cup \{\Delta_0^{\otimes d}, \Delta_1^{\otimes d}\})$. This shows that $\#\text{CSP}^d(\mathcal{F} \cup \{\Delta_0^{\otimes d}, \Delta_1^{\otimes d}\})$ is also #P-hard. By Lemma 6.22, $\#\text{CSP}^d(\mathcal{F})$ is #P-hard. □

This completes the proof of Theorem 6.16. We conclude this section with two remarks. The first is concerned with a subtlety in the proof of Lemma 6.23. The statement that the two problems $\#\text{CSP}^d(\mathcal{F})$ and $\#\text{CSP}^d(\mathcal{F} \cup \{\Delta_0, \Delta_1\})$ are either simultaneously in P or simultaneously #P-hard technically does not depend on the assumption that P is different from #P, although the presumed truth of this hypothesis is what makes the theory of this book interesting. The trivial reduction $\#\text{CSP}^d(\mathcal{F}) \le_T \#\text{CSP}^d(\mathcal{F} \cup \{\Delta_0, \Delta_1\})$ shows that if the former is #P-hard, then so is the latter; and if the latter is in P, then so is the former. The proof of the lemma shows that if the latter is #P-hard, then so is the former. Now suppose the former is in P. We want to claim that the latter is also in P. Suppose not. Then *by the dichotomy theorem* already proved, Theorem 6.18, the latter problem $\#\text{CSP}^d(\mathcal{F} \cup \{\Delta_0, \Delta_1\})$ is #P-hard. The proof of Lemma 6.23 shows that the former problem $\#\text{CSP}^d(\mathcal{F})$ is also #P-hard. Thus #P collapses to P, and therefore $\#\text{CSP}^d(\mathcal{F} \cup \{\Delta_0, \Delta_1\})$ is in P anyhow.

https://doi.org/10.1017/9781107477063.007 Published online by Cambridge University Press

The second remark is that the proof of Lemma 6.23 does not maintain planarity in general. This is because when we substitute $\Delta_0^{\otimes d}$ and $\Delta_1^{\otimes d}$ for $d$ occurrences of $\Delta_0$ and $\Delta_1$ in a signature grid of the form $d\Omega$, there is no guarantee we can do so while maintaining planarity. In the next section we will see that in some circumstances this is still possible.

Very recently Lin and Wang [LW16] showed that there is a general method to extract a constraint function $f$ from its tensor power $f^{\otimes d}$ in the Holant setting. This implies a general pinning lemma in the #CSP$^d$ setting. With their permission we present their result below. Note that this proof is nonconstructive.

**Lemma 6.24 (Lin-Wang).** *Let $f$ be a signature from Boolean variables to $\mathbb{C}$ and let $d \geq 1$ be an integer. Then we have*

$$\text{Holant}(\mathcal{F} \cup \{f\}) \leq_{\text{T}} \text{Holant}(\mathcal{F} \cup \{f^{\otimes d}\}),$$

*for any set $\mathcal{F}$ of signatures from Boolean variables to $\mathbb{C}$.*

*Proof.* We prove this lemma by induction on $d$. The base case $d = 1$ is trivial. Assume the lemma is true for all integers $1 \leq d' < d$, and we prove it for $d$.

We ask the question: Is there any signature grid $\Omega$ for Holant$(\mathcal{F} \cup \{f\})$ in which the value $\text{Holant}_\Omega \neq 0$ and the number of occurrences of $f$ is not divisible by $d$? If not then the following is a simple reduction proving the lemma: Given any signature grid $\Omega$ for Holant$(\mathcal{F} \cup \{f\})$, we return the value $\text{Holant}_\Omega = 0$ if the number of occurrences of $f$ is not divisible by $d$; otherwise, we replace every $d$ occurrences of $f$ in $\Omega$ by a single copy of $f^{\otimes d}$, thus creating a signature grid in Holant$(\mathcal{F} \cup \{f^{\otimes d}\})$.

Now suppose the answer to the question is yes, and let $\Omega_0$ be such a fixed signature grid, with the number of occurrences $m$ of $f$ not divisible by $d$. Write $m = qd + r$, where $1 \leq r < d$. Now replace the $qd$ occurrences of $f$ by $q$ copies of $f^{\otimes d}$. We can use one extra copy of $f^{\otimes d}$, applying its variables in places for the $r$ remaining copies of $f$ in $\Omega_0$; this creates an $(\mathcal{F} \cup \{f^{\otimes d}\})$-gate with dangling edges, defining a signature that is $f^{\otimes(d-r)}$, multiplied by a nonzero constant $\text{Holant}_{\Omega_0}$. Hence we have a reduction

$$\text{Holant}(\mathcal{F} \cup \{f^{\otimes(d-r)}\}) \leq_{\text{T}} \text{Holant}(\mathcal{F} \cup \{f^{\otimes d}\}).$$

Note that $1 \leq d - r < d$. Now induction completes the proof. □

Since #CSP$^d(\mathcal{F})$ is nothing but Holant$(\mathcal{F} \cup \mathcal{EQ}_d)$, where $\mathcal{EQ}_d = \{(=_{kd}) \mid k \geq 1\}$ is the set of all EQUALITIES of arity divisible by $d$, combining Lemma 6.22 obtaining $\Delta_0^{\otimes d}$ and $\Delta_1^{\otimes d}$ for #CSP$^d$, Lemma 6.24 has the following corollary.

https://doi.org/10.1017/9781107477063.007 Published online by Cambridge University Press

**Corollary 6.25.** *Let $\mathcal{F}$ be a set of signatures from Boolean variables to $\mathbb{C}$, then*

$$\#\mathrm{CSP}^d(\mathcal{F}) \equiv_{\mathrm{T}} \#\mathrm{CSP}^d(\mathcal{F} \cup \{\Delta_0, \Delta_1\}).$$

Using Corollary 6.25, Theorem 6.16 and Theorem 6.18 are obviously equivalent. In this logical derivation one can bypass the need to carefully annotate along the way when a reduction exists without using Lemma 3.13. The slight drawback is that the proof of Lemma 6.24, hence that of Corollary 6.25, is nonconstructive. It nonconstructively establishes the existence of a polynomial-time reduction. The proof given for Theorem 6.16 is constructive, in that whenever a stated polynomial-time reduction exists, the proof actually constructs one.

## 6.5 Eulerian Orientations and the Tutte Polynomial

In this section, we prove a specific problem to be #P-hard. This problem will serve as a new basic problem from which we prove #P-hardness in the dichotomy theorems that follow. We first define the problem. Recall that an orientation of an undirected graph is to assign a unique direction, out of two possible directions, to each edge.

**Definition 6.26.** Given a graph $G$, an orientation is an *Eulerian orientation* if for each vertex $v$ of $G$, the number of incoming edges of $v$ equals the number of outgoing edges of $v$.

We will prove that the problem of counting the number of Eulerian orientations, denoted by #Eulerian-Orientations, or #EO for short, is #P-hard over planar 4-regular graphs. Mihail and Winkler [MW96] proved that #EO is #P-hard over general graphs. Huang and Lu proved it for even-degree regular but not necessarily planar graphs. Guo and Williams improved it to planar 4-regular graphs. There is a related problem counting the number of Eulerian circuits. Given an undirected graph $G$, an Eulerian circuit is a closed circuit that traverses every edge exactly once. It is well known that an orientation is an Eulerian orientation iff there is an Eulerian circuit that follows the orientation. However, these problems are different. A single Eulerian orientation may correspond to many Eulerian circuits. See Section 6.6.

There is a natural expression of #EO over 4-regular graphs as a bipartite Holant problem. Given a 4-regular undirected graph $G$, we assign the binary Disequality signature on each edge, which corresponds to an orientation, then we assign the Exact-Two signature $[0, 0, 1, 0, 0]$ of arity 4 for each vertex of

https://doi.org/10.1017/9781107477063.007 Published online by Cambridge University Press

$G$. This creates a bipartite signature grid out of the edge–vertex incidence graph of $G$. Then this problem can be expressed as Holant($\neq_2 \mid [0, 0, 1, 0, 0]$). Over planar 4-regular graphs, we use the notation Pl-Holant($\neq_2 \mid [0, 0, 1, 0, 0]$) to denote this problem.

The Tutte polynomial of an undirected graph is a polynomial in two variables with integer coefficients. It contains a wealth of information about the graph.

**Definition 6.27** (cf. p. 335 of [Bol98])**.** Let $G = (V, E)$ be an undirected graph; then the Tutte polynomial of $G$ is defined as

$$\mathrm{T}(G; x, y) = \sum_{A \subseteq E} (x-1)^{\kappa(V,A)-\kappa(V,E)} (y-1)^{\kappa(V,A)+|A|-|V|}, \tag{6.17}$$

where $\kappa(V, A)$ denotes the number of connected components of the graph $(V, A)$.

To prove the problem Pl-Holant($\neq_2 \mid [0, 0, 1, 0, 0]$) #P-hard, we reduce from the problem of evaluating the Tutte polynomial at the point $(3, 3)$, which is #P-hard for planar graphs.

The following theorems give complexity classifications of the Tutte polynomial over general and planar graphs.

**Theorem 6.28** (Jaeger, Vertigan, and Welsh [JVW90])**.** *For $x, y \in \mathbb{C}$, evaluating the Tutte polynomial at $(x, y)$ is* #P*-hard over graphs unless*

$$(x-1)(y-1) = 1$$

*or* $(x, y) \in \{(1, 1), (-1, -1), (0, -1), (-1, 0), (\mathrm{i}, -\mathrm{i}), (-\mathrm{i}, \mathrm{i}), (\omega, \omega^2), (\omega^2, \omega)\}$*, where $\omega = e^{2\pi \mathrm{i}/3}$. In each exceptional case, the problem is in polynomial time.*

**Theorem 6.29** (Theorem 5.1 in [Ver05])**.** *For $x, y \in \mathbb{C}$, evaluating the Tutte polynomial at $(x, y)$ is* #P*-hard over planar graphs unless*

$$(x-1)(y-1) \in \{1, 2\} \quad \text{or} \quad (x, y) \in \{(1, 1), (-1, -1), (\omega, \omega^2), (\omega^2, \omega)\},$$

*where $\omega = e^{2\pi \mathrm{i}/3}$. In each exceptional case, the problem is in polynomial time.*

The first step in the reduction from the Tutte polynomial is concerned with a sum of weighted Eulerian orientations on the medial graph of a plane graph. Recall that a plane graph is a planar embedding of a planar graph. We define a medial graph next.

**Definition 6.30** (cf. p. 368 of [Bol98])**.** Given a connected plane graph $G$, its *medial graph* $G_m$ has a vertex $e'$ for each edge $e$ of $G$, and vertices $e'_1$ and $e'_2$

https://doi.org/10.1017/9781107477063.007 Published online by Cambridge University Press

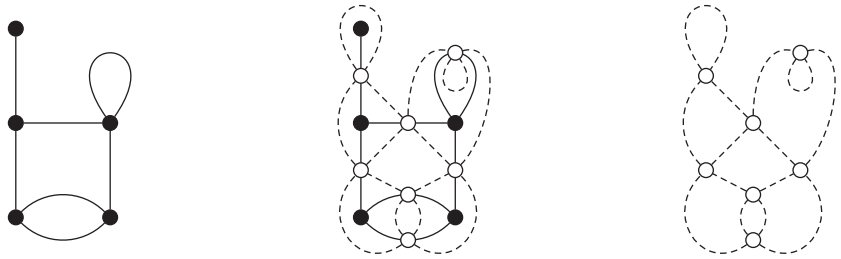

Figure 6.9. A plane graph (a), its medial graph (b), and the two graphs superimposed (c).

in $G_m$ are joined by an edge for each face of $G$ in which their corresponding edges $e_1$ and $e_2$ in $G$ occur consecutively.

Note that $G$ may have loops and parallel edges, and the vertices $e_1'$ and $e_2'$ of $G_m$ in Definition 6.30 may not be distinct. An example of a plane graph and its medial graph are given in Figure 6.9. Note that the medial graph of a plane graph is always a planar 4-regular graph. Michel Las Vergnas [Ver88] connected the evaluation of the Tutte polynomial of a planar graph $G$ at the point $(3, 3)$ with a sum of weighted Eulerian orientations on a medial graph of $G$.

**Theorem 6.31** (Theorem 2.1 in [Ver88])**.** *Let $G$ be a connected plane graph and let $\mathcal{O}(G_m)$ be the set of all Eulerian orientations in the medial graph $G_m$ of $G$. Then*

$$2 \cdot \mathrm{T}(G; 3, 3) = \sum_{O \in \mathcal{O}(G_m)} 2^{\beta(O)}, \tag{6.18}$$

*where $\beta(O)$ is the number of saddle vertices in the orientation $O$, i.e., the number of vertices in which the edges are oriented "in, out, in, out" in cyclic order.*

Although the medial graph depends on a particular embedding of the planar graph $G$, the RHS of (6.18) is invariant under different embeddings of $G$. This follows from (6.18) and the fact that the Tutte polynomial does not depend on the embedding of $G$.

Given a signature $g$ of arity 4, we abbreviate $g_{abcd}$ for the value $g(a, b, c, d)$, for $(a, b, c, d) \in \{0, 1\}^4$. Recall the definition of the *signature matrix* $M_g$ of an arity 4 signature $g$ (Definition 1.4):

$$M_g = \begin{bmatrix} g_{0000} & g_{0010} & g_{0001} & g_{0011} \\ g_{0100} & g_{0110} & g_{0101} & g_{0111} \\ g_{1000} & g_{1010} & g_{1001} & g_{1011} \\ g_{1100} & g_{1110} & g_{1101} & g_{1111} \end{bmatrix}. \tag{6.19}$$

https://doi.org/10.1017/9781107477063.007 Published online by Cambridge University Press

In general when we present $g$ pictorially with four external edges $ABCD$, we may order them counterclockwise. In $M_g$, the rows are indexed by the bits $AB$, and the columns are indexed by the bits $DC$ in the *reverse order*, both lexicographically. This reversal is for the convenience that the signature matrix of a construction that links two arity 4 signatures is the matrix product of the signature matrices of the two arity 4 signatures.

In particular, the signature matrix of a symmetric arity 4 signature has identical two middle rows and identical two middle rows. If $f = [f_0, f_1, f_2, f_3, f_4]$ is a symmetric signature of arity 4, then its signature matrix is

$$M_f = \begin{bmatrix} f_0 & f_1 & f_1 & f_2 \\ f_1 & f_2 & f_2 & f_3 \\ f_1 & f_2 & f_2 & f_3 \\ f_2 & f_3 & f_3 & f_4 \end{bmatrix}. \tag{6.20}$$

Now we can prove

**Theorem 6.32.** #Eulerian-Orientations *is* #P*-hard for planar 4-regular graphs.*

*Proof.* We reduce the problem of calculating the RHS of (6.18) to the problem Pl-Holant($\neq_2 \mid [0, 0, 1, 0, 0]$). Then by Theorems 6.29 and 6.31, we conclude that #EO is #P-hard for planar 4-regular graphs.

The RHS of (6.18) is the bipartite planar Holant problem Pl-Holant($\neq_2 \mid f$), where the signature matrix of $f$ is

$$M_f = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 1 & 2 & 0 \\ 0 & 2 & 1 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}.$$

Note that $f^{0101} = f^{1010} = 2$ are the two entries respectively listed in the second row indexed by 01 and the third column indexed by 01, and in the third row indexed by 10 and the second column indexed by 10. We perform a holographic transformation by $Z = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ \mathrm{i} & -\mathrm{i} \end{bmatrix}$ to get

$$\begin{aligned} \text{Pl-Holant}(\neq_2 \mid f) &\equiv_{\mathrm{T}} \text{Pl-Holant}((\neq_2)(Z^{-1})^{\otimes 2} \mid Z^{\otimes 4} f) \\ &\equiv_{\mathrm{T}} \text{Pl-Holant}(=_2 \mid \hat{f}) \\ &\equiv_{\mathrm{T}} \text{Pl-Holant}(\hat{f}), \end{aligned}$$

https://doi.org/10.1017/9781107477063.007 Published online by Cambridge University Press

where crucially $(=_2)Z^{\otimes 2} = (\neq_2)$, and the signature matrix of $\hat{f} = Z^{\otimes 4} f$ is

$$M_{\hat{f}} = \begin{bmatrix} 2 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 2 \end{bmatrix}, \tag{6.21}$$

and can be computed as $Z^{\otimes 2} M_f (Z^{\mathrm{T}})^{\otimes 2}$. An alternative way to compute this transformed signature is to write $Z^{\otimes 4} f$ by definition

$$\begin{aligned}
& Z^{\otimes 4} \left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix} \otimes \begin{bmatrix} 1 \\ 0 \end{bmatrix} \otimes \begin{bmatrix} 0 \\ 1 \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} 0 \\ 1 \end{bmatrix} \otimes \begin{bmatrix} 1 \\ 0 \end{bmatrix} \right. \\
& \quad + 2 \begin{bmatrix} 1 \\ 0 \end{bmatrix} \otimes \begin{bmatrix} 0 \\ 1 \end{bmatrix} \otimes \begin{bmatrix} 1 \\ 0 \end{bmatrix} \otimes \begin{bmatrix} 0 \\ 1 \end{bmatrix} + 2 \begin{bmatrix} 0 \\ 1 \end{bmatrix} \otimes \begin{bmatrix} 1 \\ 0 \end{bmatrix} \otimes \begin{bmatrix} 0 \\ 1 \end{bmatrix} \otimes \begin{bmatrix} 1 \\ 0 \end{bmatrix} \\
& \quad \left. + \begin{bmatrix} 0 \\ 1 \end{bmatrix} \otimes \begin{bmatrix} 1 \\ 0 \end{bmatrix} \otimes \begin{bmatrix} 1 \\ 0 \end{bmatrix} \otimes \begin{bmatrix} 0 \\ 1 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} \otimes \begin{bmatrix} 0 \\ 1 \end{bmatrix} \otimes \begin{bmatrix} 1 \\ 0 \end{bmatrix} \otimes \begin{bmatrix} 1 \\ 0 \end{bmatrix} \right\} \\
& = \frac{1}{4} \left\{ \begin{bmatrix} 1 \\ \mathrm{i} \end{bmatrix} \otimes \begin{bmatrix} 1 \\ \mathrm{i} \end{bmatrix} \otimes \begin{bmatrix} 1 \\ -\mathrm{i} \end{bmatrix} \otimes \begin{bmatrix} 1 \\ -\mathrm{i} \end{bmatrix} + \begin{bmatrix} 1 \\ \mathrm{i} \end{bmatrix} \otimes \begin{bmatrix} 1 \\ -\mathrm{i} \end{bmatrix} \otimes \begin{bmatrix} 1 \\ -\mathrm{i} \end{bmatrix} \otimes \begin{bmatrix} 1 \\ \mathrm{i} \end{bmatrix} \right. \\
& \quad + 2 \begin{bmatrix} 1 \\ \mathrm{i} \end{bmatrix} \otimes \begin{bmatrix} 1 \\ -\mathrm{i} \end{bmatrix} \otimes \begin{bmatrix} 1 \\ \mathrm{i} \end{bmatrix} \otimes \begin{bmatrix} 1 \\ -\mathrm{i} \end{bmatrix} + 2 \begin{bmatrix} 1 \\ -\mathrm{i} \end{bmatrix} \otimes \begin{bmatrix} 1 \\ \mathrm{i} \end{bmatrix} \otimes \begin{bmatrix} 1 \\ -\mathrm{i} \end{bmatrix} \otimes \begin{bmatrix} 1 \\ \mathrm{i} \end{bmatrix} \\
& \quad \left. + \begin{bmatrix} 1 \\ -\mathrm{i} \end{bmatrix} \otimes \begin{bmatrix} 1 \\ \mathrm{i} \end{bmatrix} \otimes \begin{bmatrix} 1 \\ \mathrm{i} \end{bmatrix} \otimes \begin{bmatrix} 1 \\ -\mathrm{i} \end{bmatrix} + \begin{bmatrix} 1 \\ -\mathrm{i} \end{bmatrix} \otimes \begin{bmatrix} 1 \\ -\mathrm{i} \end{bmatrix} \otimes \begin{bmatrix} 1 \\ \mathrm{i} \end{bmatrix} \otimes \begin{bmatrix} 1 \\ \mathrm{i} \end{bmatrix} \right\}.
\end{aligned}$$

Then it can be easily calculated that $\hat{f}^{0000} = \hat{f}^{1111} = 2$, $\hat{f}^{0011} = \hat{f}^{0101} = \hat{f}^{1010} = \hat{f}^{1100} = 1$, and all other entries are 0. (A side remark for readers who are familiar with notations from quantum computing: The function $f$ can also be denoted as $|0011\rangle + |0110\rangle + 2|0101\rangle + 2|1010\rangle + |1001\rangle + |1100\rangle$.)

We can perform the same holographic transformation by $Z$ on our target counting problem Pl-Holant$(\neq_2 \mid [0, 0, 1, 0, 0])$ and get

$$\begin{aligned}
& \text{Pl-Holant}(\neq_2 \mid [0, 0, 1, 0, 0]) \\
& \quad \equiv_{\mathrm{T}} \text{Pl-Holant}([0, 1, 0](Z^{-1})^{\otimes 2} \mid Z^{\otimes 4}[0, 0, 1, 0, 0]) \\
& \quad \equiv_{\mathrm{T}} \text{Pl-Holant}([1, 0, 1] \mid \tfrac{1}{2}[3, 0, 1, 0, 3]) \\
& \quad \equiv_{\mathrm{T}} \text{Pl-Holant}([3, 0, 1, 0, 3]).
\end{aligned}$$

We can verify $Z^{\otimes 4}[0, 0, 1, 0, 0] = \frac{1}{2}[3, 0, 1, 0, 3]$ as follows: First note that $Z^{\otimes 4}[0, 0, 1, 0, 0]$ is symmetric, and we may denote it as $[h_0, h_1, h_2, h_3, h_4]$.

https://doi.org/10.1017/9781107477063.007 Published online by Cambridge University Press

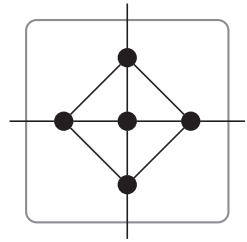

Figure 6.10. The planar tetrahedron gadget. Each vertex is assigned [3, 0, 1, 0, 3].

Writing $Z^{\otimes 4}[0, 0, 1, 0, 0]$, it is by definition $\frac{1}{4}$ of the sum over all six permutations of the tensor product $\begin{bmatrix}1\\ \mathrm{i}\end{bmatrix} \otimes \begin{bmatrix}1\\ \mathrm{i}\end{bmatrix} \otimes \begin{bmatrix}1\\ -\mathrm{i}\end{bmatrix} \otimes \begin{bmatrix}1\\ -\mathrm{i}\end{bmatrix}$. Hence it is clear that $h_0 = h_4 = 6/4 = 3/2$ and $h_2 = 1$. The entries of odd weight $h_1 = h_3 = 0$ because there is an involution $\pm\mathrm{i} \mapsto \mp\mathrm{i}$.

Our goal is to reduce Pl-Holant($\hat{f}$) to Pl-Holant([3, 0, 1, 0, 3]).

Consider the planar tetrahedron gadget in Figure 6.10. We assign [3, 0, 1, 0, 3] to every vertex and obtain a signature $32\hat{g}$, where the signature matrix of $\hat{g}$ is

$$M_{\hat{g}} = \frac{1}{2}\begin{bmatrix} 19 & 0 & 0 & 7 \\ 0 & 7 & 5 & 0 \\ 0 & 5 & 7 & 0 \\ 7 & 0 & 0 & 19 \end{bmatrix}.$$

This signature matrix requires some calculation to obtain. One can observe that all entries with odd Hamming weight from $\{0, 1\}^4$ are 0, because [3, 0, 1, 0, 3] has this property. Then, for example, the entry $\hat{g}^{0000}$ can be calculated as follows. Set the external pattern 0000. Then the four outside vertices virtually have the signature [3, 0, 1, 0]. The central vertex $v_0$ takes nonzero values at eight bit patterns: 0000, 1111, and the six patterns of Hamming weight 2. For 0000 at $v_0$, the contribution to $\hat{g}^{0000}$ is three times the trace of $\begin{bmatrix}3 & 0\\ 0 & 1\end{bmatrix}^4$, which is 246. Similarly, for 1111 at $v_0$, the contribution to $\hat{g}^{0000}$ is three times the trace of $\begin{bmatrix}0 & 1\\ 1 & 0\end{bmatrix}^4$, which is 6. For the bit patterns 0101 and 1010 of Hamming weight two at $v_0$, the contributions to $\hat{g}^{0000}$ is twice the trace of $\begin{bmatrix}3 & 0\\ 0 & 1\end{bmatrix}\begin{bmatrix}0 & 1\\ 1 & 0\end{bmatrix}\begin{bmatrix}3 & 0\\ 0 & 1\end{bmatrix}\begin{bmatrix}0 & 1\\ 1 & 0\end{bmatrix}$, which is 12. Finally, for 0011 at $v_0$, and its four cyclic permutations, the contributions to $\hat{g}^{0000}$ is four times the trace of $\begin{bmatrix}0 & 1\\ 1 & 0\end{bmatrix}\begin{bmatrix}0 & 1\\ 1 & 0\end{bmatrix}\begin{bmatrix}3 & 0\\ 0 & 1\end{bmatrix}\begin{bmatrix}3 & 0\\ 0 & 1\end{bmatrix}$, which

https://doi.org/10.1017/9781107477063.007 Published online by Cambridge University Press

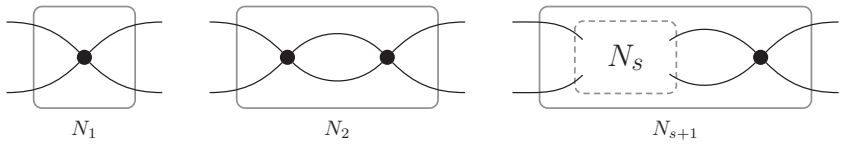


Figure 6.11. Recursive construction to interpolate $\hat{f}$. The vertices are assigned $\hat{g}$.

is 40. This gives the total value $\hat{g}^{0000} = 304$. Moreover, the signature $\hat{g}$ is invariant when we flip 0's with 1's, because [3, 0, 1, 0, 3] has this property. Thus $\hat{g}^{1111} = 304$ as well. Also, $\hat{g}$ is invariant under cyclic permutations owing to the symmetry of the tetrahedron gadget. All nonzero values of $\hat{g}$ can be computed similarly.

Now we show how to reduce Pl-Holant($\hat{f}$) to Pl-Holant($\hat{g}$) by interpolation. Consider an instance $\Omega$ of Pl-Holant($\hat{f}$). Suppose that $\hat{f}$ appears $n$ times in $\Omega$. We construct from $\Omega$ a sequence of instances $\Omega_s$ of Holant($\hat{g}$) indexed by $s \geq 1$. We obtain $\Omega_s$ from $\Omega$ by replacing each occurrence of $\hat{f}$ with the gadget $N_s$ in Figure 6.11 with $\hat{g}$ assigned to all vertices. Although $\hat{f}$ and $\hat{g}$ are asymmetric signatures, they are invariant under a cyclic permutation of their inputs. Thus, it is unnecessary to specify which edge corresponds to the first input of $\hat{f}$ and $\hat{g}$. We call such signatures *rotationally symmetric*.

To obtain $\Omega_s$ from $\Omega$, we effectively replace $M_{\hat{f}}$ with $M_{N_s} = (M_{\hat{g}})^s$, the $s$th power of the signature matrix $M_{\hat{g}}$. Let

$$T = \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & -1 & 0 \\ 1 & 0 & 0 & -1 \end{bmatrix} \quad \text{and} \quad \Lambda_{\hat{f}} = \begin{bmatrix} 3 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \ \Lambda_{\hat{g}} = \begin{bmatrix} 13 & 0 & 0 & 0 \\ 0 & 6 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 6 \end{bmatrix}.$$

Then

$$M_{\hat{f}} = T\Lambda_{\hat{f}}T^{-1} \qquad \text{and} \qquad M_{\hat{g}} = T\Lambda_{\hat{g}}T^{-1}$$

We can view our construction of $\Omega_s$ as first replacing each $M_{\hat{f}}$ by $T\Lambda_{\hat{f}}T^{-1}$ to obtain a signature grid $\Omega'$, which does not change the Holant value, and then replacing each $\Lambda_{\hat{f}}$ with $\Lambda_{\hat{g}}^s$. We stratify the assignments in $\Omega'$ based on the assignment to $\Lambda_{\hat{f}}$. Recall that the rows of $\Lambda_{\hat{f}}$ and $\Lambda_{\hat{g}}$ are indexed by 00, 01, 10, 11 and the columns are indexed by 00, 10, 01, 11, in their respective orders. We only need to consider the assignments to $\Lambda_{\hat{f}}$ that assign

- (00, 00) $j$ many times,
- (01, 10) or (11, 11) $k$ many times, and
- (10, 01) $\ell$ many times,

https://doi.org/10.1017/9781107477063.007 Published online by Cambridge University Press

where $j + k + \ell = n$, the total number of occurrences of $\Lambda_{\hat{f}}$ in $\Omega'$. Let $c_{jk\ell}$ be the sum over all such assignments of the products of evaluations from $T$ and $T^{-1}$ but excluding $\Lambda_{\hat{f}}$ on $\Omega'$. Then

$$\text{Pl-Holant}_{\Omega} = \sum_{j+k+\ell=n} 3^j c_{jk\ell}$$

and the value of the Holant on $\Omega_s$, for $s \geq 1$, is

$$\text{Pl-Holant}_{\Omega_s} = \sum_{j+k+\ell=n} (13^j 6^k)^s c_{jk\ell}. \tag{6.22}$$

This is a linear equation system with unknowns $c_{jk\ell}$, and a coefficient matrix whose rows are indexed by $s$ and columns are indexed by $(j, k)$, where $0 \leq j, k$ and $j + k \leq n$. We take $1 \leq s \leq \binom{n+2}{2}$. This coefficient matrix in the linear system of (6.22) is Vandermonde and has full rank since for any $j, k, j', k' \geq 0$, if $(j, k) \neq (j', k')$ then $13^j 6^k \neq 13^{j'} 6^{k'}$. Therefore, after obtaining the values of $\text{Pl-Holant}_{\Omega_s}$ by oracle calls to #EO, for $1 \leq s \leq \binom{n+2}{2}$, we can solve the linear system for the unknowns $c_{jk\ell}$ (where $j, k, \ell \geq 0$ and $j + k + \ell = n$) and obtain the value of $\text{Pl-Holant}_{\Omega}$. □

## 6.6 Kirchhoff's Matrix Tree Theorem

In this section we describe a celebrated formula counting the number of spanning trees in an undirected graph due to G. Kirchhoff [Kir47], and an extension to directed graphs due to W. T. Tutte [Tut48]. We then apply this to obtain a counting formula for the number of Eulerian circuits given an Eulerian orientation, known as the *BEST* theorem, originally discovered by de *B*ruijn and van Aardenne-*E*hrenfest [vAEdB51], with a special case due to *S*mith and *T*utte [TS41].

Let $G = (V, E)$ be a simple undirected graph without multiple edges and self loops. A spanning tree of a connected graph $G$ is a subgraph that forms a tree on $V$. Equivalently, it is a connected subgraph with $|V| - 1$ edges; or minimally connected; or acyclic with $|V| - 1$ edges; or maximally acyclic. The vertex–edge incidence matrix $M$ is defined as follows: The rows and columns of $M$ are indexed by $V$ and $E$ respectively. For every edge $e = \{i, j\}$, we define $M(i, e) = +1$ and $M(j, e) = -1$. (We arbitrarily pick one incident vertex of $e$ to assign $+1$ and the other incident vertex to assign $-1$.) Then $L = MM^{\mathrm{T}}$ is called a Laplacian matrix of $G$. $L$ is a symmetric positive semidefinite $|V| \times |V|$ matrix, with diagonal entries $L(i, i) = \deg(i)$, the degree of the vertex $i$, and $L(i, j) = L(j, i) = -1$ for $i \neq j$, if there is an edge $e = \{i, j\}$, and 0 otherwise. Note that $(1, 1, \ldots 1)^{\mathrm{T}}$ is an eigenvector of $L$ with eigenvalue $\lambda_0 = 0$.

https://doi.org/10.1017/9781107477063.007 Published online by Cambridge University Press

Kirchhoff's formula counts the number of labeled spanning trees of $G$ in terms of $L$.

**Theorem 6.33.** *(Matrix Tree Theorem) Suppose $G$ is a graph on $n$ vertices with Laplacian matrix $L$. Then the number of spanning trees of $G$ is the determinant of any $(n-1) \times (n-1)$ submatrix obtained by striking out row $i$ and column $i$ from $L$ (this value is the same for all $i$).*

*Equivalently the number of spanning trees of $G$ is*

$$\frac{1}{n}\lambda_1\lambda_2 \cdots \lambda_{n-1},$$

*where $0 = \lambda_0 \le \lambda_1 \le \lambda_2 \le \cdots \le \lambda_{n-1}$ are the eigenvalues of $L$.*

*Proof.* For $n = 1$, there is a unique spanning tree with no edges. By convention the determinant of the empty matrix is 1. If $|E| < n-1$, then there are no spanning trees, and the rank of $L$ is at most $|E|$ by the definition $L = MM^{\mathrm{T}}$. So we may assume $n \ge 2$ and $|E| \ge n-1$.

Consider the submatrix $M'$ obtained from $M$ by striking out row $i$. We are interested in the determinant of $M'M'^{\mathrm{T}}$. By the Cauchy–Binet formula,

$$\det(M'M'^{\mathrm{T}}) = \sum_{S \subseteq E:\ |S|=n-1} \det(M'_S)\det((M'_S)^{\mathrm{T}}) = \sum_{S \subseteq E:\ |S|=n-1} [\det(M'_S)]^2,$$

where the sum is over all subsets $S$ of $E$ of cardinality $n-1$, and $M'_S$ is the submatrix of $M'$ by choosing the columns indexed by $S$.

Suppose $(V, S)$ is disconnected, i.e., the subset of edges $S$ does not correspond to a spanning tree of $G$. Then there is a proper subset of vertices not containing $i$ that is disconnected from the rest, i.e., a subset $V' \subseteq V - \{i\}$ such that $V'$ is disconnected from $V - V'$ in $(V, S)$. Then, in $M'_S$, the sum of all rows indexed by $V'$ is 0, because for every edge $e = \{j, k\} \in S$, either both $j$ and $k$ are in $V'$ or both are not. Hence for such $S$, $\det(M'_S) = 0$.

Suppose $(V, S)$ is connected, i.e., the subset of edges $S$ corresponds to a spanning tree of $G$. Every spanning tree on $n \ge 2$ vertices has at least two leaves. If $n = 2$ then $M'_S$ has a single $\pm 1$ entry, indicating an edge incident to the remaining vertex and the removed vertex with index $i$. Then clearly $[\det(M'_S)]^2 = 1$. Suppose $n \ge 3$. There is at least one leaf $j \in [n] - \{i\}$. We can expand $\det(M'_S)$ by row $j$, which has a single $\pm 1$ entry corresponding to an edge $e$ incident to $j$, and the rest are all 0. By striking out row $j$ and column $e$, we effectively remove the leaf $j$. By induction we get $[\det(M'_S)]^2 = 1$.

It follows that $\det(M'M'^{\mathrm{T}})$ counts the number of spanning trees of $G$.

To prove the second claim, observe that $L$ is symmetric with row/column sum 0. Thus 0 is an eigenvalue, with eigenvector $v_0 = \frac{1}{\sqrt{n}}(1, 1, \ldots 1)^{\mathrm{T}}$. Being

https://doi.org/10.1017/9781107477063.007 Published online by Cambridge University Press

symmetric, the orthogonal complement $\langle v_0 \rangle^\perp$ of $v_0$ is an invariant subspace of $L$. The restriction of $L$ to $\langle v_0 \rangle^\perp$ is also symmetric. Thus we have a set of orthonormal eigenvectors $v_0, v_1, \ldots, v_{n-1}$. Let $T$ be an $n \times n$ matrix with column vectors $v_0, v_1, \ldots, v_{n-1}$, then

$$L = T \begin{bmatrix} 0 & 0 & 0 & \ldots & 0 \\ 0 & \lambda_1 & 0 & \ldots & 0 \\ 0 & 0 & \lambda_2 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \ldots & \lambda_{n-1} \end{bmatrix} T^{\mathrm{T}}. \tag{6.23}$$

We are interested in the determinant of the $(n-1) \times (n-1)$ submatrix $L_{11}$ obtained by striking out the first row and first column of $L$, which is the same as striking out the first row of $T$ and first column of $T^{\mathrm{T}}$ in the RHS of (6.23).

Notice that the first row and the first column of the diagonal matrix are both 0. Thus

$$L_{11} = T_{11} \begin{bmatrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & \lambda_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \lambda_{n-1} \end{bmatrix} (T_{11})^{\mathrm{T}},$$

where $T_{11}$ is obtained by striking out the first row and first column of $T$. Hence $\det(L_{11}) = [\det(T_{11})]^2 \lambda_1 \lambda_2 \cdots \lambda_{n-1}$.

If we add rows 2 to $n$ of $T$ to the first row, the $(1,1)$ entry of $T$ becomes $\sqrt{n}$, and all other entries of the first row become 0, because $v_1, \ldots, v_{n-1}$ are all orthogonal to $(1, 1, \ldots 1)^{\mathrm{T}}$. Thus $\det(T) = \sqrt{n} \det(T_{11})$. It follows that $[\det(T_{11})]^2 = \frac{1}{n}$, since for an orthogonal $T$, $\det(T) = \pm 1$. □

**Corollary 6.34.** *(Cayley) There are exactly $n^{n-2}$ labeled spanning trees in the complete graph $K_n$ of $n$ vertices.*

Next we extend Theorem 6.33 to directed graphs. Suppose $G$ is a directed graph. We assume $G$ has no loops and no multiple edges. Every edge $e = (i, j)$ has a direction from $i$ to $j$. The reverse ordered pair $(j, i)$, if it is present, is considered a distinct edge.

Let $r \in V$ be any vertex. An *arborescence* rooted at $r$ is a set of $|V| - 1$ directed edges, such that the underlying undirected graph is a spanning tree rooted at $r$, and all directed edges point from child to parent. Thus from any vertex $v \in V$ there is a unique directed path from $v$ to $r$ in an arborescence rooted at $r$. We want to count the number of arborescences rooted at any vertex in a directed graph.

https://doi.org/10.1017/9781107477063.007 Published online by Cambridge University Press

We define two $|V| \times |E|$ matrices $A$ and $B$. The rows and columns are indexed by $V$ and $E$ respectively. Suppose $e = (i, j)$ is a directed edge from $i$ to $j$. Then $A(i, e) = +1$, and $A(i', e) = 0$ for all $i' \neq i$ (including $i' = j$). For $B$, we define $B(i, e) = +1$, $B(j, e) = -1$, and $B(k, e) = 0$ for all $k \neq i, j$. Then we define the modified Laplacian matrix $L^+ = AB^{\mathrm{T}}$. It is easy to see that $L^+(i, i)$ is the outdegree of $i$. For $i \neq j$, $L^+(i, j) = -1$ if there is an edge $e = (i, j)$ from $i$ to $j$, and $L^+(i, j) = 0$ otherwise.

**Theorem 6.35.** *Suppose $G$ is a directed graph with an underlying connected undirected graph with $n$ vertices and the modified Laplacian matrix $L^+$. Then the number of arborescences rooted at $i$ is the determinant of the $(n-1) \times (n-1)$ submatrix obtained by striking out row $i$ and column $i$ from $L^+$.*

*Proof.* If we denote by $A'$ and $B'$ the submatrices of $A$ and $B$ obtained by deleting the $i$th rows respectively, then we are interested in $\det(A'B'^{\mathrm{T}})$.

By the Cauchy–Binet formula,

$$\det(A'B'^{\mathrm{T}}) = \sum_{S \subseteq E:\ |S| = n-1} \det(A'_S) \det(B'_S),$$

where the sum is over all subsets $S$ of $E$ of cardinality $n-1$, and $A'_S$ and $B'_S$ are submatrices of $A'$ and $B'$ by choosing the columns indexed by $S$.

Again, as in the proof of Theorem 6.33, if the underlying undirected graph defined by $S$ on $V$ disregarding directions is disconnected, then $\det(B'_S) = 0$.

Suppose $S$ is such that $\det(B'_S) \neq 0$. This happens iff the underlying undirected graph defined by $S$ on $V$ disregarding directions is a spanning tree. Then by picking off leaves recursively to expand the determinant $\det(B'_S)$, it defines an order (corresponding to a permutation $\pi$) of $V - \{i\}$, and all the entries from $B'_S$ used in the expansion are $\pm 1$. According to this order, at each step we remove a leaf of the spanning tree. If we consider expanding $\det(A'_S)$ in this same order, we find that at each step, inductively the row corresponding to the next leaf to be removed has at most one nonzero entry, and that entry is $+1$ iff the directed edge is pointing from the child (current leaf) to the parent in the spanning tree. And if these entries are all nonzero in $n-1$ steps, they are all $+1$, and the same $+1$ entries are encountered in the expansion of $\det(B'_S)$, corresponding to the same permutation $\pi$. The two factors of the same sign of the permutation $\pi$ in $\det(A'_S)$ and $\det(B'_S)$ cancel. Finally, $\det(A'_S) \det(B'_S) = 1$ for each arborescence rooted at $i$. □

An Eulerian circuit in an undirected graph $G$ is a traversal of all its edges exactly once, i.e., a cyclically ordered sequence of edges of $G$ where

https://doi.org/10.1017/9781107477063.007 Published online by Cambridge University Press

consecutive edges are adjacent (they share a common vertex) and every edge appears exactly once. An Eulerian circuit in a directed graph $G$ has the requirement that every edge must be traversed in the direction of the edge. An Eulerian circuit has a direction but no distinguished starting point.

Perhaps the earliest result in graph theory is Euler's observation [Eul36] that an undirected graph $G$ has an Eulerian circuit if and only if $G$ is connected and every vertex has an even degree. Euler proved that this is a necessary condition, and stated without proof that connected graphs with all vertices of even degree have an Eulerian circuit. The first complete proof of the sufficiency of the condition was published in 1873 by Hierholzer [Hie73]. There is a simple generalization to directed graphs, namely the underlying undirected graph is connected, and for every vertex its indegree is equal to its outdegree.

Any Eulerian circuit defines an Eulerian orientation, as there is an equal number of times an Eulerian circuit enters and exits each vertex. Using the method of Theorem 6.35 to count the number of arborescences in a directed graph, we have the following remarkable *BEST* theorem to count the number of Eulerian circuits in a graph given an Eulerian orientation.

**Theorem 6.36.** *(BEST) Let $G$ be connected and given an Eulerian orientation $\pi$. Let $r$ be any vertex of $G$ and $t_r(G)$ be the number of arborescences rooted at $r$ in $G$. Then the number of Eulerian circuits consistent with $\pi$ is*

$$t_r(G) \prod_{v \in V(G)} (d_v - 1)!$$

*where $d_v$ is both the outdegree as well as the indegree of $v$ under $\pi$.*

Notice that this theorem shows that in particular $t_r(G)$ does not depend on $r$ for any $G$ given an Eulerian orientation.

*Proof.* Fix any vertex $r$ of $G$. Consider any Eulerian circuit consistent with $\pi$ starting at $r$. It defines an arborescence $T$ rooted at $r$ as follows: For any vertex $v \neq r$, let $\text{exit}(v)$ be the last directed edge leaving $v$ on the Eulerian circuit before it completes the entire traversal and returns to $r$. Since the undirected underlying graph $G$ is connected, every vertex has $d_v > 0$ many incoming and outgoing edges. Since the Eulerian circuit traverses every edge, every vertex is visited $d_v$ times, and $\text{exit}(v)$ is well defined. It has the form $(v, w)$, for some $w$. If $\text{exit}(v) = (v, w)$ then the last visit of $w$ by the Eulerian circuit is after the last visit of $v$. Hence starting at any $v \neq r$, successively follow the directed

https://doi.org/10.1017/9781107477063.007 Published online by Cambridge University Press

edges from $\{\text{exit}(u) \mid u \neq r\}$, it cannot repeat any previously visited vertex, and thus must end in $r$. Hence these $|V| - 1$ edges define an arborescence $T$ rooted at $r$.

Now, given $\pi$ and an arborescence $T$ rooted at $r$ consistent with $\pi$, we claim there are exactly $\prod_{v \in V(G)}(d_v - 1)!$ many Eulerian circuits that define $T$. Given $\pi$ and $T$, we define a set of Eulerian circuits as follows: Starting at $r$, at any vertex follow any unused outgoing edge, with the only rule that at $v \neq r$, use $\text{exit}(v)$ only when there are no other choices. Once we prove these are indeed Eulerian circuits, it will be clear that they all define $T$ and that any Eulerian circuit that defines $T$ is among these.

This walk ends at some vertex because there are only finitely many edges. Since at any $v \neq r$, the walk has used one more incoming edge than outgoing edge, the walk cannot end there. Hence the walk ends at $r$. When it ends at $r$, it must have used up all outgoing edges at $r$. Since $\pi$ is an Eulerian orientation, there are an equal number of incoming and outgoing edges at $r$. Every time it is at $r$, it has traversed the same number of incoming and outgoing edges at $r$. So it must have traversed all incoming edges at $r$. These incoming edges include at least one tree edge belonging to $T$. Hence this walk traversed at least one tree edge. Consider the first time this walk used an edge from $T$, say $\text{exit}(v)$. We claim that this $v$ is a leaf in $T$. Otherwise, there is at least one tree edge ending in $v$, and it has not been traversed because $\text{exit}(v)$ is the first. But then $v$ has at least one outgoing edge other than $\text{exit}(v)$, since when visiting any $v \neq r$ the walk has used one more incoming edge than outgoing edge at $v$, thus one more unused outgoing edge than incoming edge at $v$. This is a contradiction by our rule of using the tree edge $\text{exit}(v)$ last at $v$. Hence, $v$ is a leaf, and when $\text{exit}(v)$ is used all incident edges of the leaf $v$ have been used.

Now we keep track of the progress of this walk by removing this leaf in $T$. Inductively we maintain an arborescence $T'$ obtained by successively removing leaves, starting from $T$. By the same argument, the walk ends at $r$, it must use some tree edge of $T'$ before it ends since it must use some tree edge of $T'$ pointing to $r$. The vertex $u$ where the first time (after forming $T'$) a tree edge $\text{exit}(u)$ is used must be a leaf in the current arborescence $T'$, at which point all incident edges of $u$ must have been used, and we may remove $u$, updating $T'$. It follows that when the walk ends at $r$, the walk has traversed every edge exactly once, in the direction consistent with the given Eulerian orientation $\pi$. Clearly this walk is an Eulerian circuit that defines the given arborescence $T$ rooted at $r$.

There are exactly $(d_v - 1)!$ choices at every $v \neq r$ and $d_r!$ choices at $r$, giving rise to $d_r! \prod_{v \neq r}(d_v - 1)!$ many Eulerian circuits consistent with $\pi$ and $T$.

https://doi.org/10.1017/9781107477063.007 Published online by Cambridge University Press

However, each such Eulerian circuit is counted exactly $d_r$ times since it passes through $r$ exactly $d_r$ times by a cyclic shift. The theorem follows. □

Regarding the problem of counting Eulerian circuits on undirected graphs, Brightwell and Winkler [BW05] proved that this problem is #P-Complete. Ge and Štefankovič [GŠ12] improved that to show that the problem remains #P-Complete for 4-regular graphs.

https://doi.org/10.1017/9781107477063.007 Published online by Cambridge University Press