COMPLEXITY CLASSIFICATION OF EXACT AND APPROXIMATE COUNTING PROBLEMS

by

Heng Guo

A dissertation submitted in partial fulfillment of
the requirements for the degree of

Doctor of Philosophy

(Computer Sciences)

at the

UNIVERSITY OF WISCONSIN–MADISON

2015

Date of final oral examination: 07/22/2015

The dissertation is approved by the following members of the Final Oral Committee:

Jin-Yi Cai (Advisor), Professor

Eric Bach, Professor

Dieter van Melkebeek, Professor

Shuchi Chawla, Associate Professor

Sebastian Roch, Associate Professor

© Copyright by Heng Guo 2015

All Rights Reserved

To my parents

Acknowledgments

First and foremost, I cannot express enough gratitude to Jin-Yi Cai, who has guided, advised, and tolerated me in the last five years. When I first came to Madison, I was an ambitious yet clueless young student. It is Jin-Yi who illustrated to me how to do research and solve problems. He also helped me to secure every possible opportunity: attending conferences, giving talks, visiting other institutions, and so on. His insights and knowledge helped me to develop my skill set, and his dedication and attitude inspired me to devote into academics. Truly my role model.

There are also other people from whom I received wisdom and to whom I am in debt. Thanks to the hospitality of Leslie Ann Goldberg, I spent six amazing months in Oxford. It was a productive period full of wonderful events and stimulating discussions, and I enjoyed a lot the relaxed atmosphere there. Another big thanks goes to Pinyan Lu, whom I have been working with before and after starting my Ph. D. Pinyan's cool tricks and observations are something I always enjoy. The area I am working in is largely due to the pioneer work of Leslie Valiant, and while working with Leslie his marvelous vision allows me to have a glimpse on how to be such an extraordinary scientist. I thank all my collaborators: Ivona Bezáková, Xi Chen, Andreas Galanis, Zhiguo Fu, Sangxia Huang, Mark Jerrum, Liang Li, Daniel Štefankovič, Eric Vigoda, Mingji Xia, and Yitong Yin. It has been a great pleasure to be able to work and exchange ideas with you all.

I thank the CS department of UW-Madison, which housed me in the past five years. In particular, I am lucky to be with the faculty and students of the theory group. I thank Eric Bach, Shuchi Chawla, Dieter van Melkebeek, and Sebastien Roch for the excellent classes they taught me, and for the burden they took to examine this dissertation. I would also like to thank Tyson Williams, a fellow student of Jin-Yi and who perhaps I have spent most time working with aside from Jin-Yi, for his help both in work and in life. In addition to brilliant ideas, Tyson's tireless and meticulous work is essential to achieve what we have done.

I am fortunate to have a group of friends who have helped me in all kinds of ways over the years. A surely incomplete list includes: Xixuan Feng, Andreas Goebel, Jiexing Li, Yinan Li, David Richerby, Wentao Wu, Xi Wu, Junming Xu, Chen Zeng, and Ce Zhang. Our "lunch group" meeting is usually the most relaxing time of the day, and it is always very enlightening to share thoughts with peers from outside of my area. In particular I would like to thank Ce Zhang, a brilliant guy who happened to be stuck under the same roof (a.k.a. roommate) with me for a few years. Without his careful explanation and patience for someone who knows nothing beyond theory, I would never be aware of the application side of some subjects I study.

I could have achieved nothing without the selfless and unconditional support from my parents. Thank you, and your love is one of my main drives to keep me moving forward.

Finally, I am very grateful to receive the Simons Awards in Theoretical Computer Science from the Simons Foundation. Their generous support opens a lot of opportunities for me and benefited my research in an unparalleled way. I also thank the financial support from NSF CCF-0914969 and NSF CCF-1217549.

Contents

Contents iv

List of Tables vii

List of Figures viii

Abstract xi

1 Introduction 1

1.1 Counting Problems and Complexity Classification 1

1.2 Approximate Counting 7

1.3 Problems and Definitions 11

1.4 Useful Reductions 13

1.5 Tractable Signature Sets 17

1.6 Some Known Dichotomies 22

2 Parity Holant Problems 25

2.1 Tractable $\mathbb{Z}_2$-Functions 26

2.2 Hardness Results 38

2.3 Parity Holant^c Dichotomy 43

2.4 Parity Vanishing Signature Sets 46

2.5 Dichotomy for Parity Holant 56

3 Holant Problems on 4-regular graphs 62

3.1 Complex Vanishing Signatures 62

3.2 Redundant 4-by-4 Matrices 75

3.3 Counting Eulerian Orientations in 4-Regular Planar Graphs 79

3.4 Redundant Signatures with Non-Singular Compressed Matrices 84

3.5 A Unary Interpolation Lemma 94

3.6 Pl-Holant Dichotomy for a Symmetric Arity 4 Signature 96

3.7 Vanishing Signatures Revisited 103

4 Planar Counting CSP 120

4.1 Domain Pairing 122

4.2 Mixing of Tractable Signatures 127

4.3 Pinning for Planar #CSP 132

4.4 Planar #CSP Dichotomy 141

5 A Closer Look at Tractable Signatures 145

5.1 Characterization of $\mathcal{A}$-transformable Signatures 146

5.2 Characterization of $\mathcal{P}$-transformable Signatures 152

5.3 Characterization of $\mathcal{M}$-transformable Signatures 154

5.4 Related Hardness Results 160

5.5 Hardness results for the Inductive Step 171

5.6 Single Signature Dichotomy 179

6 Dichotomy for Holant Problems 183

6.1 Another Planar Tractable Case 185

6.2 Complementing Hardness Results 196

6.3 Mixing $\mathcal{M}_4$ and $\mathcal{P}_2$ with Other Signatures 204

6.4 #PM in Planar Hypergraphs 210

6.5 The Full Dichotomy 211

7 Anti-Ferromagnetic 2-Spin Systems 217

7.1 Definitions and Backgrounds 217

7.2 The Self-Avoiding Walk Tree 222

7.3 The Potential Method 227

7.4 A Degree Dependent Potential Function 237
7.5 An Improved Potential Function 241
7.6 Monotonicity of the Uniqueness 249
7.7 Uniqueness Thresholds 256
7.8 Correlation Decay at the Threshold 266
7.9 Concluding and Bibliographic Remarks 269

8 Phase Transitions and Computational Hardness 272
8.1 Phase Transitions in Anti-Ferromagnetic Systems 272
8.2 Spin Systems in Bipartite Graphs 276
8.3 Balancing Nearly-Independent Phase-Correlated Spins 279
8.4 Symmetry Breaking 283
8.5 The Reduction from #BIS 285

9 Ferromagnetic 2-Spin Systems 293
9.1 The Uniqueness Condition on Regular Trees 294
9.2 The Potential Method for Ferromagnetic Systems 297
9.3 Bounded Degree Graphs 300
9.4 General Graphs 301
9.5 Correlation Decay Beyond λ_c 304
9.6 Complementary Hardness Results 307

10 Complex Weighted Ising Models 309
10.1 Approximating Complex Numbers 309
10.2 Problem Definitions and Results 310
10.3 The Tutte polynomial 317
10.4 Series and Parallel Compositions 318
10.5 Hardness Results for the Ising Model 321
10.6 Quantum Circuits and Counting Complexity 334
10.7 Complex Ising with External Fields 346

Bibliography 358

List of Tables

1.1 List of all non-degenerate affine signatures 19

List of Figures

1.1 An $\mathcal{F}$-gate with 5 dangling edges 16

2.1 A triangular gadget 37
2.2 The gadget for $[0, 1, 0, 0, 0]$ and $[1, 1, 0, 0, 0]$ 37
2.3 The gadget to construct $[1, 1, 1, 0]$ 41
2.4 An asymmetric gadget for domain pairing 41
2.5 The gadget to construct $[1, 0, 0, 0, 1]$ from $[0, 0, 1, 0, 0]$ 43
2.6 The gadget to construct $[1, 0, 0, 0, 1]$ from $[0, 0, 1, 0, 1]$ 43
2.7 The signature inner product 50
2.8 Simulating $[1, 0]$ using $[1, 0, 0]$ or $[1, 1, 0]$ 57
2.9 A tetrahedron gadget 58
2.10 The gadget for $[1, 0, 0, 0, 1, 0]$ 58

3.1 Degenerate vanishing gadget 69
3.2 Entry movement for rotations of arity 4 signatures 78
3.3 Medial graphs 80
3.4 Planar tetrahedron gadget 82
3.5 Recursive construction 83
3.6 Recursive construction to approximate $[1, 0, \frac{1}{3}, 0, 1]$ 85
3.7 Recursive construction to interpolate $\mathtt{id}_4$ 88
3.8 Unary recursive construction 96
3.9 Binary recursive construction 97
3.10 Find a planar pairing 99

3.11 Reducing from arity 3 to arity 4 107
3.12 A binary sequence 108
3.13 Arity 4 gadgets 113
3.14 Recursive construction to interpolate a signature r' 113

4.1 Venn diagram for Pl-#CSP 123
4.2 Domain pairing gadget 125
4.3 Binary gadget for $[0, 1, 0, \ldots, 0]$ 131
4.4 Arity 4 gadget for $[a, 0, 0, 0, b, c]$ 133
4.5 Binary gadget for $[1, 0, x]$ and $[1, 0, x]$ 135

5.1 Relationship among $\mathcal{A}$-, $\mathcal{P}$-, and $\mathcal{M}$-transformable signatures 158
5.2 Realizing $[1, 0, i]$ 163
5.3 Reduction to $\mathcal{M}_2 \setminus \mathcal{P}_2$ 170
5.4 Two gadgets in the Z basis 171
5.5 Arity-4 gadgets for $\overline{g}$ 174
5.6 Matrix entries movement by a partial negation 174
5.7 Two arity 4 gadgets 174

6.1 An example of a E_6-block 186
6.2 Arity 4 E_k-blocks 187
6.3 EO-Eq-4-blocks 191
6.4 Two types of the “wheel” structure 193
6.5 Degeneracies in the wheel structure 195
6.6 Two gadgets for $\textsc{ExactOne}_3$ and $=_3$ in the Z basis 197
6.7 Two gadgets combined together 197
6.8 Reduction to a graph homormorphism problem 199
6.9 The grid gadget 201
6.10 Combining two “grid” gadgets 201
6.11 A gadget to realize $\overline{g}$ 205

7.1 The universally strictly uniqueness region of the $\lambda = 1$ plane. 226

7.2 Commutative diagram between F_d and G_d 229

7.3 Monotonicity vs. non-monotonicity of the contraction ratio 251

10.1 Complexity of Complex Weighted Ising Models 313

10.2 Gate H 335

10.3 Two quantum gates U_1 and U_2 composed together 335

10.4 An $\mathbf{IQP}_{1,2}(\theta)$ circuit 338

10.5 Equivalence between Ising instances and $\mathbf{IQP}_{1,2}(\theta)$ circuits 338

10.6 The equivalent Ising instance to the circuit in Figure 10.4 342

Abstract

We study the computational complexity of counting problems, such as computing the partition functions, in both the exact and approximate sense. In the first part of the dissertation, we classify exact counting problems. We show a dichotomy theorem for Holant problems defined by any set of symmetric complex-valued functions on Boolean variables in both general and planar graphs. Problems are classified into three classes: those that are **P**-time solvable over general graphs; those that are **P**-time solvable over planar graphs but #**P**-hard over general graphs; those that remain #**P**-hard over planar graphs. It has been shown that in many other contexts, holographic algorithms with matchgates capture all counting problems in the second class. A surprising result is that we found a new class of tractable problems in the same class, but cannot be captured by holographic algorithms with matchgates. In the course of proving this dichotomy theorem, we also classify parity Holant problems and #CSP defined by any set of symmetric complex-valued functions on Boolean variables.

Then we focus on approximating partition functions of 2-spin systems, including the famous Ising model as a special case. We show a fully polynomial-time approximation scheme (FPTAS) for anti-ferromagnetic 2-spin systems up to the tree uniqueness threshold. There is no such algorithm beyond the threshold unless **NP** = **RP** [SS14]. We also generalize this hardness result to bipartite graphs, with the exception that the Ising model without fields is approximable in bipartite graphs. This hardness result helps to establish some new imapproximability results for ferromagnetic 2-spin systems [LLZ14a]. To complement those, we give near-optimal FPTAS in certain regions of ferromagnetic 2-spin systems. Furthermore, we go beyond non-negative real weights, and classify the computational complexity of the Ising model with complex weights. Using such results, we draw conclusions about strong simulation of certain quantum circuits.

Chapter 1

Introduction

1.1 Counting Problems and Complexity Classification

In this dissertation we study counting problems. A canonical example is #SAT, namely to count the number of satisfying assignments of CNF formulas. #SAT can be viewed as a special case in a more general framework, called *counting constraint satisfaction problems* (#CSP). An instance of #CSP contains variables and constraints, and the goal is to count the number of assignments satisfying all constraints. An equivalent formalism is to sum the weights over all possible assignments, where the weight is 1 if it is satisfying and 0 otherwise. Moreover, we may decompose this weight into a product over all constraints, maintaining the same semantics. Thus this sum-of-products quantity counts the number of satisfying assignment, which is usually called the partition function. Recasting in these terms, it is natural to generalize to functions taking values in $\mathbb{Q}$, $\mathbb{R}$, or $\mathbb{C}$, rather than just $\{0, 1\}$. This is a really powerful framework, capable of expressing many problems of counting local combinatorial structures, such as independent sets or vertex covers.

The name "partition function" originated from the statistic physics literature. Its study was initiated even before the formalization of polynomial time algorithms and computational complexity. Many models have been proposed to study the interaction among particles, such as the Ising model [Isi25]. The partition function is a crucial quantity for these models, as it encodes a lot of information regarding the model. These statistical models later found a lot of applications in computer science. Examples range from factor graphs in machine learning

[WJ08], signal processing [For01], to the classical simulation of quantum computation [Val02b, JV14]. The significance of these models is that they achieve the maximum entropy among all models that are consistent with the data. After learning such models, one may want to draw conclusions according to them. This is called statistical inference, and it often amounts to computing the marginal probability, expectation, or entropy of random variables. These tasks are equivalent to the evaluation of partition functions, and thus are essentially counting problems.

Efficient algorithms of evaluating partition functions would have a lot of theoretical and practical significance, but do they really exist? If there is not an universal one, then for what problems do efficient algorithms exist? The formalization of this question dates back to 1979, in which year Valiant [Val79b, Val79a] proposed the counting complexity class #**P** to capture the apparent intractability of many natural counting problems. The class #**P** is the counting counterpart to **NP**, that is, to count the number of certificates which can be verified in polynomial time. Just like **NP**-hard problems, #**P**-hard problems are the most difficult ones in #**P**. It is believed that #**P**-hard problems, including the aforementioned #Sat, do not admit efficient algorithms. It is not surprising that #Sat is #**P**-hard, since its decision version is **NP**-hard. However, what surprises people is that problems, such as counting the number of perfect matchings (#PM), whose decision counterparts have efficient algorithms, are still possible to be #**P**-hard [Val79a].

One ultimate goal we pursue is to understand which problems are easy to count, and which are not. In other words, we want to classify the computational complexity of counting problems. There can be several different meanings of an easy problem. The strongest sense is that we have an efficient algorithm to compute the exact answer. This is the most desirable but least likely one. In many contexts, we are content with good approximations. On the other hand, in the absence of an separation between #**P** and **FP** (the function counterpart of **P**), #**P**-hardness becomes our standard notion of computational intractability. After all, #**P**-hard problems are very unlikely to have efficient algorithms. We also work under other standard complexity assumptions, such as **NP** $\neq$ **RP**, when we are talking about the impossibility of efficient algorithms in a weaker sense, such as approximations.

Some of the most intriguing efficient algorithms again come from statistical physics. In particular, the Fisher-Kasteleyn-Temperley (FKT) algorithm [TF61, Kas61, Kas67] is a classical gem that counts perfect matchings over planar graphs in polynomial time. This was an important milestone in a decades-long research program in statistical physics to determine what is known as Exactly Solved Models [Bax82, Isi25, Ons44, Yan52, YL52, LY52, TF61, Kas61, Kas67, Lie67, LS81, Wel93].

For four decades, the FKT algorithm stood as *the* polynomial-time algorithm for any counting problem over planar graphs that is #**P**-hard over general graphs. Then Valiant introduced *matchgates* [Val02b, Val02a] and *holographic* reductions to the FKT algorithm [Val08, Val06]. These reductions differ from classical ones by introducing quantum-like superpositions. This novel technique extended the reach of the FKT algorithm and produced polynomial-time algorithms for a number of problems for which only exponential-time algorithms were previously known.

Since the new polynomial-time algorithms appear so exotic and unexpected, and since they solve problems that appear so close to being #**P**-hard, they challenge our faith in the well-accepted conjecture that $\mathbf{P} \neq \mathbf{NP}$. Quoting Valiant [Val06]: "The objects enumerated are sets of polynomial systems such that the solvability of any one member would give a polynomial time algorithm for a specific problem. . . . the situation with the $\mathbf{P} = \mathbf{NP}$ question is not dissimilar to that of other unresolved enumerative conjectures in mathematics. The possibility that accidental or freak objects in the enumeration exist cannot be discounted if the objects in the enumeration have not been studied systematically." Indeed, if any "freak" object exists in this framework, it would collapse #**P** to **FP**.

Therefore, over the past 10 to 15 years, this technique has been intensely studied in order to gain a systematic understanding to the limit of the trio of holographic reductions, matchgates, and the FKT algorithm [Val02a, CC07, CCL09, CL10, Val10, CL11a, LMN13, Mor11, MM13, CG14]. Without settling the **FP** versus #**P** question, the best hope is to achieve a complexity classification. This program finds its sharpest expression in a complexity dichotomy theorem, which classifies *every* problem expressible in a framework as either solvable in polynomial time or #**P**-hard, with nothing in between.

Assuming that $\mathbf{FP} \neq \#\mathbf{P}$, such a dichotomy does not hold for the whole class of #**P**, due to an easy adaption of Ladner's theorem [Lad75]. Nevertheless all known intermediate problems are unnatural and artificial. On the other hand, many natural problems are expressible as, for example, #CSPs, which we have mentioned at the beginning of this section. Hence we usually want to work within certain frameworks, such as the Tutte polynomial or #CSP, in order not to ask overly general and vague questions. Unfortunately, there are some important problems, like #PM, that are difficult to express as a point on the Tutte polynomial or as a #CSP. In fact, #PM is provably beyond the reach of the special case of vertex assignment models [FLS07, DGL$^+$12, Sch13]. However, this is the problem for which FKT was designed, and is the basis of Valiant's matchgates and holographic reductions. To address this issue, a refined framework, called Holant problems [CLX11a], was proposed. It is an edge assignment model. It naturally encodes and expresses #PM as well as Valiant's matchgates and holographic reductions, which makes it the proper framework to study the power of holographic algorithms. It is also more general than #CSP in the sense that a complete complexity classification for Holant problems implies one for #CSP.

The first success in classifying the complexity of Holant problems was achieved in the parity setting. The goal is to compute the partition function modulo 2, and the hardness is captured by $\oplus$**P**-hardness. In a joint paper with Lu and Valiant [GLV13], we showed a complete dichotomy for symmetric Boolean functions. An interesting phenomenon is observed in this work, that is, for some functions, the partition function is always even, which makes it 0 mod 2. We call them vanishing signatures. Later, it turned out that the vanishing signatures are the last missing piece in a complete computational complexity dichotomy of complex weighted Holant problems [CGW13].

The dichotomy in [CGW13], despite being the culminating result of a long line of research [CLX11a, CK12, CK13, CHL12, HL12], does not answer the question raised by Valiant, since it does not tell us about the complexity when instances are restricted to planar graphs. With respect to planar instances, a strong theme has emerged in research. For a wide variety of problems, such as those expressible as a #CSP, holographic reductions to the FKT algorithm is a *universal* technique for turning problems that are #**P**-hard in general to **P**-time solvable over planar graphs. In fact, a preponderance of evidence suggests the following putative classification of all counting problems defined by local constraints into *exactly* three categories:

(1) those that are **P**-time solvable over general graphs;

(2) those that are **P**-time solvable over planar graphs but #**P**-hard over general graphs;

(3) those that remain #**P**-hard over planar graphs.

Moreover, category (2) consists of precisely those problems that are holographically reducible to the FKT algorithm. This theme is so strong that it has become an intuitive and trusty guide for us when we investigate unknown problems and plan proof strategies. Many of the results were proved in this way. However, one is still left wondering whether the FKT algorithm is *universal*, or more precisely, is the combined algorithmic power of the trio sufficient to capture all tractable problems over planar graphs that are intractable in general? Supporting evidences for this putative classification date back to the classification of the Tutte polynomial [Ver91, Ver05]. It has also been an unfailing theme when classifying spin systems [Kow10, CKW12] and #CSP [CLX10, GW13].

What catches us off guard is when we classify for the first time the complexity of Holant problems over planar graphs in an recent paper [CFGW15], there are new planar tractable problems not expressible by a holographic reduction to matchgates and FKT. To the best of our knowledge, this is the first *primitive* extension since FKT to a problem solvable in **P** over planar instances but #**P**-hard in general. Furthermore, our dichotomy theorem says that this completes the picture: there are no more undiscovered extensions for problems expressible in this framework, unless #**P** collapses to **FP**. In particular, the putative form of the planar Holant dichotomy is *false*. This complexity dichotomy generalizes both the dichotomy for Holant [HL12, CGW13] and the dichotomy for planar #CSP [CLX10, GW13].

This new tractable class can be described as orientation problems. Given a planar graph, we allow two kinds of vertices. The first kind can be either a sink or a source while the second kind only allows one incoming edge. The goal is to compute the number of orientations satisfying these constraints. This problem can be expressed in the Holant framework under a transformation by $\left[\begin{smallmatrix} 1 & 1 \\ i & -i \end{smallmatrix}\right]$. It can be shown that this is equivalent to the Holant problem on the (planar) edge-vertex incidence graph where we assign the DISEQUALITY function to every edge, and to each vertex, we assign either the EQUALITY function or the EXACTONE function. Suppose vertices assigned EQUALITY functions all have degree k. If $k = 2$, then this problem can be solved by FKT. We show that this problem is #**P**-hard if $k = 3$ or $k = 4$, but is tractable again if $k \geqslant 5$. The algorithm involves a recursive procedure that simplifies the instance until it can be solved by known algorithms, including FKT. The algorithm crucially uses global topological properties of a planar graph, in particular Euler's characteristic formula. If the graph is not planar, then this algorithm does not work, and indeed the problem is #**P**-hard over general graphs.

More generally, we allow vertices of arbitrary degrees to be assigned EQUALITY. If all the degrees are at most 2, then the problem is tractable by the FKT algorithm. Otherwise, the complexity depends on the greatest common divisor (gcd) of the degrees. The problem is tractable if gcd $\geqslant 5$ and #**P**-hard if gcd $\leqslant 4$. It is worth noting that the criterion for tractability is not a degree lower bound. Moreover, the planarity assumption and the degree rigidity pose a formidable challenge in the hardness proofs for gcd $\leqslant 4$.

If the graph is bipartite with EQUALITY functions assigned on one side and EXACTONE functions on the other, then this is the problem of #PM over hypergraphs with planar incidence graphs. Our results imply that the complexity of this problem depends on the gcd of the hyperedge sizes. The problem is computable in polynomial time when gcd $\geqslant 5$ and is #**P**-hard when gcd $\leqslant 4$ (assuming there are hyperedges of sizes at least 3). For more details, see Section 6.4 and Theorem 6.16.

Coming back to the challenge of the **P** vs. **NP** question posed by Valiant's holographic algorithms, we venture the opinion that the dichotomy theorem provides a satisfactory answer. Indeed, it would be difficult to conceive a world where #**P** is **FP**, and yet all this algebraic theory can somehow maintain a consistent, sharp but faux division where there is none. (Consider the following Gedankenexperiment: #**P** is really equal to P, but the Supreme Fascist keeps scores on how much of #**P** we have learned to be in P. For every problem in this broad class that is yet unknown to be in P the SF lets we prove it #**P**-hard—a superfluous notion really. Nevertheless for every problem in the class known to be in P, the SF makes sure our proof method for #**P**-hardness on that problem fails, thus preventing one from making the ultimate discovery.)

In the first part of this dissertation, our main goal is to show the dichotomy theorem for Holant problems, for both general and planar graphs. This is summarized as Theorem 6.17. In Chapter 1, starting from Section 1.3, we give necessary definitions and useful background knowledge. In Chapter 2, we show the dichotomy for Holant problems modulo 2. In Chapter 3, we give a dichotomy for a single arity-4 function. In Chapter 4, a dichotomy for planar #CSP is shown. In Chapter 5, we examine more closely about what holographic transformations can do for tractable problems, and show some related hardness results. Finally in Chapter 6, we give the main dichotomy, Theorem 6.17. Results in Chapter 2 are joint work with Pinyan Lu and Leslie G. Valiant [GLV13]. To ensure the best presentation, results taken from four papers [CGW13, GW13, CGW14, CFGW15] are blended and reported in Chapters 3, 4, 5, and 6. These are joint work with Jin-Yi Cai, Zhiguo Fu, and Tyson Williams.

1.2 Approximate Counting

One thing that seems unfortunate, indicated by dichotomy theorems described in the previous section, is that algorithms for exact counting are rather scarce. The vast majority of counting problems are #**P**-hard. However, the pursuit of efficient counting algorithms does not stop there. In practice, people use algorithms like belief propagation or the junction-tree algorithm, which either works as intended only in restricted settings, or requires exponential running time on certain instances.

For provable efficient algorithms, a celebrated result is by Jerrum et al. [JS89, JSV04], which gives a fully polynomial time randomized approximation scheme (FPTAS) for the permanent of a non-negative matrix. Computing the permanent is the first nontrivial example of #**P**-hardness given by Valiant [Val79a]. The algorithm is based on a Markov Chain Monte Carlo (MCMC) sampling scheme, a technique which has its own long and fruitful history in statistical physics. Since then, this MCMC technique is studied extensively and broadly. Examples include, but are not limited to, approximately counting the number of proper colorings [Vig00], independent sets [DG00b], and to approximate the volume of a convex body [DFK91], etc.

Powerful as they are, MCMC algorithms do not always give the optimal bounds on many problems, at least not with the current analysis tools available. For example, the best Markov chain to sample independent sets works when the graph has a degree bound of 4 [DG00b], whereas a breakthrough result by Weitz [Wei06] showed that it is possible to approximately count the number of independent sets when the degree bound is 5. Weitz's idea involves a novel technique called correlation decay or strong spatial mixing, which asserts that the correlation among variables of the system decays exponentially fast in distance. When this is the case, we can recursively compute the marginal probability by unfolding the original instance into an exponentially large tree with early truncation as the influence below is small. This idea of approximate counting without sampling is independently introduced by Bandyopadhyay and Gamarnik [BG08] as well.

Counting independent sets is a special case of the partition function of the hardcore gas model, in which the same independence constraint is applied and each vertex is assigned a weight (also called fugacity or activity) when it is chosen. The hardcore gas model, in turn, is a special case of the more general 2-state spin systems. Spin systems are some of the most fundamental statistical physics systems. The goal is to model nearest neighbour interactions. For a 2-spin system and an underlying graph, we assign "+" or "-" spins to vertices. On each edge there is an interaction between the two endpoints, depending on their spins. This interaction may be contractive, where the system is called ferromagnetic, or repulsive, where it is called anti-ferromagnetic. The system is often equipped with an external field as well, which favors toward one spin or the other. We use a matrix $\begin{bmatrix} \beta & 1 \\ 1 & \gamma \end{bmatrix}$ to denote the symmetric edge interaction, and a vector $\begin{bmatrix} \lambda \\ 1 \end{bmatrix}$ to denote the external field. Then ferromagnetic systems have $\beta\gamma > 1$ and anti-ferromagnetic $\beta\gamma < 1$. For the hardcore model, the edge interaction is $\begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix}$, banning all $(+,+)$ edges, and the external field is $\begin{bmatrix} \lambda \\ 1 \end{bmatrix}$, where λ is the vertex weight when it is chosen (or assigned "+"). Another famous model in this framework is the Ising model [Isi25], where the edge interaction is totally symmetric $\begin{bmatrix} \beta & 1 \\ 1 & \beta \end{bmatrix}$.

One important phenomenon statistical physicists discovered is the phase transition, namely that these systems undergo a drastic change of behaviors as parameters change. The system may have disordered or ordered phases, depending on the uniqueness of the so-called Gibbs measures. The breakthrough by Weitz [Wei06] first illustrated that it is possible to get an FPTAS for the hardcore gas model all the way up to this uniqueness threshold in infinite regular trees (also known as the Bethe lattice or the Cayley tree). Weitz's algorithm is later generalized to all anti-ferromagnetic 2-spin models by Sinclair et al. [SST12] and Li et al. [LLY12, LLY13]. Some of my unpublished results are combined with [LLY12, LLY13] for a journal submission [GLLY15].

On the other hand, when the correlation decay fails, there are gadgets to realize some long range correlation. In these gadgets, typical configurations are ordered, in the sense that they favor one part of the graph or the other. We then can use these ordered phases to simulate states of other problems, and show that the original problem is hard to even approximate. The first tight result of this sort is due to Sly [Sly10], building upon a series of work [DFJ02, MWW09], showing that there is no FPRAS unless **NP** = **RP** for the hardcore gas model in a small interval beyond the uniqueness threshold in infinite regular trees. It is later improved by [GGS⁺14] for the hardcore model to all fields beyond the uniqueness threshold for most degrees, and by us [CCGL12] for all anti-ferromagnetic 2-spin systems to beyond the uniqueness threshold multiplied by a constant factor. It is finally confirmed [GŠV12, SS14] that this hardness always holds beyond the uniqueness threshold. Thus, a beautiful connection between the computational complexity transition and the phase transition of statistical systems has been established.

The 2-state spin models can be viewed as the simplest case of #CSP, namely with only one binary symmetric function, and a possibly unary function with respect to the field. Similar to exact counting, a lot of research has been devoted to classify the computational complexity of approximating counting problems. An important open question emerges from these work, which is counting independent sets in bipartite graphs (#BIS). There is no known efficient algorithm for #BIS, nor is there any hardness reduction. It is conjectured that neither is the case [DGGJ03]. For instance, there is a whole class of #CSP problems that are equivalent to #BIS in approximation [BDG⁺13, CDG⁺15], and approximating the ferromagnetic Potts model (a higher domain version of the Ising model) is shown to be at least as hard as #BIS [GJ12a]. The intriguing feature of #BIS is that all previous hardness reductions break, due to the bipartite structure. Striving for a better understanding of #BIS, we [CGG⁺14] showed that beyond the uniqueness threshold, anti-ferromagnetic 2-spin systems in bipartite graphs are no easier to approximate than #BIS, except for Ising models without external field, which can be reduced to ferromagnetic Ising models and thus have FPRASes [JS93]. In particular, #BIS with degree bound 6 is as hard to approximate as #BIS itself. Our results would hopefully shed some light on the approximation complexity of #BIS in future.

Although not solving the problem #BIS, our result in [CGG⁺14] does help classifying ferromagnetic 2-spin models ($\beta\gamma > 1$). All ferromagnetic 2-spin models are #BIS-easy [GJ07]. Thus #BIS-hardness is the best lower bound one should hope for. Building upon our results in

[CGG$^+$14], Liu et al. [LLZ14a] showed some unexpected #BIS-hardness for ferromagnetic 2-spin models, especially for the case of $\beta \leqslant 1 < \gamma$. Recently, we [GL15] have obtained a complementary correlation decay based algorithm for $\beta \leqslant 1 < \gamma$ as well, for external fields $\lambda \leqslant \left(\frac{\gamma}{\beta}\right)^{\frac{\Delta_c+1}{2}}$, where $\Delta_c = \frac{\sqrt{\beta\gamma}+1}{\sqrt{\beta\gamma}-1}$, improving upon previously best known bound $\lambda \leqslant \frac{\gamma}{\beta}$ [LLZ14a]. This is almost optimal, since if we allow all external fields in the range of $(0, \lambda)$ for some $\lambda > \left(\frac{\gamma}{\beta}\right)^{\frac{\lceil \Delta_c \rceil+1}{2}}$, then the problem is #BIS-hard [LLZ14a, CGG$^+$14]. Interestingly, we also provide evidence that neither $\left(\frac{\gamma}{\beta}\right)^{\frac{\Delta_c+1}{2}}$ is the tight bound for approximability, nor is $\left(\frac{\gamma}{\beta}\right)^{\frac{\lceil \Delta_c \rceil+1}{2}}$ tight for the failure of correlation decay.

Prior to this point, all results mentioned in this section are about non-negative weighted counting problems. In a recent paper [GG14], we studied the Ising model with complex parameters. In addition to its intrinsic mathematical interest, the complex weighted Ising model has connections to a quantum complexity class called **IQP** [SB09, BJS11], which stands for "Instantaneously Quantum Polynomial-time". What we are interested in is to approximate the norm and the argument of the partition function. We show that for most parameters in the absence of external fields, these problems are **NP**-hard to approximate. In fact, we also show an interesting stronger results that for many of these parameters, even approximation is #**P**-hard. (Note that all problems in #**P** can be approximated with an **NP** oracle [DGGJ03].) We also gave a complete classification when the edge interaction and the external field are both roots of unity, cases relating to **BQP**. Our work implies some new hardness results about classically simulating **IQP** circuits in the strong sense.

The second part of this dissertation, starting from Chapter 7, focuses upon approximate counting, and is organized as follows. In Chapter 7 we give necessary backgrounds and define problems, as well as show the algorithm for all anti-ferromagnetic 2-spin systems up to the uniqueness threshold. In Chapter 8 we explain the complementary hardness results, and generalize them to bipartite graphs. In Chapter 9 we present the improved algorithm for ferromagnetic 2-spin systems. In Chapter 10 complex weighted Ising models are studied. Results presented in Chapter 7 are mostly from [GLLY15] and are joint work with Liang Li, Pinyan Lu, and Yitong Yin. Results in Chapter 8 are published in [CGG$^+$14] and are joint work with Jin-Yi Cai, Andreas Galanis, Leslie Ann Goldberg, Mark Jerrum, Daniel Štefankovič, and Eric Vigoda.

Chapter 9 is summarized in [GL15], joint work with Pinyan Lu. Results in Chapter 10 are from [GG14], joint work with Leslie Ann Goldberg.

1.3 Problems and Definitions

In the rest of this chapter we will define problems to be studied in the context of exact counting and prove some necessary background knowledge. For approximate counting, see Chapter 7 and later.

The framework of Holant problems is defined for functions mapping any $[q]^n \to R$ for a finite q and some commutative semiring R. We investigate Boolean Holant problems, that is, the domain size q is 2. We are interested in complex-weighted functions $[2]^n \to \mathbb{C}$ throughout this dissertation, as well as parity functions $[2]^n \to \mathbb{Z}_2$ in Chapter 2. For consideration of models of computation, complex functions take algebraic values.

We allow multigraphs, that is, graphs may have self-loops and parallel edges. A graph without self-loops or parallel edges is a *simple* graph. Fix a set of local constraint functions $\mathcal{F}$. A *signature grid* $\Omega = (G, \pi)$ consists of a graph $G = (V, E)$, where π assigns to each vertex $v \in V$ and its incident edges some $f_v \in \mathcal{F}$ and its input variables. We say that Ω is a *planar signature grid* if G is planar, where the variables of f_v are ordered counterclockwise starting from an edge specified by π. The Holant problem on instance Ω is to evaluate

$$\text{Holant}(\Omega; \mathcal{F}) = \sum_{\sigma} \prod_{v \in V} f_v(\sigma \mid_{E(v)}),$$

a sum over all edge assignments $\sigma : E \to \{0, 1\}$, where $E(v)$ denotes the incident edges of v and $\sigma \mid_{E(v)}$ denotes the restriction of σ to $E(v)$. We write G in place of Ω when π is clear from the context. We also sometimes write Holant_Ω instead of $\text{Holant}(\Omega; \mathcal{F})$ when $\mathcal{F}$ is clear from the context.

A Holant problem is defined by a set $\mathcal{F}$ of signatures.

Name $\text{Holant}(\mathcal{F})$

Instance A *signature grid* $\Omega = (G, \pi)$.

Output $\text{Holant}(\Omega; \mathcal{F})$.

The problem Pl-Holant($\mathcal{F}$) is defined similarly using a planar signature grid.

We allow $\mathcal{F}$ to be an infinite set. For Holant($\mathcal{F}$) (or Pl-Holant($\mathcal{F}$)) to be tractable, the problem must be computable in polynomial time even when there exists an effective description of every signature in $\mathcal{F}$, and these descriptions are considered part of the input. In contrast, we say Holant($\mathcal{F}$) is #**P**-hard if there exists a finite subset of $\mathcal{F}$ for which the problem is #**P**-hard. We say a signature set $\mathcal{F}$ is tractable (resp. #**P**-hard) if the corresponding counting problem Holant($\mathcal{F}$) is tractable (resp. #**P**-hard). Similarly for a signature f, we say f is tractable (resp. #**P**-hard) if $\{f\}$ is. We follow the usual conventions about polynomial time Turing reduction $\leqslant_T$ and polynomial time Turing equivalence $\equiv_T$.

A function f_v can be represented by listing its values in lexicographical order as in a truth table, which is a vector in $\mathbb{C}^{2^{\deg(v)}}$, or as a tensor in $(\mathbb{C}^2)^{\otimes \deg(v)}$. A function is *symmetric* if its output depends only on the Hamming weight of its input. A symmetric function f of arity n can be expressed as $[f_0, f_1, \ldots, f_n]$, where f_w is the value of f on inputs of Hamming weight w. This is the signature of f, and we also use the two words “function” and “signature” interchangeably. We study symmetric functions in most cases, but we do go beyond if necessary.

An example of symmetric signatures is the EQUALITY signature $=_n$ of arity n, which can be expressed as $[1, 0, \cdots, 0, 1]$. Let $\mathcal{EQ} = \{=_n \mid n \in \mathbb{N}\}$. Another example is the EXACTONE signature EO_n of arity n, which can be expressed as $[0, 1, 0, \cdots, 0]$. Let $\mathcal{EO} = \{\mathrm{EO}_n \mid n \in \mathbb{N}\}$. Then Holant($\mathcal{EO}$) is the problem of counting perfect matchings.

A signature f of arity n is *degenerate* if there exist unary signatures $u_j \in \mathbb{C}^2$ $(1 \leqslant j \leqslant n)$ such that $f = u_1 \otimes \cdots \otimes u_n$. It is equivalent to replace f by $u_1, \ldots, u_n$, ordered appropriately. If a degenerate signature f is symmetric, then there exists an unary u such that $f = u^{\otimes n}$.

Replacing a signature $f \in \mathcal{F}$ by a constant multiple cf, where $c \neq 0$, does not change the complexity of Holant($\mathcal{F}$). It introduces a global nonzero factor to Holant($\Omega; \mathcal{F}$). We may say we obtain a signature f when in fact we have obtained a signature cf for some $c \neq 0$.

1.4 Useful Reductions

Holographic Reduction

One key technique for Holant problems is holographic reductions. To introduce the idea, it is convenient to consider bipartite graphs. For a general graph, we can always transform it into a bipartite graph while preserving the Holant value as follows. For each edge in the graph, we replace it by a path of length two. (This operation is called the *2-stretch* of the graph and yields the edge-vertex incidence graph.) Each new vertex is assigned the binary EQUALITY signature $(=_2) = [1, 0, 1]$.

We use $\text{Holant}\,(\mathcal{F} \mid \mathcal{G})$ to denote the Holant problem over signature grids with a bipartite graph $H = (U, V, E)$, where each vertex in U or V is assigned a signature in $\mathcal{F}$ or $\mathcal{G}$, respectively. Signatures in $\mathcal{F}$ are considered as row vectors (or covariant tensors); signatures in $\mathcal{G}$ are considered as column vectors (or contravariant tensors) (see, for example [DP91]). Similarly, $\text{Pl-Holant}\,(\mathcal{F} \mid \mathcal{G})$ denotes the Holant problem over signature grids with a planar bipartite graph.

For a 2-by-2 matrix T and a signature set $\mathcal{F}$, define $T\mathcal{F} = \{g \mid \exists f \in \mathcal{F} \text{ of arity } n,\ g = T^{\otimes n} f\}$, and similarly for $\mathcal{F}T$. Whenever we write $T^{\otimes n} f$ or $T\mathcal{F}$, we view the signatures as column vectors; similarly for $f T^{\otimes n}$ or $\mathcal{F}T$ as row vectors. In the special case that $T = \left[\begin{smallmatrix} 1 & 1 \\ 1 & -1 \end{smallmatrix}\right]$, we use $\widehat{\mathcal{F}}$ to denote $T\mathcal{F}$.

Let T be an invertible 2-by-2 matrix. The holographic transformation defined by T is the following operation: given a signature grid $\Omega = (H, \pi)$ of $\text{Holant}\,(\mathcal{F} \mid \mathcal{G})$, for the same bipartite graph H, we get a new grid $\Omega' = (H, \pi')$ of $\text{Holant}\left(\mathcal{F}T \mid T^{-1}\mathcal{G}\right)$ by replacing each signature in $\mathcal{F}$ or $\mathcal{G}$ with the corresponding signature in $\mathcal{F}T$ or $T^{-1}\mathcal{G}$.

Theorem 1.1 (Valiant's Holant Theorem [Val08])**.** *If* $T \in \mathbb{C}^{2\times 2}$ *is an invertible matrix, then we have* $\text{Holant}(\Omega; \mathcal{F} \mid \mathcal{G}) = \text{Holant}(\Omega'; \mathcal{F}T \mid T^{-1}\mathcal{G})$.

Therefore, an invertible holographic transformation does not change the complexity of the Holant problem in the bipartite setting. Furthermore, there is a special kind of holographic transformation, the orthogonal transformation, that preserves the binary equality and thus can be used freely in the standard setting.

Theorem 1.2 (Theorem 2.6 in [CLX11a])**.** *If* $T \in \mathbf{O}_2(\mathbb{C})$ *is an orthogonal matrix* (*i.e.* $TT^T = I_2$)*, then* $\text{Holant}(\Omega; \mathcal{F}) = \text{Holant}(\Omega'; T\mathcal{F})$*.*

We frequently apply a holographic transformation defined by the matrix $Z = \frac{1}{\sqrt{2}} \left[\begin{smallmatrix} 1 & 1 \\ i & -i \end{smallmatrix}\right]$ (or sometimes without the nonzero factor of $\frac{1}{\sqrt{2}}$ since this does not affect the complexity). This matrix has the property that the binary EQUALITY signature $(=_2) = [1, 0, 1]$ is transformed to $[1, 0, 1]Z^{\otimes 2} = [0, 1, 0] = (\neq_2)$, the binary DISEQUALITY signature.

By Theorem 1.1, we have that

$$\text{Holant}(\mathcal{F}) \equiv \text{Holant}\left([1, 0, 1]T^{\otimes 2} \mid T^{-1}\mathcal{F}\right)$$
$$\text{Pl-Holant}(\mathcal{F}) \equiv \text{Pl-Holant}\left([1, 0, 1]T^{\otimes 2} \mid T^{-1}\mathcal{F}\right),$$

where $T \in \mathbf{GL}_2(\mathbb{C})$ is nonsingular. This leads to the notion of $\mathcal{C}$*-transformable*.

Definition 1.3. *Let* $\mathcal{F}$ *and* $\mathcal{C}$ *be two sets of signatures. We say* $\mathcal{F}$ *is* $\mathcal{C}$*-transformable if there exists a* $T \in \mathbf{GL}_2(\mathbb{C})$ *such that* $[1, 0, 1]T^{\otimes 2} \in \mathcal{C}$ *and* $\mathcal{F} \subseteq T\mathcal{C}$*.*

The following lemma is immediate.

Lemma 1.4. *If* $\mathcal{F}$ *is* $\mathcal{C}$*-transformable, then we have the following reductions.*

$$\text{Holant}(\mathcal{F}) \leqslant_T \text{Holant}(\mathcal{C});$$
$$\text{Pl-Holant}(\mathcal{F}) \leqslant_T \text{Pl-Holant}(\mathcal{C}).$$

Clearly, if $\text{Holant}(\mathcal{C})$ or $\text{Pl-Holant}(\mathcal{C})$ is tractable, then $\text{Holant}(\mathcal{F})$ or $\text{Pl-Holant}(\mathcal{F})$ is tractable for any $\mathcal{C}$-transformable set $\mathcal{F}$.

Counting Constraint Satisfaction Problems

From the bipartite perspective, it is easy to express counting constraint satisfaction problems (#CSP) in the Holant framework. An instance of #CSP($\mathcal{F}$) has the following bipartite view. Create a vertex for each variable and each constraint. Connect a variable vertex to a constraint vertex if the variable appears in the constraint. This bipartite graph is also known as the *constraint* *graph*. Moreover, each variable can be viewed as an EQUALITY function, as it forces the same value for all adjacent edges. Under this view, we see that

$$\#\text{CSP}(\mathcal{F}) \equiv_T \text{Holant}\,(\mathcal{EQ} \mid \mathcal{F})\,,$$

where $\mathcal{EQ} = \{=_1, =_2, =_3, \dots\}$ is the set of EQUALITY signatures of all arities. By restricting to planar constraint graphs, we have the planar #CSP framework, which we denote by Pl-#CSP. The construction above also shows that Pl-#CSP$(\mathcal{F}) \equiv_T$ Pl-Holant$\,(\mathcal{EQ} \mid \mathcal{F})$.

For any positive integer d, the problem #CSP$^d(\mathcal{F})$ is the same as #CSP$(\mathcal{F})$ except that every variable appears a multiple of d times. Similarly for Pl-#CSP$^d(\mathcal{F})$. Clearly #CSP$^1(\mathcal{F})$ is the same as #CSP$(\mathcal{F})$. We also have that,

$$\#\text{CSP}^d(\mathcal{F}) \equiv_T \text{Holant}\,(\mathcal{EQ}_d \mid \mathcal{F})$$
$$\text{Pl-}\#\text{CSP}^d(\mathcal{F}) \equiv_T \text{Pl-Holant}\,(\mathcal{EQ}_d \mid \mathcal{F})\,,$$

where $\mathcal{EQ}_d = \{=_d, =_{2d}, =_{3d}, \dots\}$ is the set of EQUALITY signatures of arities that are a multiple of d. Furthermore, if $d \in \{1, 2\}$, then we have

$$\#\text{CSP}^d(\mathcal{F}) \equiv_T \text{Holant}(\mathcal{EQ}_d \cup \mathcal{F})$$
$$\text{Pl-}\#\text{CSP}^d(\mathcal{F}) \equiv_T \text{Pl-Holant}(\mathcal{EQ}_d \cup \mathcal{F}). \tag{1.1}$$

Reductions from left to right are trivial. For the other direction, we take a signature grid Ω for the problem on the right and create a bipartite signature grid Ω' for the problem on the left such that both signature grids have the same Holant value. We simply create the bipartite grid Ω'' of Ω as described earlier. Then we contract all EQUALITY signatures that are connected with each other, resulting in Ω' where EQUALITY signatures are on one side and signatures from $\mathcal{F}$ on the other. If $d = 1$, then this is an instance of #CSP$(\mathcal{F})$ (or Pl-#CSP$(\mathcal{F})$). If $d = 2$, then every EQUALITY in Ω'' is of even arity and contraction keeps parity. Hence Ω' is an instance of #CSP$^2(\mathcal{F})$ (or Pl-#CSP$^2(\mathcal{F})$).

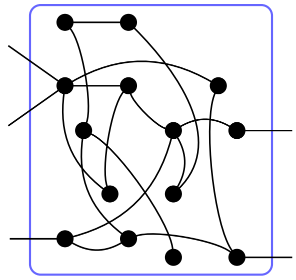


Figure 1.1: An $\mathcal{F}$-gate with 5 dangling edges.

Realization

One basic notion used extensively is realization. We say a signature f is *realizable* (or *constructible*, *implementable*) from a signature set $\mathcal{F}$ if there is a gadget with some dangling edges such that each vertex is assigned a signature from $\mathcal{F}$, and the resulting graph, when viewed as a black-box signature with inputs on the dangling edges, is exactly f. If f is realizable from a set $\mathcal{F}$, then we can freely add f into $\mathcal{F}$ while preserving the complexity.

Formally, such a notion is defined by an $\mathcal{F}$-gate [CLX10]. An $\mathcal{F}$-gate is similar to a signature grid (G, π) for Holant($\mathcal{F}$) except that $G = (V, E, D)$ is a graph with some dangling edges D. The dangling edges define external variables for the $\mathcal{F}$-gate. (See Figure 1.1 for an example.) We denote the regular edges in E by $1, 2, \ldots, m$ and the dangling edges in D by $m+1, \ldots, m+n$. Then we can define a function Γ for this $\mathcal{F}$-gate as

$$\Gamma(y_1, \ldots, y_n) = \sum_{x_1, \ldots, x_m \in \{0,1\}} H(x_1, \ldots, x_m, y_1, \ldots, y_n),$$

where $(y_1, \ldots, y_n) \in \{0,1\}^n$ is an assignment on the dangling edges and $H(x_1, \ldots, x_m, y_1, \ldots, y_n)$ is the value of the signature grid on an assignment of all edges in G, which is the product of evaluations at all internal vertices. We also call this function Γ the signature of the $\mathcal{F}$-gate.

An $\mathcal{F}$-gate is planar if the underlying graph G is a planar graph, and the dangling edges, ordered counterclockwise corresponding to the order of the input variables, are in the outer face in a planar embedding. A planar $\mathcal{F}$-gate can be used in a planar signature grid as if it is just a single vertex with the particular signature.

Using the idea of $\mathcal{F}$-gates, we can reduce one planar Holant problem to another. Suppose g is the signature of some $\mathcal{F}$-gate. Then Holant($\mathcal{F} \cup \{g\}$) $\leqslant_T$ Holant($\mathcal{F}$). The reduction is simple.

Given an instance of Holant$(\mathcal{F} \cup \{g\})$, by replacing every appearance of g by the $\mathcal{F}$-gate, we get an instance of Holant$(\mathcal{F})$. Since the signature of the $\mathcal{F}$-gate is g, the Holant values for these two signature grids are identical. This reduction clearly applies to the planar setting as well.

Although our main result is about symmetric signatures, some of our proofs utilize asymmetric signatures. When a gadget has an asymmetric signature, we place a diamond on the edge corresponding to the first input. The remaining inputs are ordered counterclockwise around the vertex.

We note that even for a very simple signature set $\mathcal{F}$, the signatures for all $\mathcal{F}$-gates can be quite complicated and expressive.

1.5 Tractable Signature Sets

We summarize several known sets of tractable Boolean functions with complex weights. The first one is very simple. If all signatures are degenerate or binary, then the problem is tractable.

For a binary signature, define its matrix as

$$M_f := \begin{bmatrix} f(00) & f(01) \\ f(10) & f(11) \end{bmatrix}. \tag{1.2}$$

Connecting f to g via one edge gives another signature h with the matrix $M_h = M_f M_g$.

Lemma 1.5. *Let $\mathcal{F}$ be a set of complex weighted symmetric signatures in Boolean variables. Then* Holant$(\mathcal{F})$ *is computable in polynomial time if all non-degenerate signatures in $\mathcal{F}$ are of arity at most 2.*

Proof. We first replace degenerate signatures by a bunch of equivalent unary signatures. Then any instance of Holant$(\mathcal{F})$ can be decomposed into paths and cycles. The Holant is a product of all paths and cycles.

For a path, we remove the two endpoints, leaving a binary signature f composed by a series of binary signatures. Compute the signature matrix M_f of f by multiplying all binary signatures along the path. Then the Holant is $v M_f u^T$, where v and u are the two unary signatures at endpoints.

For a cycle, we arbitrarily break an edge getting a path with two dangling edges. Similar to the above case, we multiply matrices of all binary signatures along this path, getting M. The trace of M is the Holant. $\square$

We further note that for a binary signature f and $T \in \mathbb{C}^{2\times 2}$, let $g = fT^{\otimes 2}$. Then

$$M_g = TM_fT^T. \tag{1.3}$$

This can be seen by viewing T as a binary, and then treating g as connecting T, f, and T^T sequentially.

Affine Signatures

Definition 1.6 (Definition 3.1 in [CLX14]). *A k-ary function $f(x_1, \dots, x_k)$ is* affine *if it has the form*

$$\lambda \cdot \chi_{Ax=0} \cdot i^{\sum_{j=1}^{n} \langle \mathbf{v}_j, x \rangle},$$

where $\lambda \in \mathbb{C}$, $x = (x_1, x_2, \dots, x_k, 1)^T$, A is a matrix over $\mathbb{F}_2$, $\mathbf{v}_j$ is a vector over $\mathbb{F}_2$, and χ is a 0-1 indicator function such that $\chi_{Ax=0}$ is 1 if and only if $Ax = 0$. Note that the dot product $\langle \mathbf{v}_j, x \rangle$ is calculated over $\mathbb{F}_2$, while the summation $\sum_{j=1}^{n}$ on the exponent of $i = \sqrt{-1}$ is evaluated as a sum mod 4 of 0-1 terms. We use $\mathcal{A}$ to denote the set of all affine functions.

The matrix A defines an affine space which is the support of the signature f (and hence the name). Notice that there is no restriction on the number of rows in the matrix A. It is permissible that A is the zero matrix so that $\chi_{Ax=0} = 1$ holds for all x. An equivalent way to express the exponent of i is as a quadratic polynomial where all cross terms have an even coefficient (cf. [CCLL10]).

It is known that the set of non-degenerate symmetric signatures in $\mathcal{A}$ is precisely the nonzero signatures ($\lambda \neq 0$) in $\mathcal{F}_1 \cup \mathcal{F}_2 \cup \mathcal{F}_3$ with arity at least 2, where $\mathcal{F}_1$, $\mathcal{F}_2$, and $\mathcal{F}_3$ are three families of signatures defined as

$$\begin{aligned}
\mathcal{F}_1 &= \left\{\lambda\left([1,0]^{\otimes k} + i^r[0,1]^{\otimes k}\right) \mid \lambda \in \mathbb{C}, k = 1,2,\ldots, r = 0,1,2,3\right\}, \\
\mathcal{F}_2 &= \left\{\lambda\left([1,1]^{\otimes k} + i^r[1,-1]^{\otimes k}\right) \mid \lambda \in \mathbb{C}, k = 1,2,\ldots, r = 0,1,2,3\right\}, \qquad (1.4) \\
\mathcal{F}_3 &= \left\{\lambda\left([1,i]^{\otimes k} + i^r[1,-i]^{\otimes k}\right) \mid \lambda \in \mathbb{C}, k = 1,2,\ldots, r = 0,1,2,3\right\}.
\end{aligned}$$

We explicitly list these signatures up to an arbitrary constant multiple from $\mathbb{C}$:

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

Table 1.1: List of all non-degenerate affine signatures.

The tractability of $\mathcal{A}$ is first shown in [CLX14]. In fact, #CSP($\mathcal{A}$) is tractable and hence so is Holant($\mathcal{A}$) by (1.1). It was later generalized to arbitrary domain size using Gauss sums [CCLL10]. Together with Lemma 1.4, we have the following.

Lemma 1.7. *Let $\mathcal{F}$ be any set of symmetric, complex-valued signatures in Boolean variables. If $\mathcal{F}$ is $\mathcal{A}$-transformable, then* Holant($\mathcal{F}$) *is computable in polynomial time.*

Product-Type Signatures

Definition 1.8 (Definition 3.3 in [CLX14])**.** *A function is of* product type *if it can be expressed as a product of unary functions, binary equality functions* $([1,0,1])$*, and binary disequality functions* $([0,1,0])$*. We use $\mathcal{P}$ to denote the set of product-type functions.*

An alternate definition for $\mathcal{P}$, implicit in [CLX11b], is the tensor closure of signatures with support on two complementary bit vectors. It is easily shown (cf. Lemma A.1 in the full version of [HL12]) that if f is a symmetric signature in $\mathcal{P}$, then f is degenerate, binary DISEQUALITY $\neq_2$, or $[a, 0, \ldots, 0, b]$ for some $a, b \in \mathbb{C}$.

The tractability of $\mathcal{P}$ is due to a straightforward propagation algorithm (see, for example [CLX14]). In fact, #CSP$(\mathcal{P})$ is tractable and hence so is Holant$(\mathcal{P})$ by (1.1). Together with Lemma 1.4, we have the following.

Lemma 1.9. *Let $\mathcal{F}$ be any set of symmetric, complex-valued signatures in Boolean variables. If $\mathcal{F}$ is $\mathcal{P}$-transformable, then* Holant$(\mathcal{F})$ *is computable in polynomial time.*

Matchgate Signatures

Matchgates were introduced by Valiant [Val02b, Val02a] to give polynomial-time algorithms for a collection of counting problems over planar graphs. As the name suggests, problems expressible by matchgates can be reduced to computing a weighted sum of perfect matchings. The latter problem is tractable over planar graphs by Kasteleyn's algorithm [Kas67], a.k.a. the FKT algorithm [TF61, Kas61]. These counting problems are naturally expressed in the Holant framework using *matchgate signatures*, denoted by $\mathcal{M}$. Thus Pl-Holant$(\mathcal{M})$ is tractable.

Formally, recall that $\mathcal{EO}$ is the set of EXACTONE$_k$ functions for all integers k. Let $\mathcal{WEO}$ be the set of weighted EXACTONE$_k$ functions for all k. Then $\mathcal{M}$ contains signatures that can be realized as an $\mathcal{WEO}$-gate. Holographic transformations extend the reach of the FKT algorithm even further by Lemma 1.4, as stated below.

Lemma 1.10. *Let $\mathcal{F}$ be any set of symmetric, complex-valued signatures in Boolean variables. If $\mathcal{F}$ is $\mathcal{M}$-transformable, then* Pl-Holant$(\mathcal{F})$ *is computable in polynomial time.*

Matchgate signatures are characterized by the matchgate identities (for an up-to-date treatment, see [CG14] for the identities and a self-contained proof). Any matchgate signature f must satisfy the *parity condition*, which asserts that the support of f has to contain entries of only even or odd Hamming weights, but not both. For symmetric matchgates, they have 0 for every other entry and form a geometric progression with the remaining entries. We explicitly list all the symmetric signatures in $\mathcal{M}$ (see [CG14]).

Proposition 1.11. *Let* f *be a symmetric signature in* $\mathcal{M}$. *Then there exists* $a, b \in \mathbb{C}$ *and* $n \in \mathbb{N}$ *such that* f *takes one of the following forms:*

1. $[a^n, 0, a^{n-1}b, 0, \ldots, 0, ab^{n-1}, 0, b^n]$ *(of arity* $2n \geqslant 2$*);*
2. $[a^n, 0, a^{n-1}b, 0, \ldots, 0, ab^{n-1}, 0, b^n, 0]$ *(of arity* $2n+1 \geqslant 1$*);*
3. $[0, a^n, 0, a^{n-1}b, 0, \ldots, 0, ab^{n-1}, 0, b^n]$ *(of arity* $2n+1 \geqslant 1$*);*
4. $[0, a^n, 0, a^{n-1}b, 0, \ldots, 0, ab^{n-1}, 0, b^n, 0]$ *(of arity* $2n+2 \geqslant 2$*).*

In the last three cases with $n = 0$, *the signatures are* $[1, 0]$, $[0, 1]$, *and* $[0, 1, 0]$. *Any multiple of these is also a matchgate signature.*

Note that perfect matching signatures, $[0, 1, 0, \cdots, 0]$, and their reversal are special cases when $b = 0$ or $a = 0$ in the last two cases.

Another useful way to view the symmetric signature in $\mathcal{M}$ is via a low tensor rank decomposition. To state these low rank decompositions, we use the following definition.

Definition 1.12. *Let* S_n *be the symmetric group of degree* n. *Then for positive integers* t *and* n *with* $t \leqslant n$ *and unary signatures* $v, v_1, \ldots, v_{n-t}$, *we define*

$$\operatorname{Sym}_n^t(v; v_1, \ldots, v_{n-t}) = \sum_{\pi \in S_n} \bigotimes_{k=1}^{n} u_{\pi(k)},$$

where the ordered sequence $(u_1, u_2, \ldots, u_n) = (\underbrace{v, \ldots, v}_{t \text{ copies}}, v_1, \ldots, v_{n-t})$.

Proposition 1.13. *Let* f *be a symmetric signature in* $\mathcal{M}$ *of arity* n. *Then there exist* $a, b, \lambda \in \mathbb{C}$ *such that* f *takes one of the following forms:*

1. $[a, b]^{\otimes n} + [a, -b]^{\otimes n} = \begin{cases} 2[a^n, 0, a^{n-2}b^2, 0, \ldots, 0, b^n] & n \textit{ is even,} \\ 2[a^n, 0, a^{n-2}b^2, 0, \ldots, 0, ab^{n-1}, 0] & n \textit{ is odd;} \end{cases}$
2. $[a, b]^{\otimes n} - [a, -b]^{\otimes n} = \begin{cases} 2[0, a^{n-1}b, 0, a^{n-3}b^3, 0, \ldots, 0, ab^{n-1}, 0] & n \textit{ is even,} \\ 2[0, a^{n-1}b, 0, a^{n-3}b^3, 0, \ldots, 0, b^n] & n \textit{ is odd;} \end{cases}$
3. $\lambda \operatorname{Sym}_n^{n-1}([1, 0]; [0, 1]) = [0, \lambda, 0, \ldots, 0]$;
4. $\lambda \operatorname{Sym}_n^{n-1}([0, 1]; [1, 0]) = [0, \ldots, 0, \lambda, 0]$.

The understanding of matchgates was further developed in [CL11a], which characterized, for every symmetric signature, the set of holographic transformations under which the transformed signature becomes a matchgate signature.

1.6 Some Known Dichotomies

We first state a dichotomy theorem for a single signature of arity 3. It will be the induction base case in various settings. It is a combination of [CHL12, Theorem 3] (standard setting) and [CLX10, Theorem V.1] (planar setting).

Theorem 1.14. *If* $f = [f_0, f_1, f_2, f_3]$ *is a non-degenerate, complex-valued signature, then* Pl-Holant (f) *is* #**P**-*hard unless* f *satisfies one of the following conditions, in which case the problem is computable in polynomial time:*

1. f *is* $\mathcal{A}$- *or* $\mathcal{P}$-*transformable;*

2. *For* $\alpha \in \{2i, -2i\}$, $f_2 = \alpha f_1 + f_0$ *and* $f_3 = \alpha f_2 + f_1$*;*

3. f *is* $\mathcal{M}$-*transformable.*

If f *satisfies condition 1 or 2, then* Holant(f) *is computable in polynomial time without planarity; otherwise* Holant(f) *is* #**P**-*hard.*

Next is a dichotomy theorem about counting complex weighted graph homomorphisms over degree prescribed graphs.

Theorem 1.15 (Theorem 3 in [CK12])**.** *Let* $S \subseteq \mathbb{Z}^+$ *containing some* $r \geqslant 3$*, let* $\mathcal{G} = \{=_k \mid k \in S\}$*, and let* $d = \gcd(S)$*. Further suppose that* $f_0, f_1, f_2 \in \mathbb{C}$*. Then* Pl-Holant $([f_0, f_1, f_2] \mid \mathcal{G})$ *is* #**P**-*hard unless one of the following conditions holds:*

1. $f_0 f_2 = f_1^2$*;*

2. $f_0 = f_2 = 0$*;*

3. $f_1 = 0$*;*

4. $f_0 f_2 = -f_1^2$ *and* $f_0^d = -f_2^d \neq 0$*;*

5. $f_0^d = f_2^d \neq 0$.

In all exceptional cases, the problem is computable in polynomial time.

In particular, in Case 5, $\operatorname{Holant}([f_0, f_1, f_2] \mid \mathcal{G})$ *is #***P***-hard unless another condition is satisfied.*

Theorem 1.15 is the original statement as in [CK12]. It is explicit and easy to apply. Conceptually, it can be restated as Theorem 1.15′.

Theorem 1.15′ (Theorem 3 in [CK12])**.** *Let* $S \subseteq \mathbb{Z}^+$ *contain* $k \geqslant 3$*, let* $\mathcal{G} = \{=_k \mid k \in S\}$*, and let* $d = \gcd(S)$*. Further suppose that* f *is a non-degenerate, symmetric, complex-valued binary signature in Boolean variables. Then* $\operatorname{Pl-Holant}(f \mid \mathcal{G})$ *is #***P***-hard unless* f *satisfies one of the following conditions, in which case the problem is computable in polynomial time:*

1. *there exists* $T \in \mathcal{T}_{4d}$ *such that* $T^{\otimes 2} f \in \mathcal{A}$*;*

2. $f \in \mathcal{P}$*;*

3. *there exists* $T \in \mathcal{T}_{2d}$ *such that* $T^{\otimes 2} f \in \widehat{\mathcal{M}}$*.*

In particular, in Case 3, $\operatorname{Holant}([f_0, f_1, f_2] \mid \mathcal{G})$ *is #***P***-hard unless another condition is satisfied.*

The following dichotomy theorems are not directly used. We list them for comparison. First is the Boolean #CSP dichotomy. Note that it is not restricted to symmetric functions.

Theorem 1.16 (Theorem 3.1 in [CLX14])**.** *Let* $\mathcal{F}$ *be a set of complex functions in Boolean variables. Then* $\#\mathrm{CSP}(\mathcal{F})$ *is #***P***-hard unless* $\mathcal{F} \subseteq \mathcal{P}$ *or* $\mathcal{F} \subseteq \mathcal{A}$*, in which case the problem is computable in polynomial time.*

The other three are about Holant problems, and we paraphrase them in the more up-to-date language of $\mathcal{C}$-transformable signatures. We have the real-valued Holant dichotomy.

Theorem 1.17 (Theorem III.2 in [HL12])**.** *Let* $\mathcal{F}$ *be any set of symmetric, real-valued signatures in Boolean variables. Then* $\operatorname{Holant}(\mathcal{F})$ *is #***P***-hard unless* $\mathcal{F}$ *satisfies one of the following conditions, in which case the problem is computable in polynomial time:*

1. *Any non-degenerate signature in* $\mathcal{F}$ *is of arity at most* 2*;*

2. $\mathcal{F}$ *is* $\mathcal{A}$*- or* $\mathcal{P}$*-transformable.*

The other two dichotomy theorems concern about Holant problems with some auxiliary functions available. Let $\mathcal{U}$ denotes the set of all unary functions, and in particular, let Δ_0 and Δ_1 denote constant unary functions $[1,0]$ and $[0,1]$. Denote by $\text{Holant}^*(\mathcal{F})$ the problem $\text{Holant}(\mathcal{F} \cup \mathcal{U})$, that is, unary functions are freely available. Similarly denote by $\text{Holant}^c(\mathcal{F})$ the problem $\text{Holant}(\mathcal{F} \cup \{\Delta_0, \Delta_1\})$, that is, constant unary functions are freely available. Complex-valued Holant^* and Holant^c are classified in [CLX11a] and [CHL12], respectively.

Theorem 1.18 (Theorem 4.1 in [CLX11a])**.** *Let $\mathcal{F}$ be any set of symmetric, complex-valued signatures in Boolean variables. Then* $\text{Holant}^*(\mathcal{F})$ *is #**P**-hard unless $\mathcal{F}$ satisfies one of the following conditions, in which case the problem is computable in polynomial time:*

1. *Any non-degenerate signature in $\mathcal{F}$ is of arity at most 2;*
2. *$\mathcal{F}$ is $\mathcal{P}$-transformable;*
3. *There exists $\alpha \in \{2i, -2i\}$, such that for any signature $f \in \mathcal{F}$ of arity n, for $0 \leqslant k \leqslant n-2$, we have $f_{k+2} = \alpha f_{k+1} + f_k$.*

Theorem 1.19 (Theorem 6 in [CHL12])**.** *Let $\mathcal{F}$ be any set of symmetric, complex-valued signatures in Boolean variables. Then* $\text{Holant}^c(\mathcal{F})$ *is #**P**-hard unless $\mathcal{F}$ satisfies one of the following conditions, in which case the problem is computable in polynomial time:*

1. *Any non-degenerate signature in $\mathcal{F}$ is of arity at most 2;*
2. *$\mathcal{F}$ is $\mathcal{P}$-transformable;*
3. *$\mathcal{F} \cup \{[1,0],[0,1]\}$ is $\mathcal{A}$-transformable;*
4. *There exists $\alpha \in \{2i, -2i\}$, such that for any non-degenerate signature $f \in \mathcal{F}$ of arity n, for $0 \leqslant k \leqslant n-2$, we have $f_{k+2} = \alpha f_{k+1} + f_k$.*

Chapter 2

Parity Holant Problems

In this chapter, we show a dichotomy theorem for parity Holant problems. The parity Holant problem concerns about Boolean functions taking values in $\mathbb{Z}_2$. To avoid confusion, we call them $\mathbb{Z}_2$-functions or $\mathbb{Z}_2$-signatures. In other words, our goal is to compute the parity of integer valued Holant problems. We write $\oplus\text{Holant}(\Omega;\mathcal{F})$ to denote the Holant value in this particular case, where $\mathcal{F}$ is a set of $\mathbb{Z}_2$-signatures.

Name $\oplus\text{Holant}(\mathcal{F})$

Instance A *signature grid* $\Omega = (G, \pi)$.

Output $\oplus\text{Holant}(\Omega;\mathcal{F})$.

Similar to Holant^c problems, $\oplus\text{Holant}^c$ denote problems with $\Delta_0 = [1, 0]$ and $\Delta_1 = [0, 1]$ available. Our proof strategy is to first prove a dichotomy for $\oplus\text{Holant}^c$ and then extend it to the general case. We call a signature f a sub-signature of g if $f_0 f_1 \cdots f_n$ is a substring of $g_0 g_1 \cdots g_m$. In the Holant^c setting any sub-signature is realizable.

Note that the only non-trivial transformation in $\mathbb{Z}_2$ is $\left[\begin{smallmatrix}1 & 1\\ 0 & 1\end{smallmatrix}\right]$. However, we shall not restrict transformations in $\mathbb{Z}_2$. As we will see in a moment, transformations in $\mathbb{R}$ are useful to derive tractability results.

2.1 Tractable $\mathbb{Z}_2$-Functions

We start with three tractable families for $\oplus$Holant^c problems, related to the three families described in Section 1.5. We need to consider each class in the specialized parity setting. The first family, *affine signatures*, is adopted directly. The second family is derived from *Fibonacci signatures* [CLX13]. As we now understand, it is actually a subclass of $\mathcal{P}$-transformable signatures. For general counting problems, Fibonacci signatures are tractable, and so are they for parity counting. This family remains tractable even with the addition of the inversion signature $[0, 1, 0]$ in the parity setting. This addition for general counting problems would entail #**P**-hardness. The third family is *matchgate signatures*. As noted in Section 1.5, matchgate signatures are tractable in planar graphs but not in general graphs. However, if we are restricted to the parity problem, then matchgate $\mathbb{Z}_2$-signatures are tractable even in general graphs since we can decide the parity of the number of perfect matchings efficiently.

Clearly for any set $\mathcal{F}$, if $\oplus$Holant$^c(\mathcal{F})$ is tractable then so is $\oplus$Holant$(\mathcal{F})$.

Affine Signatures

Unlike in Section 1.5, we only need to consider unweighted affine signatures. In this special case, affine signatures correspond to simultaneous linear equations over $\mathbb{Z}_2$ and are defined as follows.

Definition 2.1. *A $\mathbb{Z}_2$-signature is affine if its support is an affine space. Denote by $\mathcal{A}_\oplus$ the set of all affine $\mathbb{Z}_2$-signatures.*

By definition, an affine $\mathbb{Z}_2$-signature can be viewed as a constraint defined by a set of linear equations. Viewing edges as variables in $\mathbb{Z}_2$, every assignment which contributes 1 in $\oplus$Holant is a solution which satisfies all linear equations, and vice versa. Hence $\oplus$Holant is exactly the number of solutions of the linear system, which can be computed in polynomial time. Note that by definition all degenerate $\mathbb{Z}_2$-signatures are in $\mathcal{A}_\oplus$.

Lemma 2.2. *The problem $\oplus$Holant$^c(\mathcal{A}_\oplus)$ is polynomial time computable.*

We explicitly list all non-degenerate symmetric affine $\mathbb{Z}_2$-signatures, derived from Table 1.1. Notice that in the parity setting 1, -1, and i all behave the same. Every non-degenerate symmetric affine $\mathbb{Z}_2$-signature has one of the following forms:

- Equality Signatures: $[1, 0, 0, \ldots, 0, 1]$;
- Parity Signatures: $[1, 0, 1, 0, \ldots, 0/1]$ or $[0, 1, 0, 1, \ldots, 0/1]$,

where the last entry depends on the arity.

Fibonacci Signatures and $[0, 1, 0]$

The family of Fibonacci signatures was introduced in [CLX13] to characterize a new family of holographic algorithms. It is now understood as a subclass of $\mathcal{P}$-transformable signatures [CGW14]. In $\mathbb{Z}_2$, we have the following definition.

Definition 2.3. *A symmetric $\mathbb{Z}_2$-signature $[f_0, f_1, \ldots, f_n]$ is called Fibonacci if for $0 \leqslant k \leqslant n-2$, $f_k + f_{k+1} = f_{k+2}$.*

The Holant of a grid composed of Fibonacci signatures can be computed in polynomial time [CLX13]. Its parity version is therefore also tractable. Here we will show that the tractability still holds even if we extend the set with the binary DisEquality signature $[0, 1, 0]$, which is not Fibonacci. The tractability is based on the properties of Fibonacci $\mathbb{Z}_2$-signatures and an observation on $[0, 1, 0]$ as a $\mathbb{Z}_2$-signature. Denote by $\mathcal{P}_{\oplus}$ the set of all Fibonacci $\mathbb{Z}_2$-signatures plus $[0, 1, 0]$ and all degenerate $\mathbb{Z}_2$-signatures. Note that as $[0, 1]$, $[1, 0]$, and $[1, 1]$ are all Fibonacci, adding degenerate signatures in $\mathcal{P}_{\oplus}$ does not affect its tractability.

Lemma 2.4. *The problem $\oplus\text{Holant}^c(\mathcal{P}_{\oplus})$ is polynomial time computable.*

In the proof, we will go beyond $\mathbb{Z}_2$ and use transformations in $\mathbb{R}$. This is the only place in this chapter where signature entries are viewed as real numbers rather than in $\mathbb{Z}_2$.

Since our goal is the parity, $[0, 1, 0]$ can be replaced by the asymmetric signature $(0, 1, -1, 0)$ in $\mathbb{R}$. (Note that here the expression $(0, 1, -1, 0)$ is a vector listing function values, rather than the abbreviated form of symmetric signatures.) This $(0, 1, -1, 0)$ is a so-called *2-realizable* *signature*, which has a special invariant property under holographic transformations [Val06, CL11a, CL11b]. This property plays an important role in the proof.

Proof. As stated above, we replace $[0,1,0]$ by the asymmetric signature $(0,1,-1,0)$. We also replace Fibonacci $\mathbb{Z}_2$-signatures by real-valued Fibonacci signatures. For example, $[1,1,0,1]$ is replaced by $[1,1,2,3]$. After the replacement, the parity of the Holant value does not change. Denote the set of real-valued Fibonacci signatures by F. Since $\Delta_0, \Delta_1 \in F$, $\text{Holant}^c(F \cup (0,1,-1,0))$ is equivalent to $\text{Holant}(F \cup (0,1,-1,0))$. Next we show that $\text{Holant}(F \cup (0,1,-1,0))$ is computable in polynomial time.

For a Fibonacci signature $f = [f_0, f_1, \ldots, f_n]$ over $\mathbb{R}$, we have that $f_{k+2} = f_{k+1} + f_k$ for all $k = 0, 1, \ldots, n-2$. This is a second-order homogeneous linear recurrence relation. Thus we have that $f_i = A\lambda_1^i + B\lambda_2^i$ for $1 \leqslant i \leqslant n$, where $\lambda_1 = (1-\sqrt{5})/2$ and $\lambda_2 = (1+\sqrt{5})/2$ are the two roots of its characteristic polynomial $x^2 = x+1$, and A, B are two real numbers depending on f_0 and f_1. In the tensor notation, we have that $f = A\left[\begin{smallmatrix}1\\ \lambda_1\end{smallmatrix}\right]^{\otimes n} + B\left[\begin{smallmatrix}1\\ \lambda_2\end{smallmatrix}\right]^{\otimes n}$. One crucial observation is that λ_1 and λ_2 are universal for all Fibonacci signatures (while A and B can vary depending on the initial values). Therefore we can do a holographic reduction as in Theorem 1.2 under the following orthogonal matrix $T = \begin{bmatrix} \frac{1}{\sqrt{\lambda_1^2+1}} & \frac{\lambda_1}{\sqrt{\lambda_1^2+1}} \\ \frac{1}{\sqrt{\lambda_2^2+1}} & \frac{\lambda_2}{\sqrt{\lambda_2^2+1}} \end{bmatrix}$. We note that T is orthogonal because $\lambda_1\lambda_2 = -1$. This does not change the Holant value by Theorem 1.2. But all Fibonacci signatures have nicer forms since

$$\begin{aligned}
T^{\otimes n} f &= T^{\otimes n}\left(A\left[\begin{smallmatrix}1\\ \lambda_1\end{smallmatrix}\right]^{\otimes n} + B\left[\begin{smallmatrix}1\\ \lambda_2\end{smallmatrix}\right]^{\otimes n}\right) \\
&= AT^{\otimes n}\left[\begin{smallmatrix}1\\ \lambda_1\end{smallmatrix}\right]^{\otimes n} + BT^{\otimes n}\left[\begin{smallmatrix}1\\ \lambda_2\end{smallmatrix}\right]^{\otimes n} \\
&= A\left(T\left[\begin{smallmatrix}1\\ \lambda_1\end{smallmatrix}\right]\right)^{\otimes n} + B\left(T\left[\begin{smallmatrix}1\\ \lambda_2\end{smallmatrix}\right]\right)^{\otimes n} \\
&= \sqrt{\lambda_1^2+1}A\left[\begin{smallmatrix}1\\ 0\end{smallmatrix}\right]^{\otimes n} + \sqrt{\lambda_2^2+1}B\left[\begin{smallmatrix}0\\ 1\end{smallmatrix}\right]^{\otimes n} \\
&= \left[\sqrt{\lambda_1^2+1}A, 0, \ldots, 0, \sqrt{\lambda_2^2+1}B\right].
\end{aligned}$$

Note that $\left[\sqrt{\lambda_1^2+1}A, 0, \ldots, 0, \sqrt{\lambda_2^2+1}B\right] \in \mathcal{P}$. This in fact shows that Fibonacci signatures are $\mathcal{P}$-transformable.

For the signature $(0,1,-1,0)$, it is easy to verify that $T^{\otimes 2}(0,1,-1,0) = (0,-1,1,0)$. Moreover, as an asymmetric signature, $(0, -1, 1, 0) \in \mathcal{P}$ as well. Hence we have shown that

$$\operatorname{Holant}(F \cup (0, 1, -1, 0)) \leqslant_T \operatorname{Holant}(\mathcal{P}).$$

Recall that $\operatorname{Holant}(\mathcal{P})$ is tractable by Lemma 1.9. This finishes the proof. □

Next we list explicitly all non-degenerate Fibonacci $\mathbb{Z}_2$-signatures. By Definition 2.3, a Fibonacci $\mathbb{Z}_2$-signature is completely determined by its first two bits and arity. The initial bits can be $00, 01, 10$, or 11. However, the 00 case leads to the degenerate all-0 signature. Every non-degenerate Fibonacci $\mathbb{Z}_2$-signature has one of the following forms:

- $[0, 1, 1, 0, 1, 1, \ldots, 0/1]$,
- $[1, 0, 1, 1, 0, 1, \ldots, 0/1]$,
- $[1, 1, 0, 1, 1, 0, \ldots, 0/1]$.

The following lemma regarding the realizability is useful in the hardness proof later.

Lemma 2.5. *Any non-degenerate Fibonacci $\mathbb{Z}_2$-signature can be realized by any $\mathbb{Z}_2$-signature in the set* $\{[0, 1, 1, 0], [1, 0, 1, 1], [1, 1, 0, 1]\}$ *with the two unary constants* $[1, 0]$ *and* $[0, 1]$.

Before the proof, we want to point out that $\{[0, 1, 1, 0], [1, 0, 1, 1], [1, 1, 0, 1]\}$ is the set of non-degenerate ternary Fibonacci $\mathbb{Z}_2$-signature. In other words, they are the simplest non-trivial ones.

Proof. Suppose we have $f \in \{[0, 1, 1, 0], [1, 0, 1, 1], [1, 1, 0, 1]\}$ and we want to realize some non-degenerate Fibonacci $\mathbb{Z}_2$-signature g of arity n. Given another non-degenerate Fibonacci $\mathbb{Z}_2$-signature f', it is easy to verify that by connecting f' to f via one edge is still symmetric, non-degenerate, and Fibonacci. Hence we may construct non-degenerate Fibonacci $\mathbb{Z}_2$-signature of arbitrarily long arity. Moreover, observe that g is a subsignature of any non-degenerate Fibonacci $\mathbb{Z}_2$-signature of arity $n + 2$, despite its initial values. Then the construction is to get such a signature of arity $n + 2$ using f first, and get the subsignature g using $[1, 0]$ and $[0, 1]$ next. □

Matchgate Signatures

In this chapter our notion of matchgates is different from that in Section 1.5. The main difference is that we allow not necessarily perfect matchings. As we shall see later, this relaxation maintains tractability even in not necessarily planar graphs. Denote by $EO_n = [0, 1, 0, \cdots, 0]$ the ExactOne function of arity n, and $AO_n = [1, 1, 0, \cdots, 0]$ the AtMostOne function of arity n. Let $\mathcal{EO}$ and $\mathcal{AO}$ the set of all these functions, respectively.

Definition 2.6. *A $\mathbb{Z}_2$-signature f is called matchgate if f can be realized by functions from $\mathcal{EO}$ and $\mathcal{AO}$. Denote by $\mathcal{M}_\oplus$ the set of all matchgate $\mathbb{Z}_2$-signatures.*

Note that $[0, 1]$, $[1, 0]$ and $[1, 1]$ are all matchgates. Hence all degenerate $\mathbb{Z}_2$ signatures are in $\mathcal{M}_\oplus$. Also note that this notion of matchgates is in its most general sense: the gadget can be either planar or non-planar and for each node we can insist or not on whether it has to be saturated by a matching edge. We will prove the following.

Lemma 2.7. *The problem $\oplus\,\mathrm{Holant}^c(\mathcal{M}_\oplus)$ is polynomial time computable.*

As $[1, 0], [0, 1] \in \mathcal{M}_\oplus$, we define *parity matching problem* which is equivalent to $\oplus\,\mathrm{Holant}^c$ $(\mathcal{M}_\oplus)$. Reductions between these two problems are straightforward. The graph is the same. All the vertices in V_0 have signatures from $\mathcal{EO}$ and all the other vertices have signatures from $\mathcal{AO}$.

Name Parity Matching Problem

Instance A graph $G = (V, E)$ and $V_0 \subseteq V$.

Output Parity of the number of (partial) matchings that saturate all the vertexes in V_0.

The total number of such matchings can be rewritten as a summation of perfect matchings

$$\mathrm{MatchingS}(G) = \sum_{U \supseteq V_0} \mathrm{PM}(G(U)),$$

where $\mathrm{MatchingS}(G)$ is the value we what to compute, $\mathrm{PM}(G)$ is the number of perfect matchings in G, and $G(U)$ is the induced subgraph of G on vertex set U.

Before proceeding to the algorithm for the Parity Matching Problem, we need an important notion called the Pfaffian. The Pfaffian of an $n \times n$ skew-symmetric matrix A, denoted

$\mathrm{Pf}(A)$, is defined to be zero if n is odd and one if $n = 0$. If $n = 2k$ is even with $k > 0$, then:

$$\mathrm{Pf}(A) = \frac{1}{2^k k!} \sum_{\sigma \in S_{2k}} \varepsilon_\sigma \prod_{i=1}^{k} A(\sigma(2i-1), \sigma(2i)),$$

where σ is a permutation on $\{1, 2, \cdots, 2k\}$, S_{2k} is the symmetric group on $2k$ elements, and $\varepsilon_\sigma \in \{1, -1\}$ is the sign of σ.

The following fact, due to Cayley [Cay49], (see also [BR91] Theorem 9.5.2) relates the Pfaffian to the determinant.

Theorem 2.8. *For any* $2k \times 2k$ *skew-symmetric matrix* A

$$\det(A) = (\mathrm{Pf}(A))^2.$$

In the field $\mathbb{Z}_2$, we have $x \equiv -x$ and hence a skew-symmetric matrix is indeed symmetric. Moreover the sign ε_π in $\mathbb{Z}_2$ can be ignored. Let A be the adjacency matrix of a graph $G = (V, E)$, i.e. nonzero elements of A are $A_{i,j} = A_{j,i} = 1$ if $\{i, j\} \in E$. Then each monomial in the Pfaffian corresponds to a distinct perfect matching in G. Therefore, in $\mathbb{Z}_2$, $\mathrm{Pf}(A)$ is exactly the parity of the number of perfect matchings in G. We have that

$$\mathrm{PM}(G) \equiv \mathrm{Pf}(A) \equiv (\mathrm{Pf}(A))^2 \equiv \det(A) \pmod 2 \tag{2.1}$$

if G has an even number of vertices. Since $\det(A)$ is polynomial time computable, so is the parity of the number of perfect matchings.

Next we show that this tractability can be extended to partial matchings as well in $\mathbb{Z}_2$. We do this through the *Pfaffian Sum Theorem* [Val02b]. For any $n \times n$ matrix A we call a set $I = \{i_1, i_2, \ldots, i_r\} \subseteq [n]$ an index set. Further we denote by $A(I)$ the $r \times r$ sub matrix of A on rows and columns in I.

Definition 2.9. *The Pfaffian Sum of an* $n \times n$ *matrix* A *is a polynomial over indeterminates*

$\lambda_1, \lambda_2, \ldots, \lambda_n$ *such that*

$$\mathrm{PfS}(A) = \sum_{I \subseteq [n]} \left(\prod_{i \notin I} \lambda_i \right) \mathrm{Pf}(A(I)).$$

We only need instances in which each λ_i is either 0 or 1. For a given unomittable vertex set V_0, we define the characteristic vector $\vec{\lambda} = (\lambda_1, \ldots \lambda_n)$ as follows: for each i, $\lambda_i = 0$ if $i \in V_0$ and $\lambda_i = 1$ otherwise. Thus the Pfaffian Sum over $\vec{\lambda}$ is the sum of $\mathrm{Pf}(A(I))$ over index sets containing V_0.

We define the $n \times n$ matrix $\Lambda^{(n)}$ as follows:

$$\Lambda^{(n)}(i,j) = \begin{cases} (-1)^{j-i+1}\lambda_i\lambda_j, & \text{if } i < j, \\ (-1)^{i-j}\lambda_i\lambda_j, & \text{if } i > j, \\ 0, & \text{if } i = j. \end{cases}$$

Also for an $n \times n$ matrix A we define A^+ to be the $(n+1) \times (n+1)$ matrix of which the first n rows and columns equal A itself, and the $(n+1)$-st row and column entries are all zero.

The following theorem, which relates the Pfaffian Sum to a single Pfaffian, was proved in [Val02b].

Theorem 2.10. *For an $n \times n$ skew-symmetric matrix A, and indeterminates $\lambda_1, \cdots, \lambda_{n+1}$*

$$\mathrm{PfS}(A) = \begin{cases} \mathrm{Pf}(A + \Lambda^{(n)}) & \textit{if } n \textit{ is even} \\ \mathrm{Pf}(A^+ + \Lambda^{(n+1)}) & \textit{with } \lambda_{n+1} = 1, \textit{ if } n \textit{ is odd.} \end{cases}$$

Thus, a Pfaffian Sum can be computed in polynomial time. The relation (2.1) between perfect matchings and Pfaffians can be therefore generalized to one between matchings and Pfaffian Sums:

$$\mathrm{MatchingS}(G) = \sum_{U \supseteq V_0} \mathrm{PM}(G(U)) = \sum_{U \supseteq V_0} \mathrm{Pf}(A(U)) = \mathrm{PfS}(A)(\vec{\lambda}) \pmod 2. \tag{2.2}$$

This relation gives a polynomial time algorithm for the Parity Matching Problem and completes the proof of Lemma 2.7.

Now we go on to list explicitly all the non-degenerate symmetric matchgate signatures. *Useful matchgate identities* in [CCL09] is an essential tool to characterize the realizability of matchgates. For completeness we quote the identities as follows.

A pattern α is an m-bit string, i.e., $\alpha \in \{0,1\}^m$. A position vector $P = \{p_i\}, i \in [l]$, is a subsequence of $\{1,2,\ldots,m\}$, i.e., $p_i \in [m]$ and $p_1 < p_2 < \cdots < p_l$. It can also be viewed as a m-bit string, whose $(p_1,p_2,\ldots,p_l)$-th bits are 1 and the others are 0. Let $e_i \in \{0,1\}^m$ be the pattern with 1 in the i-th bit and 0 elsewhere. Let $\alpha+\beta$ denote the bitwise XOR of the patterns α and β.

Proposition 2.11 (Matchgate Identities for Signatures)**.** *For a signature f realizable by signatures from $\mathcal{EO}$, for any pattern $\alpha \in \{0,1\}^m$, any l $(0 < l \leqslant m)$, and any position vector $P = \{p_i\}, i \in [l]$, the following identity holds:*

$$\sum_{i=1}^{l} (-1)^i f(\alpha + e_{p_i}) f(\alpha + p + e_{p_i}) = 0. \tag{2.3}$$

Lemma 2.12. *Every non-degenerate symmetric $\mathbb{Z}_2$-signature in $\mathcal{M}_\oplus$ has one of the following forms:*

- ExactOne $[0,1,0,0,\ldots,0]$ *or its reversal* $[0,0,\ldots,0,1,0]$*,*
- AtMostOne $[1,1,0,0,\ldots,0]$ *or its reversal* $[0,0,\ldots,0,1,1]$*,*
- Parity*:* $[1,0,1,0,\ldots,0/1]$ *or* $[0,1,0,1,\ldots,0/1]$*.*

Proof. We first prove that every non-degenerate symmetric signature f of arity n realizable from $\mathcal{EO}$ and $\mathcal{AO}$ has one of the forms as claimed in the lemma. Suppose in general f is realized by some gadget $G = (V, E, D)$ where D is the set of dangling edges, and $V_0 \subseteq V$ is the set of omittable vertices. We construct G' where every vertex in G' has $\mathcal{EO}$ signature using Theorem 2.10 as follows. We add all edges (i,j) where $i,j \in V_0$. Moreover, we add an additional vertex u with a dangling edge, and we connect u to all vertices in V_0. Then G' defines a signature g of arity $n+1$ since we add one more dangling edge.

Suppose the input is α on D. Then it is equivalent to remove all vertices that are adjacent to a dangling edge chosen by α. If there are odd many of remaining omittable vertices, then we keep u as remained, and otherwise remove u. That is to say, the input on u is $\oplus\alpha = \oplus_{i=1}^{n}\alpha_i$, the parity of α. By Theorem 2.10, $g(\alpha, \oplus\alpha) = f(\alpha)$. Since every signature in G' is from $\mathcal{EO}$, g satisfies (2.3) by Proposition 2.11.

In the following, we assume that G has an odd number of vertices. The case of even vertices is similar.

If $n = 2$, all non-degenerate symmetric signatures are of the forms claimed in the lemma.

We now consider $n \geqslant 3$ and apply the matchgate identities (2.3) to g. Consider the pattern $100\alpha0$ where α has Hamming weight $2i$, and $0 \leqslant 2i \leqslant n-3$. Let the position vector be $111000\cdots01$. Then (2.3) gives us that

$$0 = g(000\alpha0)g(111\alpha1) - g(110\alpha0)g(001\alpha1) + g(101\alpha0)g(010\alpha1) - g(100\alpha1)g(011\alpha0).$$

Translating back to f, we get that

$$f_{2i}f_{2i+3} - f_{2i+2}f_{2i+1} = 0.$$

If $n = 3$, then this is $f_0f_3 = f_1f_2$ and is the only identity.

If $n \geqslant 4$, we use the matchgate identities (2.3) again. Consider the pattern $1000\alpha \oplus \alpha$ where α has Hamming weight i, and $0 \leqslant i \leqslant n-4$. Let the position vector be $111100\cdots0$. Then (2.3) gives us that

$$\begin{aligned} 0 = {} & g(0000\alpha, \oplus\alpha)g(1111\alpha, \oplus\alpha) - g(1100\alpha, \oplus\alpha)g(0011\alpha, \oplus\alpha) \\ & + g(1010\alpha, \oplus\alpha)g(0101\alpha, \oplus\alpha) - g(1001\alpha, \oplus\alpha)g(0110\alpha, \oplus\alpha). \end{aligned}$$

Translating back to f, we get that

$$f_if_{i+4} - f_{i+2}f_{i+2} = 0.$$

Consider the pattern 10^m and the position vector $1^m \oplus (m)$, where $\oplus(m)$ is the parity of m.

Then we have

$$
\begin{aligned}
0 &= g(0^{m+1})g(1^m \oplus (m)) - g(110^{m-1})g(001^{m-2} \oplus (m)) \\
&\quad + g(1010^{m-2})g(0101^{m-3} \oplus (m)) - g(10010^{m-3})g(01101^{m-4} \oplus (m)) \pm \cdots
\end{aligned}
$$

All terms cancel except the first two. Translating to f, we get that

$$f_0 f_m - f_2 f_{m-2} = 0.$$

Similarly, consider the pattern 10^m and the position vector $1^{m-1}0 \oplus (m-1)$. Then we have

$$
\begin{aligned}
0 &= g(0^{m+1})g(1^{m-1}0 \oplus (m-1)) - g(110^{m-1})g(001^{m-3}0 \oplus (m-1)) \\
&\quad + g(1010^{m-2})g(0101^{m-4}0 \oplus (m-1)) - g(10010^{m-3})g(01101^{m-5}0 \oplus (m-1)) \pm \cdots
\end{aligned}
$$

Again, all terms cancel except the first two. Translating to f, we get that

$$f_0 f_{m-1} - f_2 f_{m-3} = 0.$$

Similarly, we can also get that $f_1 f_m = f_3 f_{m-2}$ and $f_1 f_{m-1} = f_3 f_{m-3}$.

The relation $f_i f_{i+4} = f_{i+2}^2$ implies that the subsequence of entries from even (or odd) indices form a geometric sequence. In $\mathbb{Z}_2$, there are only four types of geometric sequences, which are

1. $[0, 0, \ldots, 0]$,
2. $[1, 0, 0, \ldots, 0]$,
3. $[0, 0, \ldots, 0, 1]$,
4. $[1, 1, \ldots, 1]$.

There are $4 \times 4 = 16$ possible combinations for even subsequence and odd subsequence. We use type (i, j) to denote the sequence whose odd subsequence is of type i and even subsequence is of type j. Types $(1, 1)$ and $(4, 4)$ are degenerate. Types $(1, 2)$, $(1, 3)$, $(1, 4)$, $(2, 1)$, $(2, 2)$, $(3, 1)$,

$(3, 3)$, and $(4, 1)$ are listed in the lemma. We only need to rule out the remaining six types $(2, 3)$, $(2, 4)$, $(3, 2)$, $(3, 4)$, $(4, 2)$, and $(4, 3)$.

- For $(2, 4)$, the first four entries are $[1, 1, 0, 1]$, which does not satisfy $f_0 f_3 = f_1 f_2$. This matchgate identity also rules out $(4, 2)$, whose first four entries are $[1, 1, 1, 0]$.
- For $(3, 4)$ and $(4, 3)$, their last four entries do not satisfy $f_{m-3} f_m = f_{m-1} f_{m-2}$.
- For $(2, 3)$, it has form $[1, 0, \ldots, 0, 1]$ or $[1, 0, \ldots, 0, 1, 0]$. It violates either $f_0 f_m = f_2 f_{m-2}$ or $f_0 f_{m-1} = f_2 f_{m-3}$.
- For $(3, 2)$, we can similarly argue that it violates either $f_1 f_m = f_3 f_{m-2}$ or $f_1 f_{m-1} = f_3 f_{m-3}$.

This completes the first part of the proof, namely that all matchgate $\mathbb{Z}_2$-signatures are of desired forms. The realizability of these signatures follow from Lemma 2.15 below. □

Next we prove some realizability properties regarding symmetric matchgate signatures, which will be useful later in the #**P**-hardness proofs. We prove progressively stronger realizability.

Lemma 2.13. *Every* Parity $\mathbb{Z}_2$*-signature can be realized by the signatures* $[0, 1]$*,* $[1, 0]$*, and* $[0, 1, 0, 1]$ (*or* $[1, 0, 1, 0]$).

Proof. Suppose we have $f = [0, 1, 0, 1]$. The case of $[1, 0, 1, 0]$ is similar. We prove inductively that we can realize $[0, 1, 0, 1, \cdots, 0/1]$ of arity $2k + 1$. The base case is $k = 1$ and is f. Suppose we have $g = [0, 1, 0, 1, \cdots, 0/1]$ of arity $2k + 1 > 3$. We may connect g to f via one single edge twice. The resulting signature has the form $[0, 1, 0, 1, \cdots, 0/1]$ and arity $2k + 3$. Hence we can construct signatures of the form of arbitrary length. Suppose we want to realize some Parity $\mathbb{Z}_2$-signature h of arity n. Then h is a subsignature of the realizable $[0, 1, 0, 1, \cdots, 0/1]$ of arity $n+1$ or $n+2$, depending on the parity of n. Using $[0, 1]$ and $[1, 0]$ to get h finishes the proof. □

Lemma 2.14. *Every* Parity *and* ExactOne (*and its reversal*) $\mathbb{Z}_2$*-signature can be realized by the signatures* $[0, 1]$*,* $[1, 0]$*, and* $[0, 1, 0, 0]$ (*or* $[0, 0, 1, 0]$).

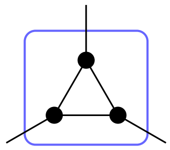


Figure 2.1: The triangular gadget.

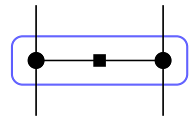


Figure 2.2: The gadget for $[0,1,0,0,0]$ and $[1,1,0,0,0]$.

Proof. By symmetry, we only need to prove the lemma for $[0,1,0,0]$. First observe that if we place $[0,1,0,0]$ at every vertex in the triangular gadget shown in Figure 2.1, the resulting signature is $[0,1,0,1]$. Then by Lemma 2.13, every PARITY signature can be constructed.

Connecting $[1,0]$ to $[0,1,0,0]$, we get $[0,1,0]$. Then we use the gadget in Figure 2.2, where we put $[0,1,0,0]$ on circle vertices and $[0,1,0]$ on the square. It is easy to verify that the resulting $\mathbb{Z}_2$-signature is $[0,1,0,0,0]$, which is symmetric. Note that in general such a construction does not entail a symmetric signature. In the same way we can connect $[0,1,0,\ldots,0]$ of arity k to $[0,1,0,0]$ via $[0,1,0]$, and the resulting $\mathbb{Z}_2$-signature is $[0,1,0,\ldots,0]$ of arity $k+1$. Therefore we can construct all $\mathbb{Z}_2$-signatures in $\mathcal{EO}$. Their reversal can be obtained by putting $[0,1,0]$ on every edge. This completes the proof. □

Lemma 2.15. *Every* PARITY, EXACTONE (*and its reversal*)*, and* ATMOSTONE (*and its reversal*)*, can be realized by the signatures* $[0,1]$*,* $[1,0]$*, and* $[1,1,0,0]$ (*or* $[0,0,1,1]$)*.*

Proof. By symmetry, we only need to prove the lemma for $[1,1,0,0]$. Note that we get $[1,1]$ from $[1,1,0,0]$ by connecting it with $[1,0]$ twice. By connecting $[1,1]$ with $[1,1,0,0]$, we get $[0,1,0]$. Then we place $[1,1,0,0]$ at circles and $[0,1,0]$ at the square on the gadget in Figure 2.2. Similar to the proof of Lemma 2.14, we get a symmetric $\mathbb{Z}_2$-signature $[0,1,0,0,0]$. Thus, by Lemma 2.14, all PARITY, $\mathcal{EO}$, and their reversal are realizable.

Moreover, connect $[0,1,0,\ldots,0]$ of arity k with $[1,1,0,0]$ via $[0,1,0]$, resulting in $[1,1,0,\ldots,0]$ of arity $k+1$. In this way, we can construct every ATMOSTONE $\mathbb{Z}_2$-signature. Again, their reversal can be obtained by putting $[0,1,0]$ on every edge. □

The realizability part of Lemma 2.12 follows from Lemma 2.15 since $[0,1]$, $[1,0]$, and $[1,1,0,0]$ are all allowed to build $\mathbb{Z}_2$-signatures in $\mathcal{M}_\oplus$.

2.2 Hardness Results

In this section, we prove several hardness results regarding $\oplus\,\mathrm{Holant}^c$ problems. These results are preparations for proving dichotomy theorems. Most of them are for low arity signature sets. As we will see later, general cases will boil down to lower arity cases.

A Hardness Starting Point

To begin with, we first consider the problem of $\oplus$Pl-Rtw-Mon-3CNF. Pl-Rtw-Mon-3CNF is a special case of the satisfying problem for 3CNF formulas, with the requirement of "PL", "Rtw", and "Mon". "PL" means that inputs are restricted to planar (constraint) graphs. "Rtw" (read twice) means that every variable only appears twice in clauses. "Mon" (monotone) means that for every variable only itself or its negation appears, but not both. Therefore we may assume all variables appear in positive forms. $\oplus$Pl-Rtw-Mon-3CNF is to compute the parity of all satisfying assignments of such instances.

To rewrite $\oplus$Pl-Rtw-Mon-3CNF in the Holant setting, we use vertices to represent all clauses and variables. Draw an edge between a clause vertex c and a variable vertex x if x appears c. Due to the requirements "Rtw" and "3CNF", the resulting graph is a 2-3 bipartite graph. Moreover, a variable that only appears positively is the same as the signature $=_2$, which means it can be viewed as an edge. The signature on each clause vertex is $[0, 1, 1, 1]$ since it is an "OR". Hence, in the Holant language, this problem becomes Planar $\oplus\,\mathrm{Holant}([0, 1, 1, 1])$.

Valiant [Val06, Theorem 2.3] showed that $\oplus$Pl-Rtw-Mon-3CNF is $\oplus$**P**-complete. Translating to the Holant framework, we have the following. Note that the problem $\oplus\,\mathrm{Holant}([0, 1, 1, 1])$ is equivalent to $\oplus\,\mathrm{Holant}([1, 1, 1, 0])$ by complementing all assignments.

Proposition 2.16. *Planar* $\oplus\,\mathrm{Holant}([0, 1, 1, 1])$ *and Planar* $\oplus\,\mathrm{Holant}([1, 1, 1, 0])$ *are* $\oplus$**P**-*complete.*

Remark All hardness results in this chapter for $\oplus\,\mathrm{Holant}^c$ hold even if we restrict the input to planar graphs. This is because Proposition 2.16 holds for planar graphs, and all reductions are also planar. On the other hand, to prove the $\oplus\,\mathrm{Holant}$ dichotomy we will employ some non-planar reductions.

To apply holographic transformations, we need to rewrite $\oplus\text{Holant}([0,1,1,1])$ in the bipartite setting, which is $\oplus\text{Holant}([1,0,1]|[0,1,1,1])$. Under the transformation $T = \left[\begin{smallmatrix}1 & 1\\ 0 & 1\end{smallmatrix}\right]$, we have that

$$\oplus\text{Holant}([1,0,1]|[0,1,1,1]) \equiv \oplus\text{Holant}([1,1,0]|[1,0,0,1]).$$

Since this is the first time to see holographic transformation in action, we will calculate it in detail. Later we shall not repeat this calculation for brevity. Let $f = [1,0,1]$, $g = [0,1,1,1]$. By Theorem 1.1, we have that $\oplus\text{Holant}(f \mid g) \equiv \oplus\text{Holant}(fT^{\otimes 2} \mid \left(T^{-1}\right)^{\otimes 3} g)$. Now we compute

$$fT^{\otimes 2} = \begin{bmatrix}1 & 0 & 0 & 1\end{bmatrix} \cdot \begin{bmatrix}1 & 1 & 1 & 1\\ 0 & 1 & 0 & 1\\ 0 & 0 & 1 & 1\\ 0 & 0 & 0 & 1\end{bmatrix} = \begin{bmatrix}1 & 1 & 1 & 2\end{bmatrix}.$$

As a $\mathbb{Z}_2$-signature, $fT^{\otimes 2} = [1,1,0]$. On the other hand, note that $T^{-1} \equiv T \pmod 2$. Hence we have

$$\left(T^{-1}\right)^{\otimes 3} g = \begin{bmatrix}1 & 1 & 1 & 1 & 1 & 1 & 1 & 1\\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1\\ 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1\\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1\\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1\\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1\\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1\\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1\end{bmatrix} \cdot \begin{bmatrix}0\\1\\1\\1\\1\\1\\1\\1\end{bmatrix} = \begin{bmatrix}7\\4\\4\\2\\4\\2\\2\\1\end{bmatrix}$$

As a $\mathbb{Z}_2$-signature, $\left(T^{-1}\right)^{\otimes 3} g = [7,4,2,1] = [1,0,0,1]$. Thus, the problem after transformation is $\oplus\text{Holant}([1,1,0]|[1,0,0,1])$. Hence we have the following corollary. Again the second problem is equivalent to the first by flipping all assignments.

Corollary 2.17. $\oplus\text{Holant}([1,1,0]|[1,0,0,1])$ *and* $\oplus\text{Holant}([0,1,1]|[1,0,0,1])$ *are* $\oplus$**P***-complete.*

In fact, the result above is also shown in [Val06]. The problem $\oplus\,\text{Holant}([1,1,0] | [1,0,0,1])$ is called $\oplus$Pl-3/2-Bip-VC, and is further equivalent to $\oplus\,\text{Holant}([1,1,0] | [1,1,1,0])$ under holographic transformations.

More Hardness Results

Next we establish some further hardness results for $\oplus\,\text{Holant}^c$ problems. First comes a quick generalization of Corollary 2.17. Recall that $=_n$ is the EQUALITY signature of arity n.

Corollary 2.18. *If $n \geqslant 3$, then $\oplus\,\text{Holant}^c([0,1,1], =_n)$ is $\oplus$**P**-complete.*

Proof. If $n = 3$, $\oplus\,\text{Holant}([0,1,1],[1,0,0,1])$ is $\oplus$**P**-complete by Corollary 2.17. Otherwise we show how to construct $=_3$ from $=_n$. Connecting $[0,1]$ with $[0,1,1]$ gives the unary $[1,1]$. Connecting $=_n$ with $(n-3)$ many $[1,1]$'s gives $=_3$. □

Then we deal with the case when the signature set intersects both $\mathcal{P}_\oplus$ and $\mathcal{M}_\oplus$. We first show a base case and then reduce the general case to it.

Lemma 2.19. *$\oplus\,\text{Holant}^c([0,1,0,1,0],[0,1,1,0])$ is $\oplus$**P**-complete.*

Proof. First we claim that $\oplus\,\text{Holant}([0,1,0,0,0],[0,1,1,0])$ is $\oplus$**P**-complete. Construct the gadget in Figure 2.3. We put $[0,1,0,0,0]$ on circles and $[0,1,1,0]$ on the square. The resulting signature is $[1,1,1,0]$. Thus, $\oplus\,\text{Holant}([1,1,1,0]) \leqslant_T \oplus\,\text{Holant}([0,1,0,0,0],[0,1,1,0])$. By Proposition 2.16, $\oplus\,\text{Holant}([1,1,1,0])$ is $\oplus$**P**-complete. The claim is proved.

Under a holographic transformation of $T = \left[\begin{smallmatrix}1 & 0\\ 1 & 1\end{smallmatrix}\right]$, we have that

$$\begin{aligned}\oplus\,\text{Holant}\left([0,1,0,0,0],[0,1,1,0]\right) &\equiv \oplus\,\text{Holant}\left([1,0,1]T^{\otimes 2} \mid \left(T^{-1}\right)^{\otimes 4}[0,1,0,0,0],\right.\\ &\qquad\qquad\left.\left(T^{-1}\right)^{\otimes 3}[0,1,1,0]\right)\\ &\equiv \oplus\,\text{Holant}\left([0,1,1] \mid [0,1,0,1,0],[0,1,1,0]\right).\end{aligned}$$

Note that we get $[0,1,1]$ by connecting $[0,1,1,0]$ to $[1,0]$. Hence

$$\oplus\,\text{Holant}\left([0,1,1] \mid [0,1,0,1,0],[0,1,1,0]\right) \leqslant_T \oplus\,\text{Holant}^c\left([0,1,0,1,0],[0,1,1,0]\right).$$

Combine everything and the lemma is proved. □

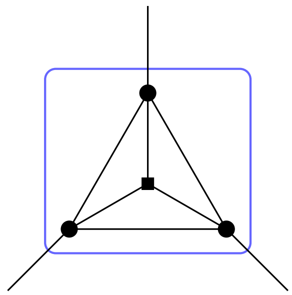


Figure 2.3: The gadget to construct $[1, 1, 1, 0]$.

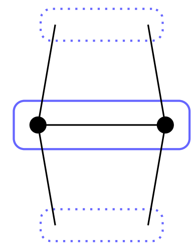


Figure 2.4: An asymmetric gadget for domain pairing.

Corollary 2.20. *Let* f *and* g *be two non-degenerate* $\mathbb{Z}_2$*-signatures of arity* n *and* m*, respectively, such that* $n, m \geqslant 3$*,* $f \in \mathcal{P}_\oplus$*, and* $g \in \mathcal{M}_\oplus$*. Then* $\oplus\text{Holant}^c(f, g)$ *is* $\oplus$**P***-complete.*

Proof. First, we connect $n-3$ many $[1, 0]$'s to f to get an arity 3 signature f'. Since $f \in \mathcal{P}_\oplus$ and is non-degenerate, f' is in the set $\{[0, 1, 1, 0], [1, 0, 1, 1], [1, 1, 0, 1]\}$. By Lemma 2.5, we can construct all signatures in $\mathcal{P}_\oplus$, including $[0, 1, 1, 0]$, from f', $[1, 0]$, and $[0, 1]$.

By Lemma 2.12, we may assume that g is in $\mathcal{EO}$, $\mathcal{AO}$, or Parity. Connect g with $m-3$ many $[1, 0]$'s. The resulting signature g' is one of $[0, 1, 0, 0]$, $[1, 1, 0, 0]$, $[0, 1, 0, 1]$, or $[1, 0, 1, 0]$. Apply Lemma 2.13, Lemma 2.14, or Lemma 2.15, and we can always construct all Parity signatures, including $[0, 1, 0, 1, 0]$, from g', $[1, 0]$, and $[0, 1]$.

Hence $\oplus\text{Holant}^c([0, 1, 0, 1, 0], [0, 1, 1, 0]) \leqslant_T \oplus\text{Holant}^c(f, g)$. By Lemma 2.19, $\oplus\text{Holant}^c(f, g)$ is $\oplus$**P**-complete. □

Corollary 2.20 implies that any mixing of non-trivial $\mathbb{Z}_2$-signatures from $\mathcal{P}_\oplus$ and $\mathcal{M}_\oplus$ leads to $\oplus$**P**-hardness. Similarly, we have the following no-mixing lemma of $\mathcal{M}_\oplus$ and $\mathcal{EQ}$.

In the following proof we will use a technique called *domain pairing*. It is first employed in [CLX10, Lemma III.2] for real weighted Pl-#CSP. It was also used in [HL12, Lemma IV.5] with real weights as well as grouping more than just two domain elements.

Given a signature f of arity $2n$, we view each pair of inputs of f as a new input. Effectively this is a signature of arity n on a domain of size $2 \times 2 = 4$. However, we will use other signatures to ensure that the values on each pair of inputs are always the same. Hence the domain size is projected down back to 2. Usually the gadget we use is not symmetric geometrically, like the one in Figure 2.4, and its signature is also asymmetric. However, by this domain pairing argument (edges are paired by the dotted circle in Figure 2.4), the resulting signature is indeed symmetric. To ensure the same value constrict, we usually work in a bipartite graph, construct an equality signature on the other side, pair its edges, and use it as an equality of half the arity.

Lemma 2.21. $\oplus\text{Holant}([0,0,1,0], =_6)$ *is* $\oplus$**P**-*complete.*

Proof. We connect one edge of two $[0,1,0,0]$'s as in Figure 2.4, getting an asymmetric signature f. As discussed above for domain pairing, we will enforce that each pair of dotted edges of f takes the same value. Under this guarantee, f is $[1,1,0]$ of arity 2 in the paired domain. In other words, the domain pairing is the following reduction

$$\oplus\text{Holant}([1,1,0] \mid =_3) \leqslant_T \oplus\text{Holant}(f \mid =_6).$$

Given an instance $\Omega = (G, \pi)$ of $\oplus\text{Holant}([1,1,0] \mid =_3)$, we replace each edge by a pair of parallel edges, and put f on the left and $=_6$ on the right. This is an instance of $\oplus\text{Holant}(f \mid =_6)$. It is easy to verify that this does not change the $\oplus$Holant. By Corollary 2.17, $\oplus\text{Holant}([1,1,0] \mid =_3)$ is $\oplus$**P**-complete and so is $\oplus\text{Holant}^c([0,1,0,0], =_n)$. □

Lemma 2.21 can be generalized for higher arity equalities and AtMostOne signatures.

Lemma 2.22. *If* $n \geqslant 3$, *then* $\oplus\text{Holant}^c([0,0,1,0], =_n)$, $\oplus\text{Holant}^c([0,1,0,0], =_n)$, $\oplus\text{Holant}^c([0,0,1,1], =_n)$ *and* $\oplus\text{Holant}^c([1,1,0,0], =_n)$ *are all* $\oplus$**P**-*complete.*

Proof. By symmetry, we only need to prove the lemma for $[0,1,0,0]$ and $[1,1,0,0]$. By Lemma 2.15, we can realize $[0,1,0,0]$ from $[1,1,0,0]$, $[0,1]$, and $[1,0]$. Hence it is sufficient to prove $\oplus$**P**-hardness for $\oplus\text{Holant}^c([0,1,0,0], =_n)$.

We reduce the arity of $=_n$ by doing self-loops. Eventually it becomes $=_3$ or $=_4$, depending on the parity of n. In either case we can realize $=_6$ by connecting four $=_3$'s or two $=_4$'s together. Hence we are done by Lemma 2.21. □

Lemma 2.22 implies the following corollary for signatures that contain both Equality and ExactOne or AtMostOne as subsignatures.

Corollary 2.23. *$\oplus\text{Holant}^c([1,0,\ldots,0,1,0])$ and $\oplus\text{Holant}^c([1,0,\ldots,0,1,1])$ are $\oplus$**P**-complete, if the number of zeroes is at least 2.*

Finally, there's still two more cases not covered by all above. They can be treated in more or less the same manner, and are summarized in the next lemma. We will use Lemma 2.21 again.

Lemma 2.24. *$\oplus\text{Holant}^c([0,0,1,0,0])$ and $\oplus\text{Holant}^c([0,0,1,0,1])$ are $\oplus$**P**-complete.*

Proof. Note that $[0,0,1,0]$ is a subsignature of both $[0,0,1,0,0]$ and $[0,0,1,0,1]$. We only need to construct $=_6$ to apply Lemma 2.21.

If we place $[0,0,1,0,0]$ at both the circle and the triangle in the gadget shown in Figure 2.5, the resulting signature is $=_4$. We get $=_6$ by connecting one edge of two $=_4$'s. Hence $\oplus\text{Holant}^c([0,0,1,0,0])$ is $\oplus$**P**-complete.

The case of $[0,0,1,0,1]$ is more complicated. First get $[0,1,0]$ as it is a subsignature of $[0,0,1,0,1]$. Hence we have $[1,0,1,0,0]$ by connecting $[0,1,0]$ to each edge of $[0,0,1,0,1]$. Then we place this $[1,0,1,0,0]$ at the triangle and $[0,0,1,0,1]$ at the circle in the gadget in Figure 2.5. The combined gadget is depicted in Figure 2.6, where circles are $[0,0,1,0,1]$ and squares are $[0,1,0]$. The resulting signature is again $[1,0,0,0,1]$. The remaining proof is the same as the case of $[0,0,1,0,0]$. □

We remark that the gadget in Figure 2.5 does not necessarily realize a symmetric $\mathbb{Z}_2$-signature, but it does with the specific signatures we put in the proof above.

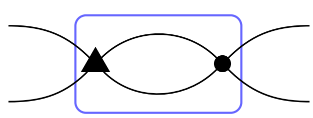


Figure 2.5: The gadget to construct $[1,0,0,0,1]$ from $[0,0,1,0,0]$.

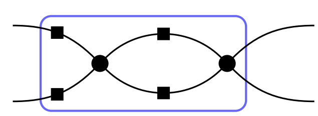


Figure 2.6: The gadget to construct $[1,0,0,0,1]$ from $[0,0,1,0,1]$.

2.3 Parity Holant^c Dichotomy

Based on the algorithms in Section 2.1 and the hardness results in Section 2.2, we are now ready to show the dichotomy theorem for $\oplus\text{Holant}^c$ problems, which is a stepping stone towards the final dichotomy. We show a single signature dichotomy first, and then generalize it to sets.

Lemma 2.25. *Let f be a $\mathbb{Z}_2$-signature such that $f \notin \mathcal{A}_\oplus \cup \mathcal{P}_\oplus \cup \mathcal{M}_\oplus$. Then* $\oplus\text{Holant}^c(f)$ *is* $\oplus$**P**-*complete.*

The proof is a case-by-case analysis. Basically we want to discuss in terms of the maximum number of consecutive '0' bits and then that of '1' bits in its symmetric form. We need to rule out some simple cases before that.

Proof. First notice that $\mathcal{M}_\oplus$ contains all signatures with arity less than or equal to two, as well as all degenerate signatures. Thus, f has arity $n \geqslant 3$ and is non-degenerate. Also note that the reversal of f has the same complexity as f by flipping. Hence we will often ignore its reversal.

We also rule out some patterns that will appear later more than once. Assume f contains one of $[0,1,1,0]$, $[1,0,1,1]$, one $[1,1,0,1]$ as a subsignature. Because $f \notin \mathcal{P}_\oplus$, it must extend that subsignature in either or both directions. Thus f must contain $[0,1,1,0,0]$, $[1,0,1,1,1]$, $[1,1,0,1,0]$ or their reversals as a subsignature. However, any of the three defines a $\oplus$**P**-complete problem, and their reversals have the same complexity by flipping.

- For $[0,1,1,0,0]$, $\oplus\text{Holant}^c(f)$ is $\oplus$**P**-complete by Corollary 2.20 since it contains both a Fibonacci $\mathbb{Z}_2$-signature $[0,1,1,0]$ and ExactOne$_3$ $[1,1,0,0]$ as subsignatures.
- For $[1,0,1,1,1]$, $\oplus\text{Holant}^c(f)$ is $\oplus$**P**-complete by Proposition 2.16 since it contains $[0,1,1,1]$ as a subsignature.
- For $[1,1,0,1,0]$, $\oplus\text{Holant}^c(f)$ is $\oplus$**P**-complete by Corollary 2.20 since it contains both a Fibonacci $\mathbb{Z}_2$-signature $[1,1,0,1]$ and Parity $[1,0,1,0]$ as subsignatures.

Hence we can assume that f does no contains any of $[0,1,1,0]$, $[1,0,1,1]$, or $[1,1,0,1]$ as its subsignature.

Now we consider the maximum number of consecutive '0' bits of f in its symmetric form. First we assume f contains at least 2 consecutive 0's. Then consider a sequence of consecutive 0's of the maximum length k_0 in f. If both ends of this sequence are 1, then f contains a subsignature of the form $[1,0,\dots,0,1,0]$, $[1,0,\dots,0,1,1]$ or their reversals, because otherwise f is an Equality signature $=_n \in \mathcal{A}_\oplus$. Then by Corollary 2.23, $\oplus\text{Holant}^c(f)$ is $\oplus$**P**-complete.

Otherwise, only one end of these 0's is 1. Hence we may assume the first k_0 bits of f are 0 as we may flip f if necessary. Consider the number of subsequent ones after these zeroes. We have the following 3 cases.

- If there are more than 2 ones, then we have $[0, 1, 1, 1]$ as its subsignature and we are done by Proposition 2.16.
- If there are 2, f cannot end here because the partial matching gate $[0, \ldots, 0, 1, 1]$ is in $\mathcal{M}$. Then we have $[0, 0, 1, 1, 0]$ as a subsignature of f. Hence $\oplus\text{Holant}^c(f)$ is $\oplus$**P**-complete as discussed earlier.
- If there is only 1, f cannot end there because $[0, \ldots, 0, 1]$ is degenerate. Also because $[0, \ldots, 0, 1, 0]$ is in $\mathcal{M}$, f must have the form $[0, \ldots, 0, 1, 0, 0]$ or $[0, \ldots, 0, 1, 0, 1]$. By Lemma 2.24 both cases are $\oplus$**P**-complete.

The case left is when f contains at most 1 consecutive 0's. Consider the maximum number k_1 of consecutive 1's in f. If $k_1 \geqslant 3$, f must contain $[0, 1, 1, 1]$ or its reversal and we get $\oplus$**P**-completeness by Proposition 2.16. If $k_1 = 1$, f must be a PARITY signature which belongs to $\mathcal{M}_\oplus$. Contradiction. Thus $k_1 = 2$. In that case f must contain a Fibonacci $\mathbb{Z}_2$-signature $[0, 1, 1, 0]$, $[1, 0, 1, 1]$ or $[1, 1, 0, 1]$ as its subsignature, which is already shown to imply $\oplus$**P**-completeness. □

Given Lemma 2.25, the remaining case to deal with is that $\mathcal{F} \subseteq \mathcal{A}_\oplus \cup \mathcal{P}_\oplus \cup \mathcal{M}_\oplus$, but $\mathcal{F}$ is not a subset of any of them. The next lemma shows that this case also implies $\oplus$**P**-complete.

Lemma 2.26. *Let $\mathcal{F}$ be a set of signatures. If $\mathcal{F} \subseteq \mathcal{A}_\oplus \cup \mathcal{P}_\oplus \cup \mathcal{M}_\oplus$, but $\mathcal{F} \nsubseteq \mathcal{A}_\oplus$, $\mathcal{F} \nsubseteq \mathcal{P}_\oplus$, and $\mathcal{F} \nsubseteq \mathcal{M}_\oplus$, then $\oplus\text{Holant}^c(\mathcal{F})$ is $\oplus$**P**-complete.*

Proof. Since $\mathcal{F} \nsubseteq \mathcal{M}_\oplus$ and every $\mathbb{Z}_2$-signature with arity at most 2 is a matchgate, there must exist $f \in \mathcal{F}$ of arity $n \geqslant 3$ which is not a matchgate. Therefore f is either $=_n$ or a Fibonacci signature.

We deal with the case that f is $=_n$ first. Since $\mathcal{F}$ is not a subset of $\mathcal{A}_\oplus$, there must exist $g \in \mathcal{F}$ such that $g \notin \mathcal{A}_\oplus$ but $g \in \mathcal{P}_\oplus \cup \mathcal{M}_\oplus$. Since $\mathcal{A}_\oplus$ contains all degenerate $\mathbb{Z}_2$-signatures, g is non-degenerate and of arity at least 2. Moreover, if g has arity 2, then g is either $[0, 1, 1]$ or $[1, 1, 0]$ as $g \notin \mathcal{A}_\oplus$. By Corollary 2.18, $\oplus\text{Holant}^c(f, g)$ is $\oplus$**P**-complete and so is $\oplus\text{Holant}^c(\mathcal{F})$. Otherwise

g has arity $\geqslant 3$. Since $g \in \mathcal{P}_\oplus \cup \mathcal{M}_\oplus$ but $g \notin \mathcal{A}_\oplus$, g is Fibonacci, ExactOne, or AtMostOne. Hence g contains a subsignature $[0,1,1]$, $[1,1,0]$, or $[0,1,0,0]$. By Corollary 2.18 and Lemma 2.22, $\oplus \text{Holant}^c(\mathcal{F})$ is $\oplus$**P**-complete in any of the cases.

The other case is when f is Fibonacci. Then f contains either $[0,1,1]$ or $[1,1,0]$ as a subsignature. Because $\mathcal{F}$ is not a subset of $\mathcal{P}_\oplus$, there must exist $g \in \mathcal{F}$ such that $g \notin \mathcal{P}_\oplus$ but $g \in \mathcal{A}_\oplus \cup \mathcal{M}_\oplus$. Since $\mathcal{P}_\oplus$ contains all degenerate $\mathbb{Z}_2$-signatures, g is non-degenerate and has arity $\geqslant 3$. If g is Equality, then by Corollary 2.18, $\oplus \text{Holant}^c(f, g)$ is $\oplus$**P**-complete and so is $\oplus \text{Holant}^c(\mathcal{F})$. Otherwise, $g \in \mathcal{M}_\oplus$. By Corollary 2.20, $\oplus \text{Holant}^c(f, g)$ is $\oplus$**P**-complete and so is $\oplus \text{Holant}^c(\mathcal{F})$. □

By Lemma 2.2, Lemma 2.4, and Lemma 2.7, if $\mathcal{F} \subseteq \mathcal{A}_\oplus$, $\mathcal{F} \subseteq \mathcal{P}_\oplus$, or $\mathcal{F} \subseteq \mathcal{M}_\oplus$, then $\oplus \text{Holant}^c(\mathcal{F})$ is computable in polynomial time. Together with Lemma 2.25 and Lemma 2.26, we have the dichotomy of $\oplus \text{Holant}^c$ problems.

Theorem 2.27. $\oplus \text{Holant}^c(\mathcal{F})$ *is* $\oplus$**P**-*complete, unless* $\mathcal{F} \subseteq \mathcal{A}_\oplus$, $\mathcal{F} \subseteq \mathcal{P}_\oplus$ *or* $\mathcal{F} \subseteq \mathcal{M}_\oplus$. *In any of the exceptional cases* $\oplus \text{Holant}^c(\mathcal{F})$ *can be computed in polynomial time. The same statement also holds for planar graphs.*

2.4 Parity Vanishing Signature Sets

In the remaining two sections we extend our results to obtain the dichotomy result for $\oplus$Holant without assuming any available functions. One key ingredient to the full dichotomy is a new tractable family which we call *Vanishing*.

Definition 2.28. *A set* $\mathcal{F}$ *of* $\mathbb{Z}_2$*-signatures is called vanishing if the value of* $\oplus \text{Holant}(\Omega; \mathcal{F})$ *is zero for every* Ω. *A single* $\mathbb{Z}_2$*-signature* f *is called vanishing if* $\{f\}$ *is. Denote by* $\mathcal{V}_\oplus$ *the class of all vanishing signature sets.*

We note that Definition 2.28 in fact defines a family of tractable $\mathbb{Z}_2$-signatures, instead of just a set. It is possible that both $\mathcal{F}_1$ and $\mathcal{F}_2$ are vanishing, but $\mathcal{F}_1 \cup \mathcal{F}_2$ is not.

We first show some basic properties of $\mathcal{V}_\oplus$. For two $\mathbb{Z}_2$-signatures f and g of the same arity, $f + g$ denotes the bitwise addition in $\mathbb{Z}_2$, i.e. for any input x, $(f+g)(x) = f(x) + g(x)$.

Lemma 2.29. *Let $\mathcal{F}$ be a vanishing $\mathbb{Z}_2$-signature set.*

- *If f is an $\mathcal{F}$-gate, then $\mathcal{F} \cup \{f\} \in \mathcal{V}_{\oplus}$.*
- *If g_0 and g_1 are two $\mathbb{Z}_2$-signatures in $\mathcal{F}$ of the same arity, then $\mathcal{F} \cup \{g_0 + g_1\} \in \mathcal{V}_{\oplus}$.*

Proof. The first statement is trivial. We prove the second, which says that a vanishing signature set is closed under linear combinations.

Let $\Omega = (G, \pi)$ be an instance of $\oplus\text{Holant}(\mathcal{F} \cup \{g_0 + g_1\})$. We want to show that $\text{Holant}_\Omega = 0$. If the signature $g_0 + g_1$ does not appear in H, then Holant_Ω is zero since $\mathcal{F} \in \mathcal{V}_{\oplus}$. Otherwise, we denote by U the set of vertices having the signature $g_0 + g_1$. Then:

$$\begin{aligned}
\text{Holant}_\Omega &= \sum_{\sigma: V \to \{0,1\}} \prod_{v \in V} f_v(\sigma|_{E(v)}) \\
&= \sum_{\sigma} \prod_{v \notin U} f_v(\sigma|_{E(v)}) \prod_{v \in U} \left(g_0(\sigma|_{E(v)}) + g_1(\sigma|_{E(v)})\right) \\
&= \sum_{\sigma} \prod_{v \notin U} f_v(\sigma|_{E(v)}) \left(\sum_{\tau: U \to \{0,1\}} \prod_{v \in U} g_{\tau(v)}(\sigma|_{E(v)}) \right) \\
&= \sum_{\tau: U \to \{0,1\}} \left(\sum_{\sigma} \prod_{v \notin U} f_v(\sigma|_{E(v)}) \prod_{v \in U} g_{\tau(v)}(\sigma|_{E(v)}) \right),
\end{aligned}$$

where σ is an assignment, $E(v)$ denotes the incident edges of v, and $\sigma|_{E(v)}$ denotes the restriction of σ to $E(v)$. In the final line, we rewrote Holant_Ω into an exponential sum over all configurations τ on U, where every term in the bracket is a Holant value on Ω but with $g_1 + g_2$ on v replaced by $g_{\tau(v)}$ for every $v \in U$. These are all instances of $\oplus\text{Holant}(\mathcal{F})$, and therefore all terms are zero since $\mathcal{F} \in \mathcal{V}_{\oplus}$. The summation Holant_Ω, albeit exponential, is also zero. This completes the proof. □

In the following two subsections, we mention some simple examples of $\mathcal{V}_{\oplus}$. They are not really used in the dichotomy proof, but one may get some intuition and there are some interesting phenomena. In the last part of this section, we will introduce *self-vanishable* $\mathbb{Z}_2$-signatures, which is crucial in the proof of the full dichotomy.

Complement Invariant Signatures

The simplest example of vanishing $\mathbb{Z}_2$-signatures are complement invariant ones. A signature f of arity n is called *complement invariant* if for $\alpha \in \{0,1\}^n$, $f(\alpha) = f(\overline{\alpha})$.

If all $\mathbb{Z}_2$-signatures involved in a Holant instance Ω are complement invariant, then any assignment of edges and its complement yield the same value, and hence they are cancelled in $\oplus$Holant. In the end $\oplus$Holant is always 0.

Proposition 2.30. *Let $\mathcal{F}$ be a set of complement invariant signatures. Then $\mathcal{F}$ is vanishing.*

As a side note, this family of signature sets corresponds to the additional tractable case of in Faben's parity CSP dichotomy [Fab08].

Matching Based Vanishing Signature Sets

Next we describe another family of vanishing signature sets. In a graph where all nodes have even degrees the parity of the number of perfect matchings is even. This can be easily shown using (2.1). The parity of perfect matchings is equal to that of the determinant of its adjacency matrix. Adding up all rows of the adjacency matrix, we get a vector composed of even numbers. Thus this matrix must be singular in $\mathbb{Z}_2$ and its determinant is zero.

Furthermore, using (2.2), we make the same claim for Parity Matching Problems defined in Section 2.1. If unomittable nodes have even degrees and omittable nodes have odd degrees, then the parity of desired matchings is always even. By (2.2), the parity of the number of such matchings equals $\mathrm{Pf}(A+\Lambda^{(n)})$ if n, the number of nodes, is even, or $\mathrm{Pf}(A^{+}+\Lambda^{(n+1)})$ if n is odd. As the number of vertices of odd degrees must be even, it is easy to verify that the summation of all rows in $A+\Lambda^{(n)}$ for even n, or the first n rows in $A^{+}+\Lambda^{(n+1)}$ for odd n is a zero vector in $\mathbb{Z}_2$. Hence, the Pfaffian, which equals the determinant, is zero in $\mathbb{Z}_2$. In the Holant language, unomittable nodes of even degrees have $\mathbb{Z}_2$-signatures $\textsc{ExactOne}_{2k}$, and omittable nodes of odd degrees have $\textsc{AtMostOne}_{2k'-1}$, for some integers $k, k' \geqslant 1$.

By Lemma 2.29, the linear combination of these matching signatures, or signatures that can be realized from them, also belong to this vanishing signature family.

Proposition 2.31. *If a $\mathbb{Z}_2$-signature set $\mathcal{F}$ is composed of* ExactOne *of even arities,* AtMostOne *of odd arities, $\mathbb{Z}_2$-signatures realizable from them, and linear combinations of all above, then $\mathcal{F} \in \mathcal{V}_\oplus$.*

Note that all $\mathbb{Z}_2$-signatures realized from ExactOne_{2k} and $\text{AtMostOne}_{2k'-1}$ are in $\mathcal{M}_\oplus$, and they do not provide any new tractable set. However, the linear combination can bring us outside of $\mathcal{M}_\oplus$. Then $\mathbb{Z}_2$-signatures realized from these linear combinations are also outside of $\mathcal{M}_\oplus$. Indeed, there are some sets $\mathcal{F}$ such that $\oplus\text{Holant}^c(\mathcal{F})$ is $\oplus$**P**-complete, but $\mathcal{F}$ is vanishing, and thus $\oplus\text{Holant}(\mathcal{F})$ is tractable.

Example 2.32. *By Theorem 2.27 $\oplus\text{Holant}^c(\{[1,1,1,0,1]\})$ is $\oplus$**P**-complete. However, $[1,1,1,0,1]$ is vanishing because $[1,1,1,0,1] = [0,1,0,0,0] + [1,0,1,0,1]$, where $[0,1,0,0,0]$ is ExactOne_4 and $[1,0,1,0,1]$ can be realized from ExactOne_4 (the construction is similar to that in Lemma 2.14).*

Self-Vanishable Signatures

In this section, we introduce a new notion called *self-vanishable signatures* which plays an important role in the proof of the full dichotomy. First, we introduce an extended version of the inner product for two signatures with not necessarily the same arity.

Definition 2.33. *Let f and g be two signatures with arities n and m ($n \geqslant m$) respectively. Their inner product $h = \langle f, g \rangle$ is a signature with arity $n - m$ defined as follows:*

$$h(\alpha) = \sum_{\beta \in \{0,1\}^m} f(\beta, \alpha) g(\beta),$$

for every $\alpha \in \{0,1\}^{n-m}$.

If f is symmetric, then $h = \langle f, g \rangle$ is also symmetric. If both f and g are symmetric, then their inner product $h = [h_0, h_1, \ldots, h_{n-m}]$ has the following form: $h_i = \sum_{j=0}^{m} \binom{m}{j} f_{j+i} g_j$ for $0 \leqslant i \leqslant n - m$. Hence, in $\mathbb{Z}_2$,

$$\langle f, [1,1]^{\otimes 2} \rangle = \langle f, [1,1,1] \rangle = \langle f, [1,0,1] \rangle, \tag{2.4}$$

since $\binom{2}{1} = 2$. We will need this simple fact in future.

We can also view this inner product in a combinatorial way. Given two gates with signatures f and g, connecting (the first) m dangling edges of f to g (see Figure 2.7), results in a gadget with signature $\langle f, g \rangle$.

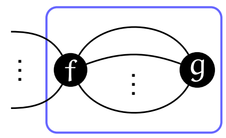


Figure 2.7: The signature inner product.

Definition 2.34. *A $\mathbb{Z}_2$-signature f of arity n is called self-vanishable of degree k if there exists a unique non-negative integer $k < n$ such that $\langle f, [1,1]^{\otimes k+1} \rangle = \mathbf{0}$ and $\langle f, [1,1]^{\otimes k} \rangle \neq \mathbf{0}$, denoted by $\mathrm{rd}(f) = k$. Define $\mathrm{rd}(\mathbf{0}) := -1$. If such a k does not exist, then f is not self-vanishable and define $\mathrm{rd}(f) := n$.*

The notation, $\mathrm{rd}(\cdot)$, stands for the recurrence degree. We will see in Lemma 2.37 that entries of f with $\mathrm{rd}(f) = d$ satisfy a linear recurrence of degree d.

A vanishing $\mathbb{Z}_2$-signature is necessarily self-vanishable, as shown in the following proposition. It also partly explains the intuition why we define this notion in this way.

Proposition 2.35. *If f is vanishing, then f is self-vanishable.*

Proof. Let n be the arity of f. If $n = 2k$ is even, construct an instance of one single vertex of k many self loops, and put f on it. The resulting signature is of arity 0, which means it is a single value, and the value is $\langle f, [1,0,1]^{\otimes k} \rangle$. However, (2.4) implies that

$$\langle f, [1,1]^{\otimes 2k} \rangle = \langle f, [1,1,1]^{\otimes k} \rangle = \langle f, [1,0,1]^{\otimes k} \rangle.$$

Since f is not self-vanishable, $\langle f, [1,1]^{\otimes 2k} \rangle \neq 0$. Hence f is not vanishing, and neither is $\mathcal{F}$.

If $n = 2k+1$ is odd, then we do k many self-loops of f. The resulting $\mathbb{Z}_2$-signature is arity 1, and its signature is $g = \langle f, [1,0,1]^{\otimes k} \rangle = \langle f, [1,1]^{\otimes 2k} \rangle$. This g cannot be $[0,0]$ or $[1,1]$ for f is not self-vanishable. Hence g must be either $[0,1]$ or $[1,0]$. In either case, connect two copies of g. We get an instance whose $\oplus$Holant is 1. Again f is not vanishing, and neither is $\mathcal{F}$. $\square$

The following lemma is immediate by induction.

Lemma 2.36. *Let* f *be self-vanishable of degree* $d \geqslant 0$*. Let* r *be an integer that* $0 < r \leqslant d+1$*. Then* $\mathrm{rd}(\langle f, [1,1]^{\otimes r}\rangle) = d - r$*.*

It is easy to verify that for a symmetric $\mathbb{Z}_2$-signature $f = [f_0, f_1, \ldots, f_n]$, we have that

$$\langle f, [1,1]\rangle = [f_0 + f_1, f_1 + f_2, \ldots, f_{n-1} + f_n].$$

Hence the only symmetric $\mathbb{Z}_2$-signature of arity n with $\mathrm{rd}(f) = 0$ is $[1,1]^{\otimes n}$. In other words, it satisfies a linear relation $f_i = 1$ for any $0 \leqslant i \leqslant n$. In general, we have the following.

Lemma 2.37. *Let* $f = [f_0, f_1, \cdots, f_n]$ *be a symmetric* $\mathbb{Z}_2$*-signature with arity* n*,* $\mathrm{rd}(f) = d$ *and* $0 \leqslant d < n$*. Then we have that*

$$\sum_{j=0}^{d} \binom{d}{j} f_{i+j} = 1, \tag{2.5}$$

for any $0 \leqslant i \leqslant n - d$*.*

Proof. We prove it by induction on d. The base case is when $d = 0$. As shown above, the only possible $\mathbb{Z}_2$-signature is $[1,1]^{\otimes n}$, and it satisfies $f_i = 1$ for any $0 \leqslant i \leqslant n$.

Now suppose $d > 0$ and the lemma holds for any $0 \leqslant k < d$. Let $f' = \langle f, [1,1]^{\otimes r}\rangle$ of arity $n-1$. By Lemma 2.29, $\mathrm{rd}(f') = d - 1$. By induction hypothesis,

$$\sum_{j=0}^{d-1} \binom{d-1}{j} f'_{i+j} = 1, \tag{2.6}$$

for any $0 \leqslant i \leqslant n - d$. Moreover, $f'_i = f_i + f_{i+1}$. Plugging it in (2.6) the lemma is proved. □

Lemma 2.37 implies that any self-vanishable signature of degree $d \geqslant 0$ is completely determined by its first d entries. We call a self-vanishable $\mathbb{Z}_2$-signature f with arity n and $\mathrm{rd}(f) = d < n$ has the *canonical form* if $f_i = 0$ for any $0 \leqslant i \leqslant d-1$. Using Lemma 2.36, it is easy to verify that if f has the canonical form for symmetric self-vanishable $\mathbb{Z}_2$-signatures of degree d, then $\langle f, [1,1]^{\otimes r}\rangle$ has the canonical form of degree $d - r$ where $0 < r \leqslant d+1$.

Denote by $\nu^{d,n}$ the canonical symmetric self-vanishable $\mathbb{Z}_2$-signature of degree d and arity $n > d$. We also write ν^d when the arity is clear from the context. Clearly $\nu^d_d = 1$ by (2.5) and the fact that $\nu^d_i = 0$ for any $0 \leqslant i \leqslant d-1$. We will use this observation in the following lemma.

Lemma 2.38. *Let f be a symmetric self-vanishable $\mathbb{Z}_2$-signature of degree d and arity $n > d$. Then there exist $c_i \in \{0,1\}$ where $0 \leqslant i \leqslant d-1$ such that*

$$f = \nu^{d,n} + \sum_{i=0}^{d-1} c_i \nu^{i,n}.$$

Proof. Since all canonical self-vanishables have the same arity n, we will drop the superscript n. Let c_i's be the unique solution to the linear system $\sum_{j=0}^{d-1} x_j \nu^j_i = f_i$ for $0 \leqslant i \leqslant d-1$. The solution $\{c_i\}$ always exists because the coefficient matrix A is of full rank, as

$$A = \begin{bmatrix} \nu^0_0 & \nu^0_1 & \dots & \nu^0_{d-1} \\ \nu^1_0 & \nu^1_1 & \dots & \nu^1_{d-1} \\ \vdots & \vdots & & \vdots \\ \nu^{d-1}_0 & \nu^{d-1}_1 & \dots & \nu^{d-1}_{d-1} \end{bmatrix} = \begin{bmatrix} 1 & 1 & \dots & 1 \\ 0 & 1 & \dots & 0/1 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix}$$

where we use the facts that $\nu^i_j = 0$ for all $0 \leqslant j < i$ and $\nu^i_i = 1$ for all $0 \leqslant i \leqslant d-1$. Let

$$f' := \nu^d + \sum_{i=0}^{d-1} c_i \nu^i.$$

We will show that $f' = f$. By the definition of f' and $\{c_i\}$, it is easy to verify that $f'_i = f_i$ for all $0 \leqslant i \leqslant d-1$.

Then by Lemma 2.37, it is sufficient to prove that f' is self-vanishable of degree d. This can be verified as follows:

$$\langle f', [1,1]^{\otimes d+1} \rangle = \langle \nu^d, [1,1]^{\otimes d+1} \rangle + \sum_{i=0}^{d-1} c_i \langle \nu^i, [1,1]^{\otimes d+1} \rangle = \mathbf{0},$$

and

$$\langle f', [1,1]^{\otimes d}\rangle = \langle \nu^d, [1,1]^{\otimes d}\rangle + \sum_{i=0}^{d-1} c_i \langle \nu^i, [1,1]^{\otimes d}\rangle = \langle \nu^d, [1,1]^{\otimes d}\rangle \neq \mathbf{0}. \qquad \square$$

The canonical form is useful due to the following decomposition lemma.

Lemma 2.39. *The canonical symmetric self-vanishable $\mathbb{Z}_2$-signature of degree $d \geqslant 0$ can be expressed as follows:*

$$\nu^d = \sum_{S\subseteq[n],|S|=d} \bigotimes_{i=1}^{n} u_{[i\in S]},$$

where $[i \in S] = 1$ if and only if $i \in S$, $u_0 = [1,1]$ and $u_1 = [0,1]$.

Proof. We prove by induction on d. It is obvious for $d = 0$. Now we assume that the lemma holds for $d-1$. For $d > 0$, let

$$f := \sum_{S\subseteq[n],|S|=d} \bigotimes_{i=1}^{n} u_{[i\in S]}.$$

Then,

$$\begin{aligned}\langle f, [1,1]\rangle &= \left\langle \sum_{S\subseteq[n],|S|=d} \bigotimes_{i=1}^{n} u_{[i\in S]}, [1,1] \right\rangle = \sum_{S\subseteq[n],|S|=d} \left\langle \bigotimes_{i=1}^{n} u_{[i\in S]}, [1,1] \right\rangle \\ &= \sum_{S\subseteq[n],|S|=d} \left\langle u_{[n\in S]}, [1,1]\right\rangle \otimes \bigotimes_{i=1}^{n-1} u_{[i\in S]} \\ &= \sum_{\substack{S\subseteq[n],|S|=d\\ n\in S}} \bigotimes_{i=1}^{n-1} u_{[i\in S]} = \sum_{S\subseteq[n-1],|S|=d-1} \bigotimes_{i=1}^{n-1} u_{[i\in S]}.\end{aligned}$$

By the induction hypothesis, we conclude that $\langle f, [1,1]\rangle = \nu^{d-1}$. Hence f is self-vanishable of degree d by Definition 2.34. Moreover, we have that $f_i + f_{i+1} = 0$ for all $0 \leqslant i \leqslant d-2$. It implies that $f_i = f_{i+1}$ for all $0 \leqslant i \leqslant d-2$. Since $d > 0$, it is easy to verify that $f_0 = 0$. Hence $f_i = 0$ for all $0 \leqslant i \leqslant d-1$. $\square$

Remark Note that this exponential sum is very similar to the symmetrization function $\mathrm{Sym}_n^t(\cdot\,;\cdot)$

of Definition 1.12. However since we are working in $\mathbb{Z}_2$, we have to be more careful to avoid even factors introduced by redundant permutations.

We define strongly and weakly self-vanishable $\mathbb{Z}_2$-signatures depending on $\mathrm{rd}(\cdot)$. As we will see shortly, strongly self-vanishable $\mathbb{Z}_2$-signatures are vanishing.

Definition 2.40. *Let* f *be a self-vanishable* $\mathbb{Z}_2$*-signature with arity* n *and* $\mathrm{rd}(f) = d < n$*. We call* f *strongly self-vanishable if* $-1 \leqslant d \leqslant \frac{n}{2}$*, and weakly self-vanishable otherwise.*

A set $\mathcal{F}$ of strongly self-vanishable $\mathbb{Z}_2$-signatures is vanishing. The idea is that using Lemma 2.38 and Lemma 2.39, $\oplus\mathrm{Holant}(\Omega; \mathcal{F})$ can be always decomposed into smaller $\oplus$Holant's where each one contains only unary signatures, and there are more than half $[1, 1]$'s. As a result two $[1, 1]$'s must be matched, and makes the $\oplus$Holant0.

Lemma 2.41. *Let* $\mathcal{F}$ *be a set of symmetric strongly self-vanishable* $\mathbb{Z}_2$*-signatures. Then* $\mathcal{F}$ *is vanishing, that is,* $\mathcal{F} \in \mathcal{V}_\oplus$*.*

Proof. Clearly we can ignore all $\mathbb{Z}_2$-signatures in $\mathcal{F}$ that are identically 0, which are f such that $\mathrm{rd}(f) = -1$. Then by Lemma 2.29 and Lemma 2.38 it is sufficient to prove the theorem for canonical strongly self-vanishable $\mathbb{Z}_2$-signatures. They can be decomposed as in Lemma 2.39. Each term of the decomposition is a degenerate signature, a tensor product of $[1, 1]$'s and $[0, 1]$'s. For a strongly self-vanishable $\mathbb{Z}_2$-signature, we have that $d \leqslant \frac{n}{2}$. It implies that in each term, the number of $[1, 1]$'s is larger than or equal to the number of $[0, 1]$'s.

Let $\Omega = (G, \pi)$ be an instance of $\oplus\mathrm{Holant}(\mathcal{F})$. Suppose that there is at least one $f \in \mathcal{F}$ of arity n appearing in Ω such that $\mathrm{rd}(f) = d < \frac{n}{2}$. It implies that in each term of the decomposition there are strictly more $[1, 1]$'s than $[0, 1]$'s. In this case, we further decompose the Holant value as in Lemma 2.29 into a sum of several (possibly exponentially many) Holant values according to the decomposition of canonical $\mathbb{Z}_2$-signatures in Lemma 2.39. Then in every such Holant, every signature appeared is degenerate. A vertex of arity n can be viewed as n unary signatures ($[1, 1]$'s or $[0, 1]$'s). Therefore the whole graph is decomposed into isolated edges. For each edge, its two endpoints are either $[1, 1]$ or $[0, 1]$. The Holant is the product over all these edges. If both ends of one edge are $[1, 1]$, then the value for this edge is 0 and so is the Holant. However, in every Holant, such cancellation must happen at some edge because there are strictly more

$[1,1]$'s than $[0,1]$'s. Hence, in total, the whole Holant is a sum of (possibly exponentially many) 0s, which is still 0.

If there is a signature of odd arity, then $d < \frac{n}{2}$ always holds. The remaining case is that of all $\mathbb{Z}_2$-signatures have even arity and satisfy $d = \frac{n}{2}$. In that case we do the same decomposition as in the previous paragraph. The numbers of $[1,1]$s and $[0,1]$s are now exactly equal. Now it is possible to have some Holant in the decomposition that equals 1. In this case, we need to look further into the structure of the decomposition

$$f = \sum_{S \subseteq [n], |S|=d} \bigotimes_{i=1}^{n} v_{[i \in S]} = \sum_{S \subseteq [n], |S|=\frac{n}{2}} \bigotimes_{i=1}^{n} u_{[i \in S]}.$$

Let $G = (V, E)$. As in the proof of Lemma 2.29, we have that

$$\begin{aligned}
\text{Holant}_{\Omega} &= \sum_{\sigma} \prod_{v \in V} f_v(\sigma |_{E(v)}) \\
&= \sum_{\substack{S_v \subseteq [n_v],\ |S_v| = \frac{n_v}{2} \\ \forall\ v \in V}} \sum_{\sigma} \prod_{v \in V} \bigotimes_{i=1}^{n} u_{[i \in S_i]}(\sigma |_{E(v)}) \\
&= \sum_{\substack{S_v \subseteq [n_v],\ |S_v| = \frac{n_v}{2} \\ \forall\ v \in V}} \prod_{(i,j) \in E} \left\langle u_{\left[t^i_{(i,j)} \in S_i\right]}, u_{\left[t^j_{(i,j)} \in S_j\right]} \right\rangle,
\end{aligned}$$

where $t^i_{(i,j)}$ and $t^j_{(i,j)}$ are indices of the edge (i,j) in the local numbering of edges at vertices i and j, respectively. Note that the summation is indexed by a vector $\{S_v \mid v \in V\}$. A term indexed by some $\{S_v\}$ in the summation contributes 1 if and only if $\{S_v\}$ satisfies the condition that for all edges (i,j), exactly one of $t^i_{(i,j)} \in S_i$ and $t^j_{(i,j)} \in S_j$ is true. The crucial observation is that if $\{S_v\}$ satisfies this condition, its complement $\{\overline{S_v}\}$ also satisfies it. Hence, if a term indexed by $\{S_v\}$ is 1, it will be cancelled out with the term indexed by the $\{\overline{S_v}\}$. (Here we use the fact $|\overline{S_v}| = |S_v| = \frac{n_v}{2}$.) This completes the proof. □

As a final remark we note that the family $\mathcal{V}_{\oplus}$ has the following difference from $\mathcal{A}_{\oplus}$, $\mathcal{P}_{\oplus}$, and $\mathcal{M}_{\oplus}$. The union of two sets in $\mathcal{V}_{\oplus}$ is not necessarily in $\mathcal{V}_{\oplus}$. For example, $[0,0,1,1,0]$ is strongly self-vanishable, and $[1,0,1,1,1]$ is matching based vanishing, but the set of $[0,0,1,1,0]$ and $[1,0,1,1,1]$ is not vanishing.

2.5 Dichotomy for Parity Holant

In the final section of this chapter, we prove the dichotomy for $\oplus$Holant problems with symmetric $\mathbb{Z}_2$-signatures. We will reduce a lot of cases into $\oplus\text{Holant}^c$ problems, and then apply Theorem 2.27. To do so, we need to do pinning, that is, to realize $[0,1]$ and $[1,0]$. However, pinning is not always possible. We show that for exceptional cases, signature sets must be vanishing, and are therefore tractable.

We remark that some reductions in this chapter are not planar (for example, Lemma 2.42). As a result, a dichotomy for planar graphs does not follow directly.

Our first goal is to show that realizing either one of $[0,1]$ and $[1,0]$ is sufficient to realize the other. Before that we show that degenerate $\mathbb{Z}_2$-signatures $[0,0,1]$ or $[1,0,0]$ can be used as $[0,1]$ or $[1,0]$.

Lemma 2.42. *For any $\mathbb{Z}_2$-signature set $\mathcal{F}$, we have that*

$$\oplus\text{Holant}(\mathcal{F}\cup\{[1,0]\}) \equiv_T \oplus\text{Holant}(\mathcal{F}\cup\{[1,0,0]\})$$
$$\oplus\text{Holant}(\mathcal{F}\cup\{[0,1]\}) \equiv_T \oplus\text{Holant}(\mathcal{F}\cup\{[0,0,1]\}).$$

Proof. In both claims reductions from the right to the left are trivial. We will show $\oplus\text{Holant}(\mathcal{F}\cup\{[1,0]\}) \leqslant_T \oplus\text{Holant}(\mathcal{F}\cup\{[1,0,0]\})$. The other claim is similar.

Let $\Omega = (G,\pi)$ be an instance of $\oplus\text{Holant}(\mathcal{F}\cup\{[1,0]\})$. We make another copy G' of G and put them together disjointly. Then use $[1,0,0]$ to replace every occurrence of $[1,0]$ in G and its corresponding one in G'. This is depicted in Figure 2.8. All circles are put $[1,0]$ and all squares $[1,0,0]$. View the part of G excluding $[1,0]$'s as a (possibly exponentially large) signature f. The Holant of the left is $f(00\dots0)$, while the Holant of the right is $(f(00\dots0))^2 \equiv f(00\dots0) \pmod 2$. □

It may not be always possible that having only one of $[0,1]$ and $[1,0]$ we can construct the other. However, if the signature set is of one of the three tractable family, then we don't need to worry. Otherwise we show that if the other unary signature is not easy to construct, the signature set itself is hard already.

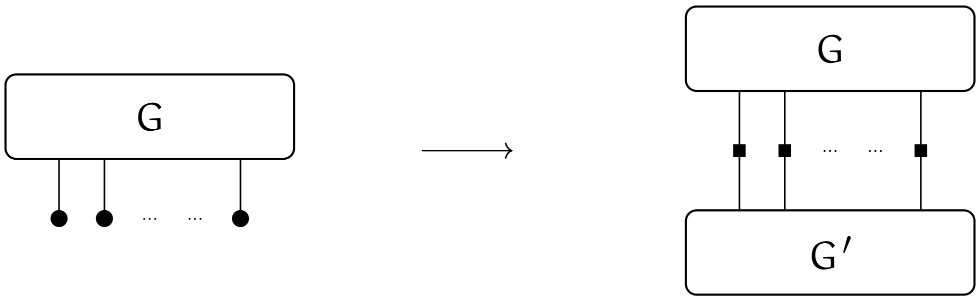


Figure 2.8: Simulating $[1,0]$ using $[1,0,0]$ or $[1,1,0]$.

Lemma 2.43. *Let $\mathcal{F}$ be a set of symmetric $\mathbb{Z}_2$-signatures. Problems* $\oplus$Holant$(\mathcal{F}\cup\{[1,0]\})$, $\oplus$Holant $(\mathcal{F}\cup\{[1,0,0]\})$, $\oplus$Holant$(\mathcal{F}\cup\{[0,1]\})$ *and* $\oplus$Holant$(\mathcal{F}\cup\{[0,0,1]\})$ *are* $\oplus$**P**-*complete unless* $\mathcal{F} \subseteq \mathcal{A}_\oplus$, $\mathcal{F} \subseteq \mathcal{P}_\oplus$, *or* $\mathcal{F} \subseteq \mathcal{M}_\oplus$. *Moreover, in any of the exceptional cases, these problems are computable in polynomial time.*

Proof. By Lemma 2.42 and symmetry, we only need to prove the lemma for $\oplus$Holant$(\mathcal{F}\cup\{[1,0]\})$. If $\mathcal{F}$ is a subset of $\mathcal{A}$, $\mathcal{M}$, or $\mathcal{F} \cup \{[0,1,0]\}$, then by Lemma 2.2, Lemma 2.4, or Lemma 2.7, $\oplus$Holant$(\mathcal{F}\cup\{[1,0]\})$ is computable in polynomial time.

Now suppose otherwise. If we can simulate $[0,1]$ or $[0,0,1]$, then by Lemma 2.42, $\oplus$Holant^c $(\mathcal{F}) \leqslant_T \oplus$Holant$(\mathcal{F}\cup\{[1,0]\})$. Hence by Theorem 2.27, $\oplus$Holant$(\mathcal{F}\cup\{[1,0]\})$ is $\oplus$**P**-complete.

Since $\mathcal{F}$ is not a subset of $\mathcal{A}_\oplus$, there exists $f \in \mathcal{F}$, which is non-degenerate and $f \notin \mathcal{A}_\oplus$. Consider the first bit of f. Assume it is 0. If the next bit is 1, we are done, since using $[1,0]$ we can get any prefix of f, and in particular, $[0,1]$ in this case. Otherwise it begins with, say, k many successive 0's followed by 1. Using $[1,0]$ we get $[0,0,\dots,1]$ of arity k. If $k=2$, then it is $[0,0,1]$ and we are done. Otherwise we do self-loops until it is unary or binary. The resulting signature is either $[0,1]$ or $[0,0,1]$. In either case, we are done with the leading bit 0 case.

Next assume that the leading bit is 1. Similar to the leading bit 0 case, we can get a $\mathbb{Z}_2$-signature g of the form $[1,1,\dots,1,0]$ of arity k. Depending on k, we have three cases.

- If $k \geqslant 3$, one self-loop on g results in a $\mathbb{Z}_2$-signature of the form $[0,0,\dots,0,1]$ of arity $k-2$. This has been dealt with in the leading bit 0 case.

- If $k=2$, then $g=[1,1,0]$. Connecting two copies of g sequentially gives us $[0,1,1]$. We get $[0,1]$ by connecting $[0,1,1]$ to $[1,0]$.

- Otherwise $k = 1$ and f begins with $1, 0$. Consider the number of successive 0's afterwards in f.
  - If there is only one 0 afterwards, f have a prefix $[1, 0, 1]$. Since f cannot be $[1, 0, 1, 0, \ldots, 0/1] \in \mathscr{A}_\oplus$, f must have a prefix of the form $[1, 0, 1, 0, \ldots, 1, 0, 0]$ or $[1, 0, 1, 0, \ldots, 0, 1, 1]$. In either case, one self-loop on this prefix gives us $[0, 0, \ldots, 0, 1]$. This has been dealt with in the leading bit 0 case.
  - Otherwise f starts with $[1, 0, \cdots, 0, 1]$ where there are at lest two 0's since f is non-degenerate. Moreover, f cannot be $[1, 0, \ldots, 0, 1] \in \mathscr{A}_\oplus$. Hence f must have a prefix h of the form either $[1, 0, \ldots, 0, 1, 0]$ or $[1, 0, \ldots, 0, 1, 1]$ of arity $k \geqslant 4$. In either case, we do a self-loop on h to get h'. Note that h' has the same form as h, but with two less arities and 0's. Repeat this process until we get a $\mathbb{Z}_2$-signature of arity 4 or 5, depending on the parity of k. It is one of the following four: $[1, 0, 0, 1, 0]$, $[1, 0, 0, 1, 1]$, $[1, 0, 0, 0, 1, 0]$ or $[1, 0, 0, 0, 1, 1]$.
    - For $[1, 0, 0, 1, 0]$, one self-loop gives $[1, 1, 0]$ which is dealt with above.
    - For $[1, 0, 0, 1, 1]$, put it on every vertex in the tetrahedron gadget of Figure 2.9. The resulting signature is $[0, 1, 0, 1, 0]$. We get $[0, 1]$ from $[0, 1, 0, 1, 0]$ by connecting it with three $[1, 0]$'s.
    - For $[1, 0, 0, 0, 1, 0]$, put it on every vertex in the gadget of Figure 2.10. The resulting signature is $[0, 0, 1]$.
    - For $[1, 0, 0, 0, 1, 1]$, with two self-loops we get $[0, 1]$. □

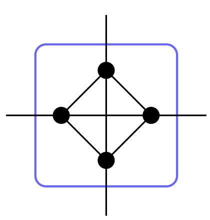


Figure 2.9: The tetrahedron gadget.

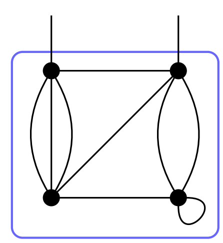


Figure 2.10: The gadget for $[1, 0, 0, 0, 1, 0]$.

Using the idea of replication in Lemma 2.42, $[1, 1, 0]$ can also be used as $[1, 0]$. Formally we have the following corollary.

Corollary 2.44. *Let $\mathcal{F}$ be a set of symmetric $\mathbb{Z}_2$-signatures. Problems* $\oplus\,\text{Holant}(\mathcal{F}\cup\{[1,1,0]\})$ *and* $\oplus\,\text{Holant}(\mathcal{F}\cup\{[0,1,1]\})$ *are* $\oplus$**P**-*complete, unless* $\mathcal{F}\subseteq\mathcal{P}_\oplus$ *or* $\mathcal{F}\subseteq\mathcal{M}_\oplus$. *In either of the exceptional cases, both problems are computable in polynomial time.*

Notice that if $\mathcal{F}$ is a subset of $\mathcal{A}_\oplus$, then $\oplus\,\text{Holant}(\mathcal{F}\cup\{[1,1,0]\})$ is still possibly $\oplus$**P**-complete, because for example $[1,1,0]$ together with $=_3\in\mathcal{A}_\oplus$ gives $\oplus$**P**-hardness by Corollary 2.17.

Proof. We first prove that

$$\oplus\,\text{Holant}\left((\mathcal{F}\cup\{[1,1,0]\})\cup[1,0]\}\right)\leqslant_T\oplus\,\text{Holant}(\mathcal{F}\cup\{[1,1,0]\}). \tag{2.7}$$

We use the same idea as in Lemma 2.42. Given an instance $\Omega=(G,\pi)$ of $\oplus\,\text{Holant}((\mathcal{F}\cup[1,1,0])\cup[1,0])$, we replicate G and replace every pair of corresponding occurrences of $[1,0]$ by $[1,1,0]$, as depicted in Figure 2.8. Circles are $[1,0]$'s and squares $[1,1,0]$'s. Call the new instance Ω'. Suppose there are n many occurrences of $[1,0]$ in G. Again view the part of G excluding $[1,0]$'s as an arity n signature f. Hence we have

$$\oplus\,\text{Holant}_{\Omega'}=\sum_{\substack{\alpha,\beta\in\{0,1\}^n\\ \alpha\wedge\beta=\mathbf{0}}}f(\alpha)f(\beta),$$

where $\alpha\wedge\beta$ is the bit-wise "and" of α and β. The requirement $\alpha\wedge\beta=\mathbf{0}$ is due to $[1,1,0]$'s. In the summation, if $\alpha\neq\beta$ and $f(\alpha)f(\beta)=1$, then $f(\beta)f(\alpha)=1$ as well, so their contributions are cancelled in $\oplus\,\text{Holant}_{\Omega'}$. If $\alpha=\beta$, then it must be that $\alpha=\beta=\mathbf{0}$. Hence,

$$\oplus\,\text{Holant}_{\Omega'}=(f(\mathbf{0}))^2\equiv f(\mathbf{0})=\oplus\,\text{Holant}_{\Omega}\,.$$

By Lemma 2.43 and (2.7), $\oplus\,\text{Holant}(\mathcal{F}\cup\{[1,1,0]\})$ is $\oplus$**P**-complete unless $(\mathcal{F}\cup\{[1,1,0]\})\subseteq\mathcal{A}_\oplus$, $(\mathcal{F}\cup\{[1,1,0]\})\subseteq\mathcal{P}_\oplus$, or $(\mathcal{F}\cup\{[1,1,0]\})\subseteq\mathcal{M}_\oplus$. Since $[1,1,0]\in\mathcal{P}_\oplus$, $[1,1,0]\in\mathcal{M}_\oplus$ whereas $[1,1,0]\notin\mathcal{A}_\oplus$, the condition above simplifies to $\mathcal{F}\subseteq\mathcal{P}_\oplus$ or $\mathcal{F}\subseteq\mathcal{M}_\oplus$. □

Finally we are ready to show the full $\oplus\,$Holant dichotomy.

Theorem 2.45. *Let $\mathcal{F}$ be a set of symmetric $\mathbb{Z}_2$-signatures. The problem $\oplus$Holant$(\mathcal{F})$ is $\oplus$**P**-complete unless $\mathcal{F} \subseteq \mathcal{A}_\oplus$, $\mathcal{F} \subseteq \mathcal{P}_\oplus$, $\mathcal{F} \subseteq \mathcal{M}_\oplus$, or $\mathcal{F} \in \mathcal{V}_\oplus$. Moreover, in any of the exceptional cases, $\oplus$Holant$(\mathcal{F})$ is polynomial time computable.*

Proof. If $\mathcal{F} \in \mathcal{V}_\oplus$ then $\oplus$Holant$(\mathcal{F})$ is trivially computable in polynomial time since we just return 0 for any input. The other three cases are tractable by Lemma 2.2, Lemma 2.4, and Lemma 2.7 (note that $\oplus\text{Holant}(\mathcal{F}) \leqslant_T \oplus\text{Holant}^c(\mathcal{F})$).

Now we assume that $\mathcal{F} \notin \mathcal{V}_\oplus$. By Definition 2.28, there is an instance $\Omega = (G, \pi)$ of $\oplus$Holant$(\mathcal{F})$ such that $\oplus\text{Holant}(\Omega; \mathcal{F}) = 1$. We shall use G as a gadget to realize $\mathbb{Z}_2$-signatures. Breaking the graph of G at one arbitrary edge, we get a signature of arity 2. Call it g. Hence $\oplus\text{Holant}(\Omega; \mathcal{F}) = g(00) + g(11) = 1$. By symmetry, we may assume that $g(00) = 1$ and $g(11) = 0$. If $g(01) = g(10) = 0$, then we have a $\mathbb{Z}_2$-signature $[1, 0, 0]$ and we are done by Lemma 2.43. If $g(01) = g(10) = 1$, then we have a $\mathbb{Z}_2$-signature $[1, 1, 0]$ and we are done by Corollary 2.44.

The remaining cases are $g(01) = 1$, $g(10) = 0$ and $g(01) = 0$, $g(10) = 1$. These two $\mathbb{Z}_2$-signatures are the same up to a reordering of the two dangling edges. We may assume that $g(01) = 1$, $g(10) = 0$. Then g is $[1, 0] \otimes [1, 1]$. By connecting two copies of this g through their first edge, we get a $\mathbb{Z}_2$-signature $[1, 1, 1] = [1, 1] \otimes [1, 1]$. By the same replication argument as in Lemma 2.42, we can use the $\mathbb{Z}_2$-signature $[1, 1]$ freely.

If all $\mathbb{Z}_2$-signatures in $\mathcal{F}$ are strongly self-vanishable, then $\mathcal{F} \in \mathcal{V}_\oplus$ by Lemma 2.41, a contradiction. Therefore there exists $f \in \mathcal{F}$ of arity n which is weakly self-vanishable or not self-vanishable.

We first assume that f is not self-vanishable. If $n = 2k + 1$ is odd, by k many self-loops on f, we get a unary $\langle f, [1, 0, 1]^{\otimes k} \rangle = \langle f, [1, 1]^{2k} \rangle$. This is not $[0, 0]$ or $[1, 1]$ since f is not self-vanishable. So it must be $[0, 1]$ or $[1, 0]$ and we are done by Lemma 2.43. If $n = 2k$ is even, then $k - 1$ many self-loops on f gives us a binary $\langle f, [1, 0, 1]^{\otimes k-1} \rangle = \langle f, [1, 1]^{\otimes 2k-2} \rangle = [a, b, c]$, where $a \neq c$ since f is not self-vanishable. It has to be one of $[1, 0, 0]$, $[0, 0, 1]$, $[1, 1, 0]$ and $[0, 1, 1]$. Again we are done by Lemma 2.43 or Corollary 2.44.

Henceforth we may assume that all $\mathbb{Z}_2$-signatures in $\mathcal{F}$ are self-vanishable. In particular $f \in \mathcal{F}$ is weakly self-vanishable such that $\text{rd}(f) = d$. Then $\frac{n}{2} < d \leqslant n - 1$ by Definition 2.40. We will realize either $[1, 0]$ or $[0, 1]$. Since d is an integer and $d > \frac{n}{2}$, $2d - n - 1 \geqslant 0$. We connect f

to $(n-d)$ many copies of g's and $2d-n-1$ many copies of $[1,1]$'s to get f'. This construction is valid because f' has arity $n-2(n-d)-(2d-n-1)=1$. Recall that $g=[1,0]\otimes[1,1]$. We calculate f' as follows:

$$f' = \langle f, g^{\otimes n-d}\otimes[1,1]^{\otimes 2d-n-1}\rangle = \langle f, [1,1]^{\otimes d-1}\otimes[1,0]^{\otimes n-d}\rangle = \langle\langle f, [1,1]^{\otimes d-1}\rangle, [1,0]^{\otimes n-d}\rangle.$$

Since f' is unary, it is in fact the first two entries of $\langle f, [1,1]^{\otimes d-1}\rangle$. By Lemma 2.36, $\langle f, [1,1]^{\otimes d-1}\rangle$ is a self-vanishable signature of degree $d-(d-1)=1$. Therefore by Lemma 2.37 it must be Parity $[1,0,1,0,\ldots,0/1]$ or $[0,1,0,1,\ldots,0/1]$, whose first two entries are either $[1,0]$ or $[0,1]$. □

Concluding Remarks

Results reported in this chapter are mainly from [GLV13], joint work with Pinyan Lu and Leslie G. Valiant. However, the presentation, especially regarding vanishing $\mathbb{Z}_2$-signatures, is largely rewritten using more up-to-date language.

The major open problem left is to characterize all vanishing $\mathbb{Z}_2$-signatures. Although we have provided three sufficient conditions including Lemma 2.41, as well as a necessary condition Proposition 2.35, it is not clear what a complete characterization of $\mathcal{V}_\oplus$ might look like. The main obstacle is that some cases, such as matching based vanishing $\mathbb{Z}_2$-signatures, can become very complicated. As we will see in Chapter 3, in contrast, we have a complete characterization of complex vanishing signatures, which basically corresponads to strongly self-vanishable signatures in this chapter.

Related work includes Faben's parity Boolean CSP dichotomy [Fab08], which was later generalized to all integer moduli [GHLX11], as well as computing the parity of graph homomorphisms [FJ15, GGR14, GGR15]. We note that in terms of graph homomorphisms, a full dichotomy is still open. Only special cases are solved such as trees and other graph families.

Chapter 3

Holant Problems on 4-regular graphs

In the following several chapters we will classify all complex weighted Boolean Holant problems on both general graphs and planar graphs. From this point on, all functions are $[2]^n \to \mathbb{C}$. In this chapter we will first develop a theory about vanishing signatures in $\mathbb{C}$, similar to the one developed in Chapter 2. However, unlike in Chapter 2, we will complete characterize all vanishing signatures in $\mathbb{C}$. These vanishing signatures happen to account for some isolated tractable cases in previous work [CLX11a, CHL12]. We will use the characterization and continue to show a dichotomy for $\operatorname{Holant}(f)$ and $\operatorname{Pl-Holant}(f)$ where f is a symmetric signature of arity 4.

3.1 Complex Vanishing Signatures

We define complex vanishing signatures similar to Definition 2.28.

Definition 3.1. *A set of signatures $\mathcal{F}$ is called* vanishing *if* $\operatorname{Holant}(\Omega; \mathcal{F}) = 0$ *for every signature grid Ω. A signature f is called* vanishing *if the singleton set $\{f\}$ is vanishing.*

The trivial example of vanishing signatures is the identically zero signature. Similar to the case in $\mathbb{Z}_2$, there are non-trivial vanishing signatures in $\mathbb{C}$. We will characterize all sets of symmetric vanishing signatures. Note that we do not have a complete characterization in Chapter 2.

First we observe that Lemma 2.29 still holds in $\mathbb{C}$, with the same proof. Note that it does not require signatures to be symmetric. Recall that $f + g$ denotes the bit-wise addition of two signatures f and g of the same arity, i.e. $(f+g)(x) = f(x) + g(x)$ for any input x.

Lemma 3.2. *Let $\mathcal{F}$ be a vanishing signature set.*

- *If f is an $\mathcal{F}$-gate, then $\mathcal{F} \cup \{f\}$ is vanishing.*
- *If g_0 and g_1 are two signatures in $\mathcal{F}$ of the same arity, then $\mathcal{F} \cup \{g_0 + g_1\}$ is vanishing.*

Our generalization is mostly based on the idea in Lemma 2.41. Basically we want to show that for certain signatures $\mathcal{F}$, Holant$(\Omega; \mathcal{F})$ can be always decomposed into smaller Holant's where each one contains only unary signatures, and two unary signatures vanishing each other must be matched. In Lemma 2.41, the vanishing unary is $[1, 1]$. In $\mathbb{C}$, apparently $[1, 1]$ is not vanishing. Instead, we will use $[1, i]$ and $[1, -i]$, which are the only two vanishing unary signatures in $\mathbb{C}$. Also note that there is no vanishing unary in $\mathbb{R}$. That explains why there is no vanishing signature shown up in the dichotomy Theorem 1.17 of real-weighted Holant .

More concretely, consider a signature set $\mathcal{F}$ where every signature of arity n is degenerate. That is, every signature of arity n is a tensor product of unary signatures. Moreover, for each signature, suppose that more than half of the unary signatures in the tensor product are $[1, i]$. For any signature grid Ω with signatures from $\mathcal{F}$, it can be decomposed into many pairs of unary signatures. The total Holant value is the product of the Holant on each pair. Since more than half of the unaries in each signature are $[1, i]$, more than half of the unaries in Ω are $[1, i]$. Then two $[1, i]$'s must be paired up and hence $\text{Holant}_\Omega = 0$. Thus, all such signatures form a vanishing set. Clearly this argument holds when $[1, i]$ is replaced by $[1, -i]$.

These signatures described above are generally asymmetric. To characterize symmetric vanishing signatures, we will use the symmetrization operation in Definition 1.12. Let t and n be two positive integers such that $t \leqslant n$. Let $v, v_1, \ldots, v_{n-t}$ be unary signatures, and S_n be the symmetric group of degree n. Recall that

$$\text{Sym}_n^t(v; v_1, \ldots, v_{n-t}) = \sum_{\pi \in S_n} \bigotimes_{k=1}^{n} u_{\pi(k)},$$

where the ordered sequence $(u_1, u_2, \ldots, u_n) = (\underbrace{v, \ldots, v}_{t \text{ copies}}, v_1, \ldots, v_{n-t})$.

Compare Definition 1.12 to the symmetrization used in Lemma 2.39. The difference is that in Definition 1.12 we allow redundant permutations of v. Moreover, equivalent v_i's also induce redundant permutations. These redundant permutations introduce a nonzero constant factor, which affects the complexity in $\mathbb{Z}_2$, but not in $\mathbb{C}$. On the other hand, they simplify our calculations in $\mathbb{C}$. We will mainly use $v = [1, i]$ or $v = [1, -i]$. An illustrative example of Definition 1.12 is

$$\begin{aligned}\mathrm{Sym}_3^2([1,i];[a,b]) &= 2[a,b]\otimes[1,i]\otimes[1,i]+2[1,i]\otimes[a,b]\otimes[1,i]+2[1,i]\otimes[1,i]\otimes[a,b]\\ &= 2[3a, 2ia+b, -a+2ib, -3b].\end{aligned}$$

Next we define the vanishing degree, which is a dual of the recurrence degree in Definition 2.34.

Definition 3.3. *A nonzero symmetric signature* f *of arity* n *has* positive vanishing degree $k \geqslant 1$, *denoted by* $\mathrm{vd}^+(f) = k$*, if* $k \leqslant n$ *is the largest positive integer such that there exists* $n-k$ *unary signatures* $v_1, \ldots, v_{n-k}$ *such that*

$$f = \mathrm{Sym}_n^k([1,i]; v_1, \ldots, v_{n-k}).$$

If f *cannot be expressed as such a symmetrization form, we define* $\mathrm{vd}^+(f) = 0$*. If* f *is the all zero signature, define* $\mathrm{vd}^+(f) = n+1$*.*

We define negative vanishing degree vd^- *similarly, using* $-i$ *instead of* i*.*

It is possible that both $\mathrm{vd}^+(f)$ and $\mathrm{vd}^-(f)$ are nonzero. For example, $\mathrm{vd}^+(=_2) = \mathrm{vd}^-(=_2) = 1$.

We define analogues to strongly self-vanishable $\mathbb{Z}_2$-signatures of Definition 2.40. Note that unlike in Definition 2.40, in $\mathbb{C}$ a signature of vanishing degree exactly $\frac{n}{2}$ does not vanish, since we do not have the nice cancellation in $\mathbb{Z}_2$.

Definition 3.4. *For* $\sigma \in \{+,-\}$*, define* $\mathcal{V}^\sigma := \{f \mid 2\,\mathrm{vd}^\sigma(f) > \mathrm{arity}(f)\}$*.*

Similar to Lemma 2.41, $\mathcal{V}^+$ and $\mathcal{V}^-$ are vanishing.

Lemma 3.5. *Let* $\mathcal{F}$ *be a set of symmetric signatures. If* $\mathcal{F} \subseteq \mathcal{V}^+$ *or* $\mathcal{F} \subseteq \mathcal{V}^-$*, then* $\mathcal{F}$ *is vanishing.*

In Theorem 3.12, we will show that $\mathcal{V}^+$ and $\mathcal{V}^-$ capture all symmetric vanishing signature sets in $\mathbb{C}$.

Characterizing Vanishing Signatures using Recurrence Relations

Similar to Lemma 2.37, we have an equivalent characterization of vanishing signatures. It also uses linear recurrence relations, but the recurrence is different from (2.5).

Definition 3.6. *A symmetric signature* $f = [f_0, f_1, \dots, f_n]$ *of arity* n *is in* $\mathcal{R}_t^+$ *for a nonnegative integer* $t \geqslant 0$ *if* $t > n$ *or for any* $0 \leqslant k \leqslant n-t$, $f_k, \dots, f_{k+t}$ *satisfy the recurrence relation*

$$\binom{t}{t} i^t f_{k+t} + \binom{t}{t-1} i^{t-1} f_{k+t-1} + \cdots + \binom{t}{0} i^0 f_k = 0. \tag{3.1}$$

We define $\mathcal{R}_t^-$ *similarly but with* $-i$ *in place of* i *in* (3.1).

It is easy to see that $\mathcal{R}_0^+ = \mathcal{R}_0^-$ is the set of all zero signatures. Also, for $\sigma \in \{+,-\}$, we have $\mathcal{R}_t^\sigma \subseteq \mathcal{R}_{t'}^\sigma$ when $t \leqslant t'$. By definition, if $\text{arity}(f) = n$ then $f \in \mathcal{R}_{n+1}^\sigma$.

Let $f = [f_0, f_1, \dots, f_n] \in \mathcal{R}_t^+$ with $0 < t \leqslant n+1$. Then the characteristic polynomial of its recurrence relation is $(1+xi)^t$. Thus there exists a polynomial $p(x)$ of degree at most $t-1$ such that $f_k = i^k p(k)$, for $0 \leqslant k \leqslant n$. Furthermore, $p(x)$ is unique. If there are two polynomials $p(x)$ and $q(x)$, both of degree at most $t-1 \leqslant n$, such that $f_k = i^k p(k) = i^k q(k)$ for $0 \leqslant k \leqslant n$, then $p(x)$ and $q(x)$ must be identical. Now suppose $f_k = i^k p(k)$ ($0 \leqslant k \leqslant n$) for some polynomial p of degree at most $t-1$, where $0 < t \leqslant n$. Then f satisfies the recurrence (3.1) of order t. Hence $f \in \mathcal{R}_t^+$.

Thus $f \in \mathcal{R}_{t+1}^+$ if and only if there exists a polynomials $p(x)$ of degree at most t such that $f_k = i^k p(k)$ ($0 \leqslant k \leqslant n$), for all $0 \leqslant t \leqslant n$. For $\mathcal{R}_{t+1}^-$, just replace i by $-i$.

Definition 3.7. *For a nonzero symmetric signature* f *of arity* n, *it is of* positive *(resp.* negative*)* recurrence degree $t \leqslant n$, *denoted by* $\text{rd}^+(f) = t$ *(resp.* $\text{rd}^-(f) = t$*), if and only if* $f \in \mathcal{R}_{t+1}^+ - \mathcal{R}_t^+$ *(resp.* $f \in \mathcal{R}_{t+1}^- - \mathcal{R}_t^-$*). If* f *is the all zero signature, we define* $\text{rd}^+(f) = \text{rd}^-(f) = -1$.

Note that although we call it the recurrence degree, it refers to a special kind of recurrence relation. For any nonzero symmetric signature f, by the uniqueness of $p(x)$, it follows that

$\mathrm{rd}^\sigma(f) = t$ if and only if $\deg(p) = t$, where $0 \leqslant t \leqslant n$. We remark that $\mathrm{rd}^\sigma(f)$ is the maximum integer t such that f does *not* belong to $\mathcal{R}_t^\sigma$. Also, for an arity n signature f, $\mathrm{rd}^\sigma(f) = n$ if and only if f does not satisfy any such recurrence relation (3.1) of order $t \leqslant n$ for $\sigma \in \{+, -\}$.

Lemma 3.8. *Let* $f = [f_0, \dots, f_n]$ *be a symmetric signature of arity* n*, not identically 0. Let* t *be a non-negative integer such that* $0 \leqslant t < n$*. For* $\sigma \in \{+, -\}$*, the following two are equivalent:*

(i) There exist t *unary signatures* $v_1, \dots, v_t$*, such that*

$$f = \mathrm{Sym}_n^{n-t}([1, \sigma i]; v_1, \dots, v_t). \tag{3.2}$$

(ii) $f \in \mathcal{R}_{t+1}^\sigma$.

Proof. We consider $\sigma = +$ since the other case is similar, so let $v = [1, i]$.

We start with $(i) \implies (ii)$ and proceed via induction on both t and n. Assume that $f = [f_0, \dots, f_n] = \mathrm{Sym}_n^{n-t}(v; v_1, \dots, v_t)$. For the first base case of $t = 0$, $\mathrm{Sym}_n^n(v) = [1, i]^{\otimes n} = [1, i, -1, -i, \dots, i^n]$, so $f_{k+1} = i f_k$ for all $0 \leqslant k \leqslant n-1$ and $f \in \mathcal{R}_1^+$.

The other base case is that $t = n-1$. Let $\mathrm{Sym}_n^1(v; v_1, \dots, v_t) = [f_0, \dots, f_n]$ where $v_i = [a_i, b_i]$ for $1 \leqslant i \leqslant t$, and $S = i^n f_n + \dots + \binom{n}{1} i f_1 + \binom{n}{0} i^0 f_0$. We need to show that $S = 0$. First notice that any entry in f is a linear combination of terms of the form $a_{i_1} a_{i_2} \cdots a_{i_{n-1-k}} b_{j_1} \cdots b_{j_k}$, where $0 \leqslant k \leqslant n-1$, and $\{i_1, \dots, i_{n-1-k}, j_1, \dots, j_k\} = \{1, 2, \dots, n-1\}$. Thus S is a linear combination of such terms as well. Now we compute the coefficient of each of these terms in S.

Each term $a_{i_1} a_{i_2} \cdots a_{i_{n-1-k}} b_{j_1} \cdots b_{j_k}$ appears twice in S, once in f_k and the other time in f_{k+1}. Its coefficient is $k!(n-k)!$ in f_k, and is $i(k+1)!(n-k-1)!$ in f_{k+1}. Thus, its coefficient in S is

$$\binom{n}{k+1} i^{k+1} i (k+1)!(n-k-1)! + \binom{n}{k} i^k k!(n-k)! = 0.$$

The above computation works for any such term due to the symmetry of f, so all their coefficients in S are 0. Hence $S = 0$.

Now assume for any $t' < t$ or for the same t and any $n' < n$, the statement holds. For (n, t), where $n > t+1$, let $g = [g_0, \dots, g_{n-1}]$ be a signature such that $g = \mathrm{Sym}_{n-1}^{n-t-1}(v; v_1, \dots, v_t)$, and for any $1 \leqslant j \leqslant t$, let $h^{(j)} = [h_0^{(j)}, \dots, h_{n-1}^{(j)}]$ be a signature such that

$$h^{(j)} = \mathrm{Sym}_{n-1}^{n-t}(v; v_1, \dots, v_{j-1}, v_{j+1}, \dots, v_t).$$

By the induction hypothesis, g satisfies (3.1) of order $t+1$, namely $g \in \mathcal{R}_{t+1}^{+}$. Also for any j, $h^{(j)}$ satisfies (3.1) of order t, namely $h^{(j)} \in \mathcal{R}_t^{+} \subseteq \mathcal{R}_{t+1}^{+}$.

We have the recurrence relation

$$\begin{aligned} \mathrm{Sym}_n^{n-t}(v; v_1, \dots, v_t) &= (n-t)v \otimes \mathrm{Sym}_{n-1}^{n-t-1}(v; v_1, \dots, v_t) \\ &\quad + \sum_{j=1}^{t} v_j \otimes \mathrm{Sym}_{n-1}^{n-t}(v; v_1, \dots, v_{j-1}, v_{j+1}, \dots, v_t). \end{aligned} \tag{3.3}$$

By (3.3), the entry of weight k in f for any $k > 0$ is

$$f_k = (n-t) i g_{k-1} + \sum_{j=1}^{t} b_j h_{k-1}^{(j)}.$$

We know that $\{g_i\}$ and $\{h_i^{(j)}\}$ satisfy the recurrence relation (3.1) of order $t+1$. Thus, their linear combination $\{f_i\}$ also satisfies (3.1) starting from $i = k > 0$ for any $k > 0$.

We also observe that by (3.3), the entry of weight k in f for any $k < n$ is

$$f_k = (n-t) g_k + \sum_{j=1}^{t} a_j h_k^{(j)}.$$

Since $t < n-1$, by the same argument, (3.1) holds for f when $k = 0$ as well.

Now we show (ii) $\implies$ (i). Notice that we only need to find unary signatures $\{v_i\}$ for $1 \leqslant i \leqslant t$ such that $\mathrm{Sym}_n^{n-t}(v; v_1, \dots, v_t)$ matches the first $t+1$ entries of f. The theorem follows from this since we have shown that $\mathrm{Sym}_n^{n-t}(v; v_1, \dots, v_t)$ satisfies (3.1) of order $t+1$ and any such signature is determined by the first $t+1$ entries.

We show that there exist $v_i = [a_i, b_i]$ ($1 \leqslant i \leqslant t$) satisfying the requirement above. Since f is not identically 0, by (3.1), some nonzero term occurs among $\{f_0, \dots, f_t\}$. Let $f_s \neq 0$, for $0 \leqslant s \leqslant t$, be the first nonzero term. By a nonzero constant multiplier, we may normalize $f_s = s!(n-s)!$, and set $v_j = [0, 1]$, for $1 \leqslant j \leqslant s$ (which is vacuous if $s = 0$), and set $v_{s+j} = [1, b_{s+j}]$, for

$1 \leqslant j \leqslant t-s$ (which is vacuous if $s=t$). Let f' be the function defined in (3.2) using these ν_i's. Then $f'_k = f_k = 0$ for $0 \leqslant k < s$ (which is vacuous if $s = 0$). By expanding the symmetrization function, for $s \leqslant k \leqslant t$, we get

$$f'_k = k!(n-k)! \sum_{j=0}^{k-s} \binom{n-t}{k-s-j} \Delta_j i^{k-s-j},$$

where Δ_j is the elementary symmetric polynomial in $\{b_{s+1}, \ldots, b_t\}$ of degree j for $0 \leqslant j \leqslant t-s$. By definition, $\Delta_0 = 1$ and $f'_s = f_s$. Setting $f'_k = f_k$ for $s+1 \leqslant k \leqslant t$, this is a linear system in Δ_j's ($1 \leqslant j \leqslant t-s$), with a triangular matrix and nonzero diagonals. We can hence solve Δ_j's for $1 \leqslant j \leqslant t-s$. It is sufficient to find b_j's ($s+1 \leqslant j \leqslant t$) to satisfy Δ_j's ($1 \leqslant j \leqslant t-s$) which we have just solved. We pick the $(t-s)$ many roots of the equation $\sum_{j=0}^{t-s} (-1)^j \Delta_j x^{t-s-j} = 0$ to be b_j's ($s+1 \leqslant j \leqslant t$). It is easy to see that such b_j's ensure that $f'_k = f_k$ for $s+1 \leqslant k \leqslant t$, and hence $f' = f$. □

Corollary 3.9. *If* f *is a symmetric signature and* $\sigma \in \{+,-\}$*, then* $\mathrm{vd}^\sigma(f) + \mathrm{rd}^\sigma(f) = \mathrm{arity}(f)$.

Thus we have an equivalent form of $\mathcal{V}^\sigma$ for $\sigma \in \{+,-\}$. Namely,

$$\mathcal{V}^\sigma = \{f \mid 2\,\mathrm{rd}^\sigma(f) < \mathrm{arity}(f)\}.$$

Characterizing Vanishing Signature Sets

Now we show that $\mathcal{V}^+$ and $\mathcal{V}^-$ capture all symmetric vanishing signature sets. To begin, we show that a vanishing signature set cannot contain both types of nontrivial vanishing signatures.

Lemma 3.10. *Let* $f_+ \in \mathcal{V}^+$ *and* $f_- \in \mathcal{V}^-$*. If neither* f_+ *nor* f_- *is the all zero signature, then the signature set* $\{f_+, f_-\}$ *is not vanishing.*

Proof. Let $\mathrm{arity}(f_+) = n$ and $\mathrm{rd}^+(f_+) = t$, so $2t < n$. Consider the gadget with two vertices and $2t$ edges between two copies of f_+. (See Figure 3.1 for an example of this gadget.) View f_+ in the symmetrized form. Since $\mathrm{vd}^+(f_+) = n-t$, in each term, there are $n-t$ many $[1,i]$'s and t many unary signatures not equal to (a multiple of) $[1,i]$. This is a superposition of many degenerate signatures. Then the only non-vanishing contributions come from the cases where the $n-2t$ dangling edges on both sides are all assigned $[1, i]$, while inside, the t copies of $[1, i]$ pair up with t unary signatures not equal to $[1, i]$ from the other side perfectly. Notice that for any such contribution, the Holant value of the inside part is always the same constant and this constant is not 0 because $[1, i]$ paired up with any unary signature other than (a multiple of) $[1, i]$ is not 0. Then the superposition of all of the permutations is a degenerate signature $[1, i]^{\otimes 2(n-2t)}$ up to a nonzero constant factor.

Similarly, we can do this for f_- of arity n' and $\mathrm{rd}^-(f_-) = t'$, where $2t' < n'$, and get a degenerate signature $[1, -i]^{\otimes 2(n'-2t')}$, up to a nonzero constant factor. Then form a bipartite signature grid with $(n' - 2t')$ vertices on one side, each assigned $[1, i]^{\otimes 2(n-2t)}$, and $(n - 2t)$ vertices on the other side, each assigned $[1, -i]^{\otimes 2(n'-2t')}$. Connect edges between the two sides arbitrarily as long as it is a 1-1 correspondence. The resulting Holant is a power of 2, which is not vanishing. □

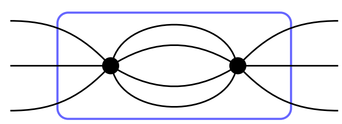


Figure 3.1: Example of a gadget used to create a degenerate vanishing signature from some general vanishing signature. This example is for a signature of arity 7 and recurrence degree 2, which is assigned to both vertices.

Lemma 3.11. *Every symmetric vanishing signature is in $\mathcal{V}^+ \cup \mathcal{V}^-$.*

Proof. Let f be a symmetric vanishing signature. We prove this by induction on n, the arity of f. For $n = 1$, by connecting $f = [f_0, f_1]$ to itself, we have $f_0^2 + f_1^2 = 0$. Then up to a constant factor, we have either $f = [1, i]$ or $f = [1, -i]$. The lemma holds.

For $n = 2$, first we do a self loop. The Holant is $f_0 + f_2$. Also, we can connect two copies of f, in which case the Holant is $f_0^2 + 2f_1^2 + f_2^2$. Since f is vanishing, we have that $f_0 + f_2 = 0$ and $f_0^2 + 2f_1^2 + f_2^2 = 0$. Solving them, we get $f = [1, i, -1] = [1, i]^{\otimes 2}$ or $f = [1, -i, -1] = [1, -i]^{\otimes 2}$ up to a constant factor.

Now assume $n > 2$ and the lemma holds for any signature of arity $k < n$. Let $f = [f_0, f_1, \ldots, f_n]$ be a vanishing signature. A self loop on f gives $f' = [f'_0, f'_1, \ldots, f'_{n-2}]$, where $f'_j = f_j + f_{j+2}$ for $0 \leqslant j \leqslant n-2$. Since f is vanishing, f' is vanishing as well. By the induction hypothesis, $f' \in \mathcal{V}^+ \cup \mathcal{V}^-$.

If f' is identically zero, then we have $f_j + f_{j+2} = 0$ for $0 \leqslant j \leqslant n-2$. This means that the f_j's satisfy a recurrence relation with characteristic polynomial x^2+1, so we have $f_j = a i^j + b(-i)^j$ for some a and b. Then we perform a holographic transformation with $Z = \frac{1}{\sqrt{2}} \left[\begin{smallmatrix} 1 & 1 \\ i & -i \end{smallmatrix}\right]$,

$$\begin{aligned} \operatorname{Holant}(=_2 \mid f) &\equiv_T \operatorname{Holant}\left([1,0,1]Z^{\otimes 2} \mid (Z^{-1})^{\otimes n} f\right) \\ &\equiv_T \operatorname{Holant}\left([0,1,0] \mid \bar{f}\right), \end{aligned}$$

where $\bar{f} = [a, 0, \ldots, 0, b]$. The problem $\operatorname{Holant}\left([0,1,0] \mid \bar{f}\right)$ is a weighted version of testing if a graph is bipartite. Now consider a graph with only two vertices, both assigned f, and n edges between them. The Holant of this graph is $2ab$. However, we know that it must be vanishing, so $ab = 0$. If $a = 0$, then $f \in \mathcal{V}^-$. Otherwise, $b = 0$ and $f \in \mathcal{V}^+$.

Now suppose that f' is in $\mathcal{V}^+ \cup \mathcal{V}^-$ but not identically zero. We consider $f' \in \mathcal{V}^+$ since the other case is similar. Then $\operatorname{rd}^+(f') = t$, so $2t < n-2$. Consider the gadget which has only two vertices, both assigned f', and has $2t$ edges between them. (It is the same one as in Lemma 3.10. See Figure 3.1 for an example.) It forms a signature of arity $d = 2(n-2-2t)$. This gadget is valid because $n-2 > 2t$. By the combinatorial view as in the proof of Lemma 3.10, this signature is $[1,i]^{\otimes d}$.

Moreover, $\operatorname{rd}^+(f') = t$ implies that the entries of f' satisfy (3.1) of order $t+1$. Replacing f'_j by $f_j + f_{j+2}$, we get a recurrence relation for the entries of f with characteristic polynomial $(x^2+1)(x-i)^{t+1} = (x+i)(x-i)^{t+2}$. Thus, $f_j = i^j p(j) + c(-i)^j$ for some polynomial $p(x)$ of degree at most $t+1$ and some constant c. It suffices to show that $c = 0$ since $2(t+1) < n$ as $2t < n-2$.

Consider the signature $h = [h_0, \ldots, h_{n-1}]$ created by connecting f with a single unary signature $[1,i]$. For any $(n-1)$–regular graph $G = (V, E)$ with h assigned to every vertex, we can define a duplicate graph of $(d+1)|V|$ vertices as follows. First for each $v \in V$, define vertices v', $v_1, \ldots, v_d$. For each i, $1 \leqslant i \leqslant d$, we make a copy of G on $\{v_i \mid v \in V\}$, i.e., for each edge $(u,v) \in E$, include the edge (u_i, v_i) in the new graph. Next for each $v \in V$, we introduce edges between v' and v_i for all $1 \leqslant i \leqslant d$. For each $v \in V$, assign the degenerate signature $[1,i]^{\otimes d}$ that we have constructed in the last paragraph to the vertices v'; assign f to all the vertices $v_1, \ldots, v_d$. Let H be the Holant of G with h assigned to every vertex. Then for the new graph with the given signature assignments, the Holant is H^d. By our assumption, f is vanishing, so $H^d = 0$. Thus, $H = 0$. This holds for any graph G, so h is vanishing.

Notice that $h_k = f_k + if_{k+1}$ for any $0 \leqslant k \leqslant n-1$. If h is identically zero, then $f_k + if_{k+1} = 0$ for any $0 \leqslant k \leqslant n-1$, which means $f = [1, i]^{\otimes n}$ up to a constant factor and we are done. Otherwise, suppose that h is not identically zero. By the inductive hypothesis, $h \in \mathcal{V}^+ \cup \mathcal{V}^-$. We claim h cannot be from $\mathcal{V}^-$. This is because, although we do not directly construct h from f, we can always realize it by the method depicted in the previous paragraph. Therefore the set $\{f', h\}$ is vanishing. As both f' and h are nonzero, and $f' \in \mathcal{V}^+$, we have $h \notin \mathcal{V}^-$, by Lemma 3.10.

Hence $h \in \mathcal{V}^+$. There exists a polynomial $q(x)$ of degree at most $t' = \lfloor \frac{n-1}{2} \rfloor$ such that $h_k = i^k q(k)$, for any $0 \leqslant k \leqslant n-1$. Since $2t < n-2$, we have $t \leqslant t'$. On the other hand, $h_k = f_k + if_{k+1}$ for any $0 \leqslant k \leqslant n-1$, so we have

$$\begin{aligned}
i^k q(k) = h_k &= f_k + if_{k+1} \\
&= i^k p(k) + c(-i)^k + i\left(i^{k+1} p(k+1) + c(-i)^{k+1}\right) \\
&= i^k \left(p(k) - p(k+1)\right) + 2c(-i)^k \\
&= i^k r(k) + 2c(-i)^k,
\end{aligned}$$

where $r(x) = p(x) - p(x+1)$ is another polynomial of degree at most t. Then we have

$$q(k) - r(k) = 2c(-1)^k,$$

which holds for all $0 \leqslant k \leqslant n-1$. Notice that the left hand side is a polynomial of degree at most t', call it $s(x)$. However, for all even $k \in [n]$, $s(k) = 2c$. There are exactly $\lceil \frac{n}{2} \rceil > \lfloor \frac{n-1}{2} \rfloor = t'$ many even k within the range $\{0, \dots, n-1\}$. Thus $s(x) = 2c$ for any x. Now we pick $k = 1$, so $s(1) = -2c = 2c$, which implies $c = 0$. This completes the proof. □

Combining Lemma 3.5, Lemma 3.10, and Lemma 3.11, we obtain the following theorem that characterizes all symmetric vanishing signature sets.

Theorem 3.12. *Let $\mathcal{F}$ be a set of symmetric signatures. Then $\mathcal{F}$ is vanishing if and only if $\mathcal{F} \subseteq \mathcal{V}^+$ or $\mathcal{F} \subseteq \mathcal{V}^-$.*

The set of vanishing signatures is closed under orthogonal transformations. This is because under any orthogonal transformation, the unary signatures $[1, i]$ and $[1, -i]$ are either invariant or transformed into each other. Then considering the symmetrized form of any signature, we have the following lemma.

Lemma 3.13. *For a symmetric signature* f *of arity* n, $\sigma \in \{+, -\}$, *and an orthogonal matrix* $T \in \mathbb{C}^{2\times 2}$, *either* $\operatorname{vd}^{\sigma}(f) = \operatorname{vd}^{\sigma}(T^{\otimes n} f)$ *or* $\operatorname{vd}^{\sigma}(f) = \operatorname{vd}^{-\sigma}(T^{\otimes n} f)$.

Characterizing Vanishing Signatures via a Holographic Transformation

There is another explanation for the vanishing signatures. Given $f \in \mathcal{V}^{+}$ with $\operatorname{arity}(f) = n$ and $\operatorname{rd}^{+}(f) = d$, we perform a holographic transformation with $Z = \frac{1}{\sqrt{2}} \left[\begin{smallmatrix} 1 & 1 \\ i & -i \end{smallmatrix}\right]$,

$$\begin{aligned} \operatorname{Holant}(=_2 \mid f) &\equiv_T \operatorname{Holant}\left([1,0,1] Z^{\otimes 2} \mid (Z^{-1})^{\otimes n} f\right) \\ &\equiv_T \operatorname{Holant}\left([0,1,0] \mid \bar{f}\right), \end{aligned}$$

where $\bar{f}$ is of the form $[\bar{f}_0, \bar{f}_1, \ldots, \bar{f}_d, 0, \ldots, 0]$, and $\bar{f}_d \neq 0$. To see this, note that $Z^{-1} = \frac{1}{\sqrt{2}} \left[\begin{smallmatrix} 1 & -i \\ 1 & i \end{smallmatrix}\right]$ and $Z^{-1} \left[\begin{smallmatrix} 1 \\ i \end{smallmatrix}\right] = \sqrt{2} \left[\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right]$. We know that f has a symmetrized form, such as $\operatorname{Sym}_n^{n-d}(\left[\begin{smallmatrix} 1 \\ i \end{smallmatrix}\right]; v_1, \ldots, v_d)$. Then up to a factor of $2^{n/2}$, we have $\bar{f} = (Z^{-1})^{\otimes n} f = \operatorname{Sym}_n^{n-d}(\left[\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right]; u_1, \ldots, u_d)$, where $u_i = Z^{-1} v_i$ for $1 \leqslant i \leqslant d$ and u_i and v_i are column vectors in $\mathbb{C}^2$. From this expression for $\bar{f}$, it is clear that all entries of Hamming weight greater than d in $\bar{f}$ are 0. Moreover, if $\bar{f}_d = 0$, then one of the u_i has to be a multiple of $[1, 0]$. This contradicts the degree assumption of f, namely $\operatorname{vd}^{+}(f) = n - \operatorname{rd}^{+}(f) = n - d$ but not any higher. Formally we have the following.

Lemma 3.14. *Suppose* f *is a symmetric signature of arity* n. *Let* $\bar{f} = (Z^{-1})^{\otimes n} f$. *If* $\operatorname{rd}^{+}(f) = d$, *then* $\bar{f} = [\bar{f}_0, \bar{f}_1, \ldots, \bar{f}_d, 0, \ldots, 0]$ *and* $\bar{f}_d \neq 0$. *Also* $f \in \mathcal{R}_d^{+}$ *if and only if all nonzero entries of* $\bar{f}$ *are among the first* d *entries in its symmetric signature notation.*

Similarly, if $\operatorname{rd}^{-}(f) = d$, *then* $\bar{f} = [0, \ldots, 0, \bar{f}_{n-d}, \ldots, \bar{f}_n]$ *and* $\bar{f}_{n-d} \neq 0$. *Also* $f \in \mathcal{R}_d^{-}$ *if and only if all nonzero entries of* $\bar{f}$ *are among the last* d *entries in its symmetric signature notation.*

By linearity, Lemma 3.14 implies the following fact. If $f = g + h$ is of arity n, where $\operatorname{rd}^{+}(g) = d$, $\operatorname{rd}^{-}(h) = d'$, and $d + d' < n$, then after a holographic transformation by Z, $\bar{f} = (Z^{-1})^{\otimes n} f$

takes the form $[\overline{g}_0, \ldots, \overline{g}_d, 0, \ldots, 0, \overline{h}_{d'}, \ldots, \overline{h}_0]$, with $n - d - d' - 1 \geqslant 0$ zeros in the middle of the signature.

In any instance of $\operatorname{Holant}([0,1,0] \mid \overline{f})$, the binary DISEQUALITY $(\neq_2) = [0,1,0]$ on the left imposes the condition that half of the edges must take the value 0 and the other half must take the value 1. On the right side, by $f \in \mathcal{V}^+$, we have $d < n/2$, thus $\overline{f}$ requires that less than half of the edges are assigned the value 1. Therefore the Holant is always 0. A similar conclusion was reached in [CLX12] for certain 2-3 bipartite Holant problems with Boolean signatures. However, the importance was not realized at that time.

Under this transformation, one can observe another interesting phenomenon. For any $a, b \in \mathbb{C}$,

$$\operatorname{Holant}([0,1,0] \mid [a,b,1,0,0]) \qquad \text{and} \qquad \operatorname{Holant}([0,1,0] \mid [0,0,1,0,0])$$

take exactly the same value on every signature grid. This is because, to contribute a nonzero term in the Holant, exactly half of the edges must be assigned 1. Then for the first problem, the signature on the right can never contribute a nonzero value involving a or b. Thus the Holant values of these two problems on any signature grid are always the same. Nevertheless, there exist $a, b \in \mathbb{C}$ such that there is no holographic transformation between these two problems. We note that this is the first counterexample involving non-unary signatures in the Boolean domain to the converse of Theorem 1.1, Valiant's Holant Theorem. It provides a negative answer to a conjecture made by Xia in [Xia11, Conjecture 4.1].

Moreover, $\operatorname{Holant}([0,1,0] \mid [0,0,1,0,0])$ counts Eulerian orientations in a 4-regular graph. This problem was shown #**P**-hard [HL12, Theorem V.10]. In this chapter we will strengthen this result to the planar setting. Undoing the Z transformation, the problem of counting Eulerian orientations in a 4-regular graph is $\operatorname{Holant}([3,0,1,0,3])$. The other problem $\operatorname{Holant}([0,1,0] \mid [a,b,1,0,0])$ corresponds to a Holant problem defined by $f = Z^{\otimes 4}[a,b,1,0,0]$ of arity 4 with $\operatorname{rd}(f) = 2$. Therefore, for any $a, b \in \mathbb{C}$, f is #**P**-hard as well.

Tractable cases involving vanishing signatures

We note that some particular categories of tractable cases in previous dichotomies (case 2 of Theorem 1.14, case 3 of Theorem 1.18, and case 4 of Theorem 1.19) are in $\mathcal{R}_2^{\pm}$. At the time they were discovered, it was not known that those cases are vanishing. In fact, we may add more signatures in vanishing sets and preserve its tractability. First, we can include binary signatures in $\mathcal{R}_2^\sigma$ for $\sigma \in \{+,-\}$.

Lemma 3.15. *Let $\mathcal{F}$ be a set of complex weighted symmetric signatures in Boolean variables. Then* $\operatorname{Holant}(\mathcal{F})$ *is computable in polynomial time if* $\mathcal{F} \subseteq \mathcal{V}^\sigma \cup \{f \in \mathcal{R}_2^\sigma \mid \operatorname{arity}(f) = 2\}$ *for* $\sigma \in \{+,-\}$*.*

Proof. Any binary signature $g \in \mathcal{R}_2^\sigma$ has $\operatorname{rd}^\sigma(g) \leqslant 1$, and thus $\operatorname{vd}^\sigma(g) \geqslant 1 = \operatorname{arity}(g)/2$. Any signature $f \in \mathcal{V}^\sigma$ has $\operatorname{vd}^\sigma(f) > \operatorname{arity}(f)/2$. If $\mathcal{F}$ contains a signature f of arity at least 3, then it must belong to $\mathcal{V}^\sigma$. Given an instance Ω of $\operatorname{Holant}(\mathcal{F})$, if there is f appearing in Ω such that $\operatorname{arity}(f) \geqslant 3$, then $f \in \mathcal{V}^\sigma$ and by the combinatorial view of Lemma 3.5, more than half of the unary signatures are $[1, \sigma i]$, so $\operatorname{Holant}_\Omega$ vanishes. On the other hand, if none of arity 3 signatures shown up in Ω, then every signature is of arity at most 2. This is the case of Lemma 1.5. □

Moreover, we may combine all unary and degenerate signatures with $\mathcal{R}_2^\sigma$ for $\sigma \in \{+,-\}$.

Lemma 3.16. *Let $\mathcal{F}$ be a set of complex weighted symmetric signatures in Boolean variables. Then* $\operatorname{Holant}(\mathcal{F})$ *is computable in polynomial time if all non-degenerate signatures in $\mathcal{F}$ are in $\mathcal{R}_2^\sigma$ for $\sigma \in \{+,-\}$.*

Note that any signature in $\mathcal{R}_2^\sigma$ having arity at least 3 is a vanishing signature. Thus all signatures of arity at least 3 in Lemma 3.16 are vanishing. Lemma 3.16 is in fact case 3 of Theorem 1.18 [CLX11a]. It is also a special case of Fibonacci gates [CLX13, CLX11b]. Here we give a different proof based on vanishing signatures.

Proof. Let Ω be an instance of $\operatorname{Holant}(\mathcal{F})$. Decompose all degenerate signatures into unary ones. Then recursively absorb any unary signature into its neighboring signature. If it is connected to another unary signature, then this produces a global constant factor. If it is connected to a binary signature, then this creates another unary signature. We observe that if $f \in \mathcal{R}_2^\sigma$ has $\operatorname{arity}(f) \geqslant 2$, then for any unary signature u, after connecting f to u, the signature $\langle f, u\rangle$ still belongs to $\mathcal{R}_2^\sigma$. Hence after recursively absorbing all unary signatures in the above process, we still have a signature grid where all signatures belong to $\mathcal{R}_2^\sigma$. Any remaining signature f that has arity at least 3 belongs to $\mathcal{V}^\sigma$ since $\text{rd}^\sigma(f) \leqslant 1$ and thus $\text{vd}^\sigma(f) \geqslant \text{arity}(f) - 1 > \text{arity}(f)/2$. Thus we can apply Lemma 3.15. □

Lemma 3.15 and Lemma 3.16 are both based on vanishing signatures, but they can be very different. In Lemma 3.15, all signatures in $\mathcal{F}$, including unary signatures but excluding binary signatures, must be in $\mathcal{V}^\sigma$ for $\sigma \in \{+, -\}$; the binary signatures are only required to be in $\mathcal{R}_2^\sigma$. In contrast, Lemma 3.16 has no requirement placed on degenerate signatures which include all unary signatures. All non-degenerate binary signatures are required to be in $\mathcal{R}_2^\sigma$. Moreover, all non-degenerate signatures of arity at least 3 are also required to be in $\mathcal{R}_2^\sigma$, which is a strong form of vanishing; they must have a large vanishing degree of type σ.

3.2 Redundant 4-by-4 Matrices

We need some language to talk about arity 4 signatures, which will be the main subject in the rest of this chapter. For a binary signature, we have a matrix representation defined in (1.2). We extend this notion to arity 4 signatures. We define a 4-by-4 matrix as the signature matrix of an arity 4 signature, index by two "input" bits and two "output" bits. Note that there is no real input and output. It is just that we view two wires as inputs and the other two as outputs.

Definition 3.17. *The* signature matrix *of a symmetric arity 4 signature* $f = [f_0, f_1, f_2, f_3, f_4]$ *is*

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

This definition extends to an asymmetric signature g *as*

$$M_g = \begin{bmatrix} g^{0000} & g^{0010} & g^{0001} & g^{0011} \\ g^{0100} & g^{0110} & g^{0101} & g^{0111} \\ g^{1000} & g^{1010} & g^{1001} & g^{1011} \\ g^{1100} & g^{1110} & g^{1101} & g^{1111} \end{bmatrix},$$

where we use $g^{\mathbf{x}}$ *to denote* $g(\mathbf{x})$ *for a vector* $\mathbf{x} \in \{0,1\}^4$.

When we present g as an $\mathcal{F}$-gate, we order the four external edges ABCD counterclockwise. In M_g, the row index bits are ordered AB and the column index bits are ordered DC, in reverse order. This is for convenience so that the signature matrix of the linking of two arity 4 $\mathcal{F}$-gates is the matrix product of the signature matrices of the two $\mathcal{F}$-gates.

One important property of the signature matrix of a symmetric signature is that its middle two rows and two columns are identical. If we connect two such signatures together by two wires, the resulting one is not necessarily symmetric, but it still has this property. Due to this observation, we define redundant matrices.

Definition 3.18. *A 4-by-4 matrix is* redundant *if its middle two rows and middle two columns are the same. Denote the set of all redundant 4-by-4 matrices over a field* $\mathbb{F}$ *by* $\mathrm{RM}_4(\mathbb{F})$.

Consider the function $\varphi : \mathbb{C}^{4\times 4} \to \mathbb{C}^{3\times 3}$ defined by

$$\varphi(M) = AMB,$$

where

$$A = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Intuitively, the operation φ replaces the middle two columns of M with their sum and then the middle two rows of M with their average. (These two steps commute.) Conversely, we have the following function $\psi : \mathbb{C}^{3\times 3} \to \mathrm{RM}_4(\mathbb{C})$ defined by

$$\psi(N) = BNA.$$

Intuitively, the operation ψ duplicates the middle row of N and then splits the middle column evenly into two columns. Notice that $\varphi(\psi(N)) = N$. When restricted to $\mathrm{RM}_4(\mathbb{C})$, φ is an isomorphism between the semi-group of 4-by-4 redundant matrices and the semi-group of 3-by-3 matrices, under matrix multiplication, and ψ is its inverse. To see this, just notice that

$$AB = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \text{and} \quad BA = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

are the identity elements of their respective semi-groups.

For an (not necessarily symmetric) signature f, if M_f is redundant, define the *compressed signature matrix* of f as $\widetilde{M_f} := \varphi(M_f)$.

We are particularly interested in the signature id_4 with signature matrix

$$M_{\mathrm{id}_4} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \tag{3.4}$$

the identity element in the semi-group of redundant matrices. Thus, $\widetilde{M_{\mathrm{id}_4}} = I_3$.

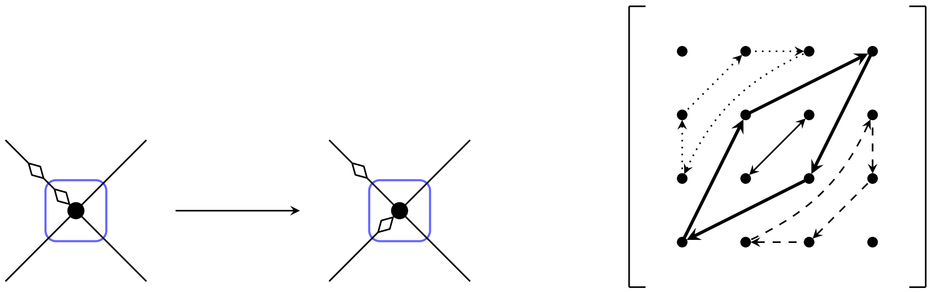
Although our main focus is symmetric signatures, to achieve our result we have to go beyond to asymmetric signatures. For an asymmetric signature (of a fixed ordering), we use a diamond to illustrate its most significant bit. We often want to reorder the input bits under a circular permutation. For a single counterclockwise rotation by 90°, the effect on the entries of the signature matrix of an arity 4 signature is given in Figure 3.2.

Redundancy of a signature matrix and non-singularity of a compressed signature matrix are invariant under invertible holographic transformations.

Lemma 3.19. *Let* f *be an arity 4 signature with complex weights,* $T \in \mathbb{C}^{2\times 2}$ *a matrix, and* $\hat{f} = T^{\otimes 4} f$*. If* M_f *is redundant, then* $M_{\hat{f}}$ *is also redundant and* $\det(\varphi(M_{\hat{f}})) = \det(\varphi(M_f)) \det(T)^6$*.*

Proof. Since $\hat{f} = T^{\otimes 4} f$, we can express $M_{\hat{f}}$ in terms of M_f and T as

$$M_{\hat{f}} = T^{\otimes 2} M_f \left(T^{\mathsf{T}}\right)^{\otimes 2}. \tag{3.5}$$



(a) A counterclockwise rotation (b) Movement of signature matrix entries

Figure 3.2: The movement of the entries in the signature matrix of an arity 4 signature under a counterclockwise rotation of the input edges. Entires of Hamming weight 1 are in the dotted cycle, entires of Hamming weight 2 are in the two solid cycles (one has length 4 and the other one is a swap), and entries of Hamming weight 3 are in the dashed cycle.

This can be directly checked. Alternatively, this relation is known (and can also be directly checked) had we not introduced the flip of the middle two columns, i.e., if the columns were ordered $00, 01, 10, 11$ by the last two bits in f and $\overline{f}$. Instead, the columns are ordered by $00, 10, 01, 11$ in M_f and $M_{\overline{f}}$. Let $T = (t^i_j)$, where row index i and column index j range from $\{0, 1\}$. Then $T^{\otimes 2} = (t^i_j t^{i'}_{j'})$, with row index ii' and column index jj'. Let

$$\mathcal{E} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

then $\mathcal{E} T^{\otimes 2} \mathcal{E} = T^{\otimes 2}$, i.e., a simultaneous row flip $ii' \leftrightarrow i'i$ and column flip $jj' \leftrightarrow j'j$ keep $T^{\otimes 2}$ unchanged. Then the known relations $M_{\overline{f}} \mathcal{E} = T^{\otimes 2} M_f \mathcal{E} \left(T^{\mathsf{T}}\right)^{\otimes 2}$ and $\mathcal{E} \left(T^{\mathsf{T}}\right)^{\otimes 2} \mathcal{E} = \left(T^{\mathsf{T}}\right)^{\otimes 2}$ imply (3.5).

Now $X \in \mathrm{RM}_4(\mathbb{C})$ if and only if $\mathcal{E} X = X = X \mathcal{E}$. Then it follows that $M_{\overline{f}} \in \mathrm{RM}_4(\mathbb{C})$ if $M_f \in \mathrm{RM}_4(\mathbb{C})$. For the two matrices A and B in the definition of φ, we note that $BA = M_{\mathrm{id}_4}$, where M_{id_4} given in (3.4) is the identity element of the semi-group $\mathrm{RM}_4(\mathbb{C})$. Since $M_f \in \mathrm{RM}_4(\mathbb{C})$, we have $BAM_f = M_f = M_f BA$. Then we have

$$\begin{aligned} \varphi(M_{\bar{f}}) &= AM_{\bar{f}}B = A\left(T^{\otimes 2}M_f\,(T^{\intercal})^{\otimes 2}\right)B \\ &= (AT^{\otimes 2}B)(AM_fB)(A\,(T^{\intercal})^{\otimes 2}\,B) \\ &= \varphi(T^{\otimes 2})\varphi(M_f)\varphi((T^{\intercal})^{\otimes 2}). \end{aligned} \tag{3.6}$$

Another direct calculation shows that

$$\det(\varphi(T^{\otimes 2})) = \det(T)^3 = \det(\varphi((T^{\intercal})^{\otimes 2})).$$

Thus, by applying determinant to both sides of (3.6), we have

$$\det(\varphi(M_{\bar{f}})) = \det(\varphi(M_f))\det(T)^6$$

as claimed. □

3.3 Counting Eulerian Orientations in 4-Regular Planar Graphs

Counting (unweighted) Eulerian orientations over 4-regular graphs was shown to be #**P**-hard [HL12, Theorem V.10]. We will strengthen this result by showing that this problem remains #**P**-hard when the graph is also planar. Recall the definition of an Eulerian orientation.

Definition 3.20. *Given a graph* G*, an orientation of its edges 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*.*

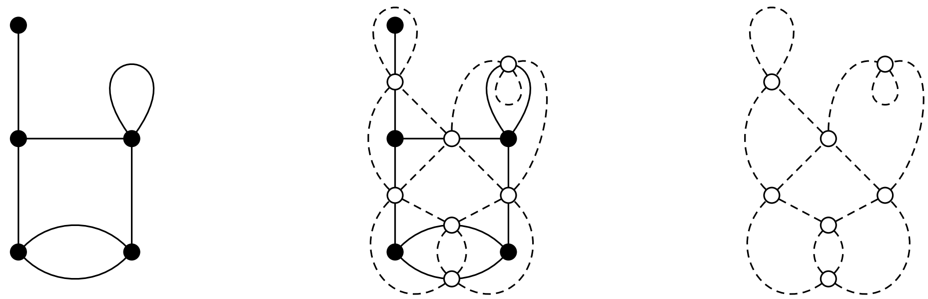
We have the following problem and result.

Name PL-4REG-#EO

Instance A 4-regular-planar graphs G.

Output The number of Eulerian orientations in G.

Theorem 3.21. PL-4REG-#EO *is #**P**-hard.*



(a) (b) (c)

Figure 3.3: A plane graph (a), its medial graph (c), and both graphs superimposed (b).

The reduction begins with the problem of evaluating the Tutte polynomial at the point (3,3), which is #**P**-hard even over planar graphs.

Theorem 3.22 (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* $(x-1)(y-1) \in \{1, 2\}$ *or* $(x, y) \in \{(1, 1), (-1, -1), (\omega, \omega^2), (\omega^2, \omega)\}$*, where* $\omega = e^{2\pi i/3}$*. In each exceptional case, the computation can be done in polynomial time.*

The first step in the reduction concerns a sum of weighted Eulerian orientations on a medial graph of a planar graph. Recall the definition of a medial graph.

Definition 3.23 (cf. [BO92])**.** *For a connected plane graph* G *(i.e. a planar embedding of a connected planar graph), its* medial graph H *has a vertex for each edge of* G *and two vertices in* H *are joined by an edge for each face of* G *in which their corresponding edges occur consecutively.*

An example of a plane graph and its medial graph are given in Figure 3.3. Notice that a medial graph of a planar graph is always a planar 4-regular graph. Las Vergnas 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 [Las88].

Theorem 3.24 (Theorem 2.1 in [Las88])**.** *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*

$$2 \cdot T(G; 3, 3) = \sum_{O \in \mathcal{O}(G_m)} 2^{\beta(O)}, \tag{3.7}$$

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 right side of (3.7) is invariant under different embeddings of G. This follows from (3.7) and the fact that the Tutte polynomial does not depend on the embedding of G.

Now we are ready to prove Theorem 3.21.

Proof of Theorem 3.21. In the Holant language, PL-4REG-#EO is the problem Pl-Holant($\neq_2$ | $[0,0,1,0,0]$). We reduce calculating the right side of (3.7) to Pl-Holant $(\neq_2 \mid [0,0,1,0,0])$. Once finished the reduction the theorem follows from Theorem 3.22 and Theorem 3.24.

The right side of (3.7) can be expressed as Pl-Holant $(\neq_2 \mid f)$, where the signature matrix of f is

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

We perform a holographic transformation by $Z = \left[\begin{smallmatrix} 1 & 1 \\ i & -i \end{smallmatrix}\right]$ to get

$$\begin{aligned} \text{Pl-Holant}\,(\neq_2 \mid f) &\equiv_T \text{Pl-Holant}\left([0,1,0](Z^{-1})^{\otimes 2} \mid Z^{\otimes 4} f\right) \\ &\equiv_T \text{Pl-Holant}\left([1,0,1]/2 \mid 4\bar{f}\right) \\ &\equiv_T \text{Pl-Holant}(\bar{f}), \end{aligned}$$

where the signature matrix of $\bar{f}$ is

$$M_{\bar{f}} = \begin{bmatrix} 2 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 2 \end{bmatrix}.$$

We also transform Pl-Holant $(\neq_2 \mid [0,0,1,0,0])$ by Z to get

$$\begin{aligned}\text{Pl-Holant}\,(\neq_2 \mid [0,0,1,0,0]) &\equiv_T \text{Pl-Holant}\left([0,1,0](Z^{-1})^{\otimes 2} \mid Z^{\otimes 4}[0,0,1,0,0]\right)\\ &\equiv_T \text{Pl-Holant}\,([1,0,1]/2 \mid 2[3,0,1,0,3])\\ &\equiv_T \text{Pl-Holant}([3,0,1,0,3]).\end{aligned}$$

Using the planar tetrahedron gadget in Figure 3.4, we assign $[3,0,1,0,3]$ to every vertex and obtain a gadget with signature $32\overline{g}$, where the signature matrix of $\overline{g}$ is

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

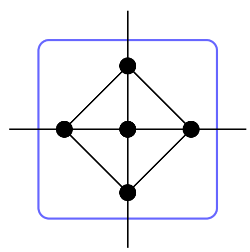


Figure 3.4: The planar tetrahedron gadget. Each vertex is assigned $[3,0,1,0,3]$.

Now we show how to reduce Pl-Holant$(\overline{f})$ to Pl-Holant$(\overline{g})$ by interpolation. Consider an instance Ω of Pl-Holant$(\overline{f})$. Suppose that $\overline{f}$ appears n times in Ω. We construct from Ω a sequence of instances Ω_s of Holant$(\overline{g})$ indexed by $s \geqslant 1$. We obtain Ω_s from Ω by replacing each occurrence of $\overline{f}$ with the gadget N_s in Figure 3.5 with $\overline{g}$ assigned to all vertices. Although $\overline{f}$ and $\overline{g}$ are asymmetric signatures, they are invariant under a cyclic permutation of their inputs. Thus, it is not necessary to specify which edge corresponds to which input. We call such signatures *rotationally symmetric*. In other words, a rotationally symmetric signature has the same matrix under the operation in Figure 3.2.

To obtain Ω_s from Ω, we effectively replace $M_{\overline{f}}$ with $M_{N_s} = (M_{\overline{g}})^s$, the sth power of the

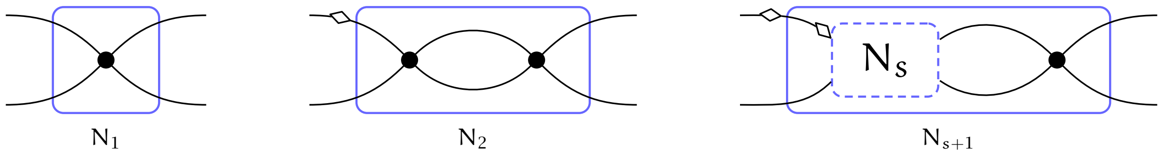


Figure 3.5: Recursive construction to interpolate $\overline{f}$. The vertices are assigned $\overline{g}$.

signature matrix $M_{\overline{g}}$. Let

$$T = \begin{bmatrix} 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ 0 & 0 & -1 & 1 \end{bmatrix}.$$

Then

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

We can view our construction of Ω_s as first replacing each $M_{\overline{f}}$ by $T\Lambda_{\overline{f}}T^{-1}$ to obtain a signature grid Ω', which does not change the Holant value, and then replacing each $\Lambda_{\overline{f}}$ with $\Lambda_{\overline{g}}^{s}$. We stratify the assignments in Ω' based on the assignment to $\Lambda_{\overline{f}}$. We only need to consider the assignments to $\Lambda_{\overline{f}}$ that assign

- 0000 j many times,
- 0110 or 1001 k many times, and
- 1111 ℓ many times.

Let $c_{jk\ell}$ be the sum over all such assignments of the products of evaluations from T and T^{-1} but excluding $\Lambda_{\overline{f}}$ on Ω'. Then

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

and the value of the Holant on Ω_s, for $s \geqslant 1$, is

$$\text{Holant}_{\Omega_s} = \sum_{j+k+\ell=n} (6^k 13^\ell)^s c_{jk\ell}. \tag{3.8}$$

This coefficient matrix in the linear system of (3.8) is Vandermonde and of full rank since for any $0 \leqslant k+\ell \leqslant n$ and $0 \leqslant k'+\ell' \leqslant n$ such that $(k,\ell) \neq (k',\ell')$, $6^k 13^\ell \neq 6^{k'} 13^{\ell'}$. Therefore, we can solve the linear system for the unknown $c_{jk\ell}$'s and obtain the value of Holant_Ω. □

3.4 Redundant Signatures with Non-Singular Compressed Matrices

In this section, we will show that all redundant signatures with non-singular compressed matrices are #**P**-hard, even in planar graphs. We begin with the identity element id_4 of $\mathrm{RM}_4(\mathbb{C})$.

Lemma 3.25. *Let* id_4 *be the arity 4 signature with* M_{id_4} *given in (3.4) so that* $\widetilde{M_{\mathrm{id}_4}} = I_3$. *Then* Pl-Holant$(\mathrm{id}_4)$ *is #**P**-hard.*

Proof. We reduce from Pl-Holant(f), where $f = [1, 0, \frac{1}{3}, 0, 1]$. Under the holographic transformation $Z = \frac{1}{\sqrt{2}} \left[\begin{smallmatrix} 1 & 1 \\ i & -i \end{smallmatrix}\right]$, Pl-Holant$(f)$ is equivalent to Pl-4reg-#EO. Hence Pl-Holant(f) is #**P**-hard by Theorem 3.21. The reduction is via an arbitrarily close approximation using the recursive construction in Figure 3.6 with g assigned to every vertex.

We claim that the signature matrix M_{N_k} of Gadget N_k is

$$M_{N_k} = \begin{bmatrix} 1 & 0 & 0 & a_k \\ 0 & a_{k+1} & a_{k+1} & 0 \\ 0 & a_{k+1} & a_{k+1} & 0 \\ a_k & 0 & 0 & 1 \end{bmatrix},$$

where $a_k = \frac{1}{3} - \frac{1}{3}\left(-\frac{1}{2}\right)^k$. One can directly verify this for N_0. Inductively assume M_{N_k} has this

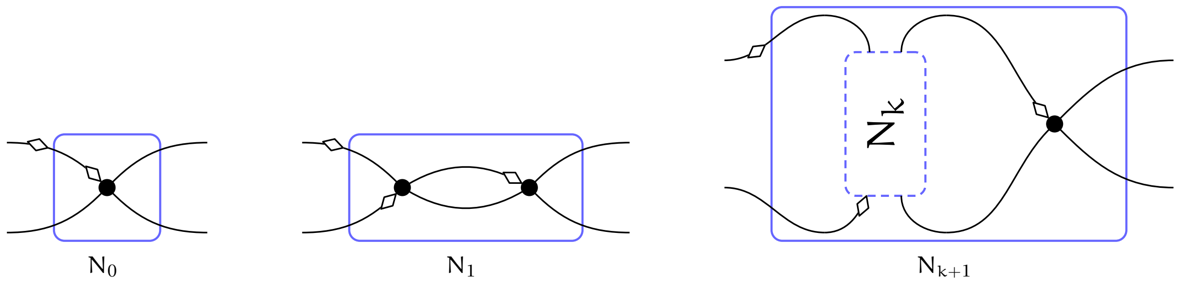


Figure 3.6: Recursive construction to approximate $[1, 0, \frac{1}{3}, 0, 1]$. Vertices are assigned g.

form. Then the rotated form of the signature matrix for N_k, as described in Figure 3.2, is

$$\begin{bmatrix} 1 & 0 & 0 & a_{k+1} \\ 0 & a_k & a_{k+1} & 0 \\ 0 & a_{k+1} & a_k & 0 \\ a_{k+1} & 0 & 0 & 1 \end{bmatrix}. \tag{3.9}$$

The action of g on the far right side of N_{k+1} is to replace each of the middle two entries in the middle two rows of the matrix in (3.9) with their average, $(a_k + a_{k+1})/2 = a_{k+2}$. This gives $M_{N_{k+1}}$.

Clearly this construction is planar.

Let $\Omega = (G, \pi)$ be an instance of Pl-Holant(f). Suppose $G = (V, E)$ and $|V| = n$. Then $|E| = 2n$. Let $H_f = \text{Holant}(\Omega; f)$ be its Holant value. Let $H_{N_k} = \text{Holant}(\Omega; N_k)$ be the Holant where every vertex in G is assigned N_k. Since each signature entry in f can be expressed as a rational number with denominator 3, each term in H_f can be expressed as a rational number with denominator 3^n, and H_{N_k} itself is a sum of 2^{2n} such terms, as $|E| = 2n$. If the error $|H_{N_k} - H_f|$ is at most $1/3^{n+1}$, then we can recover H_f from H_{N_k} by selecting the nearest rational number to H_{N_k} with denominator 3^n.

For each signature entry x in M_f, its corresponding entry $\tilde{x}$ in M_{N_k} satisfies $|\tilde{x} - x| \leqslant x/2^k$. Then for each term t in the Holant sum H_f, its corresponding term $\tilde{t}$ in the sum H_{N_k} satisfies $t(1 - 1/2^k)^n \leqslant \tilde{t} \leqslant t(1 + 1/2^k)^n$, thus $-t(1 - (1 - 1/2^k)^n) \leqslant \tilde{t} - t \leqslant t((1 + 1/2^k)^n - 1)$. Since $1 - (1 - 1/2^k)^n \leqslant (1 + 1/2^k)^n - 1$, we get $|\tilde{t} - t| \leqslant t((1 + 1/2^k)^n - 1)$. Also each term $t \leqslant 1$.

Hence

$$|H_{N_k} - H_{\mathcal{O}}| \leqslant 2^{2n}((1 + 1/2^k)^n - 1) < 1/3^{n+1},$$

if we take $k = 4n$. The construction is of linear size, and hence the reduction is in polynomial time. □

In Lemma 3.27, we will show that every signature of arity 4 with non-singular compressed matrix is able to interpolate the identity element id_4. There are three cases in Lemma 3.27 and one of them requires the following technical lemma.

Lemma 3.26. *Let* $M = [B_0\ B_1\ \cdots\ B_t]$ *be an* n*-by-*n *block matrix such that there exists a* $\lambda \in \mathbb{C}$*, for all integers* $0 \leqslant k \leqslant t$*, block* B_k *is an* n*-by-*c_k *matrix for some integer* $c_k \geqslant 0$*, and the entry of* B_k *at row* r *and column* c *is* $(B_k)_{rc} = r^{c-1}\lambda^{kr}$*, where* $r, c \geqslant 1$*. If* λ *is nonzero and is not a root of unity, then* M *is nonsingular.*

Proof. We prove by induction on n. If $n = 1$, then the sole entry is λ^k for some nonnegative integer k. This is nonzero since $\lambda \neq 0$. Assume $n > 1$ and let the left-most nonempty block be B_j. We divide row r by λ^{jr}, which is allowed since $\lambda \neq 0$. This effectively changes block B_ℓ into a block of the form $B_{\ell-j}$. Thus, we have another matrix of the same form as M but with a nonempty block B_0. To simplify notation, we also denote this matrix again by M. The first column of B_0 is all 1's. We subtract row $r-1$ from row r, for r from n down to 2. This gives us a new matrix M', and $\det M = \det M'$. Then $\det M'$ is the determinant of the $(n-1)$-by-$(n-1)$ submatrix M'' obtained from M' by removing the first row and column. Now we do column operations (on M'') to return the blocks to the proper form so that we can invoke the induction hypothesis.

For any block B'_k different from B'_0, we prove by induction on the number of columns in B'_k that B'_k can be repaired. In the base case, the rth element of the first column is $(B'_k)_{r1} = \lambda^{kr} - \lambda^{k(r-1)} = \lambda^{k(r-1)}(\lambda^k - 1)$ for $r \geqslant 2$. We divide this column by $\lambda^k - 1$ to obtain $\lambda^{k(r-1)}$, which is valid since λ is not a root of unity and $k \neq 0$. This is now the correct form for the rth element of the first column of a block in M''.

Now for the inductive step, assume that the first $d-1$ columns of block B'_k are in the correct form to be a block in M''. That is, for row index $r \geqslant 2$, which denotes the $(r-1)$-th row of M'', the rth element in the first $d-1$ columns of B'_k have the form $(B'_k)_{rc} = (r-1)^{c-1}\lambda^{k(r-1)}$. The rth element in column d of B'_k currently has the form $(B'_k)_{rd} = r^{d-1}\lambda^{kr} - (r-1)^{d-1}\lambda^{k(r-1)}$. Then we do column operations

$$\begin{aligned}(B'_k)_{rd} - \sum_{c=1}^{d-1}\binom{d-1}{c-1}(B'_k)_{rc} &= r^{d-1}\lambda^{kr} - (r-1)^{d-1}\lambda^{k(r-1)} - \sum_{c=1}^{d-1}\binom{d-1}{c-1}(r-1)^{c-1}\lambda^{k(r-1)} \\ &= r^{d-1}\lambda^{kr} - r^{d-1}\lambda^{k(r-1)} \\ &= r^{d-1}\lambda^{k(r-1)}(\lambda^k - 1)\end{aligned}$$

and divide by $(\lambda^k - 1)$ to get $r^{d-1}\lambda^{k(r-1)}$. Once again, this is valid since λ is not a root of unity and $k \neq 0$. Then more (of the same) column operations yield

$$\begin{aligned}&r^{d-1}\lambda^{k(r-1)} - \sum_{c=1}^{d-1}\binom{d-1}{c-1}(r-1)^{c-1}\lambda^{k(r-1)} \\ &\qquad = \lambda^{k(r-1)}\left(r^{d-1} + (r-1)^{d-1} - \sum_{c=1}^{d}\binom{d-1}{c-1}(r-1)^{c-1}\right)\end{aligned}$$

and the term in parentheses is precisely $(r-1)^{d-1}$. This gives the correct form for the rth element in column d of B'_k in M''.

Now we repair the columns in B'_0, also by induction on the number of columns. In the base case, if B'_0 only has one column, then there is nothing to prove, since this block has disappeared in M''. Otherwise, $(B'_0)_{r2} = r - (r-1) = 1$, so the second column is already in the correct form to be the first column in M'', and there is still nothing to prove. For the inductive step, assume that columns 2 to $d-1$ are in the correct form to be the first block in M'' for $d \geqslant 3$. That is, the entry at row $r \geqslant 2$ and column c from 2 through $d-1$ has the form $(B'_0)_{rc} = (r-1)^{c-2}$. The rth element in column d currently has the form $(B'_0)_{rd} = r^{d-1} - (r-1)^{d-1}$. Then we do the column operations

$$\begin{aligned}(B'_0)_{rd} - \sum_{c=2}^{d-1}\binom{d-1}{c-2}(B'_0)_{rc} &= r^{d-1} - (r-1)^{d-1} - \sum_{c=2}^{d-1}\binom{d-1}{c-2}(r-1)^{c-2} \\ &= (d-1)(r-1)^{d-2}\end{aligned}$$

and divide by $d-1$, which is nonzero, to get $(r-1)^{d-2}$. This is the correct form for the rth element in column d of B_0' in M''. Therefore, we invoke our original induction hypothesis that the $(n-1)$-by-$(n-1)$ matrix M'' has a nonzero determinant, which completes the proof. □

Lemma 3.27. *Let* id_4 *be the arity* 4 *signature with* M_{id_4} *given in* (3.4) *and let* f *be an arity* 4 *signature. If* M_f *is redundant and* $\widetilde{M_f}$ *is nonsingular, then for any set* $\mathcal{F}$ *containing* f*, we have that*

$$\operatorname{Holant}(\mathcal{F} \cup \{\mathrm{id}_4\}) \leqslant_T \operatorname{Holant}(\mathcal{F}).$$

Proof. Consider an instance Ω of $\operatorname{Holant}(\mathcal{F} \cup \{\mathrm{id}_4\})$. Suppose that id_4 appears n times in Ω. We construct from Ω a sequence of instances Ω_s of $\operatorname{Holant}(\mathcal{F})$ indexed by $s \geqslant 1$. We obtain Ω_s from Ω by replacing each occurrence of id_4 with the gadget N_s in Figure 3.7 with f assigned to all vertices. In Ω_s, the edge corresponding to the ith significant index bit of N_s connects to the same location as the edge corresponding to the ith significant index bit of id_4 in Ω.

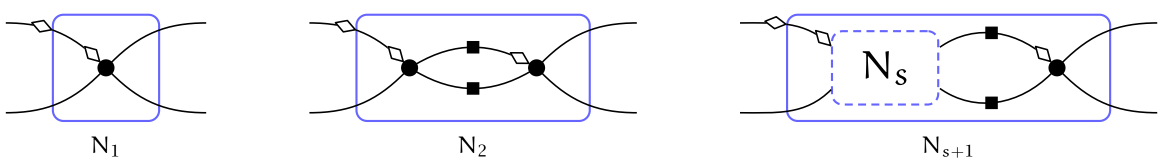


Figure 3.7: Recursive construction to interpolate id_4. All vertices are assigned f.

Now to determine the relationship between $\operatorname{Holant}_\Omega$ and $\operatorname{Holant}_{\Omega_s}$, we use the isomorphism between redundant 4-by-4 matrices and 3-by-3 matrices. To obtain Ω_s from Ω, we effectively replace M_{id_4} with $M_{N_s} = (M_f)^s$, the sth power of the signature matrix M_f. By the Jordan normal form of $\widetilde{M_f}$, there exist $T, \Lambda \in \mathbb{C}^{3\times 3}$ such that

$$\widetilde{M_f} = T\Lambda T^{-1} = T \begin{bmatrix} \lambda_1 & b_1 & 0 \\ 0 & \lambda_2 & b_2 \\ 0 & 0 & \lambda_3 \end{bmatrix} T^{-1},$$

where $b_1, b_2 \in \{0, 1\}$. Note that $\lambda_1\lambda_2\lambda_3 = \det(\widetilde{M_f}) \neq 0$. Also since $\widetilde{M_{\mathrm{id}_4}} = \varphi(M_{\mathrm{id}_4}) = I_3$, and $TI_3T^{-1} = I_3$, we have $\psi(T)M_{\mathrm{id}_4}\psi(T^{-1}) = M_{\mathrm{id}_4}$. We can view our construction of Ω_s as first replacing each M_{id_4} by $\psi(T)M_{id_4}\psi(T^{-1})$, which does not change the Holant value, and then replacing each new M_{id_4} with $\psi(\Lambda^s) = \psi(\Lambda)^s$ to obtain Ω_s. Observe that

$$\varphi(\psi(T)\psi(\Lambda^s)\psi(T^{-1})) = T\Lambda^s T^{-1} = (\widetilde{M_f})^s = (\varphi(M_f))^s = \varphi((M_f)^s).$$

Hence, $\psi(T)\psi(\Lambda^s)\psi(T^{-1}) = M_{N_s}$. (Since $M_{id_4} = \psi(T)M_{id_4}\psi(T^{-1})$ and $M_{N_s} = \psi(T)\psi(\Lambda^s)$ $\psi(T^{-1})$, replacing each M_{id_4}, sandwiched between $\psi(T)$ and $\psi(T^{-1})$, by $\psi(\Lambda^s)$ indeed transforms Ω to Ω_s. We also note that, by the isomorphism, $\psi(T^{-1})$ *is* the multiplicative inverse of $\psi(T)$ within the semi-group of redundant 4-by-4 matrices; but we prefer not to write it as $\psi(T)^{-1}$ since it is not the usual matrix inverse as a 4-by-4 matrix. Indeed, $\psi(T)$ is not invertible as a 4-by-4 matrix.)

In the case analysis below, we stratify the assignments in Ω_s based on the assignment to $\psi(\Lambda^s)$. The inputs to $\psi(\Lambda^s)$ are from $\{0,1\}^2 \times \{0,1\}^2$. However, we can combine the input 01 and 10, since $\psi(\Lambda^s)$ is redundant. Thus we actually stratify the assignments in Ω_s based on the assignment to Λ^s, which takes inputs from $\{0,1,2\} \times \{0,1,2\}$. In this compressed form, the row and column assignments to Λ^s are the Hamming weight of the two actual binary valued inputs to the uncompressed form $\psi(\Lambda^s)$.

Now we begin the case analysis on the values of b_1 and b_2.

1. Assume $b_1 = b_2 = 0$. We only need to consider the assignments to Λ^s that assign

   - $(0,0)$ i many times,
   - $(1,1)$ j many times, and
   - $(2,2)$ k many times

   since any other assignment contributes a factor of 0. Let c_{ijk} be the sum over all such assignments of the products of evaluations of all signatures in Ω_s except for Λ^s (including the contributions from T and T^{-1}). Note that this quantity is the same in Ω as in Ω_s. In particular it does not depend on s. Then

$$\text{Holant}_\Omega = \sum_{i+j+k=n} \frac{c_{ijk}}{2^j}$$

and the value of the Holant on Ω_s, for $s \geqslant 1$, is

$$\text{Holant}_{\Omega_s} = \sum_{i+j+k=n} \left(\lambda_1^i \lambda_2^j \lambda_3^k\right)^s \left(\frac{c_{ijk}}{2^j}\right).$$

The coefficient matrix is Vandermonde, but it may not have full rank because it is possible that $\lambda_1^i \lambda_2^j \lambda_3^k = \lambda_1^{i'} \lambda_2^{j'} \lambda_3^{k'}$ for some $(i, j, k) \neq (i', j', k')$, where $i + j + k = i' + j' + k' = n$. However, this is not a problem since we are only interested in the sum $\sum \frac{c_{ijk}}{2^j}$. If two coefficients are the same, we replace their corresponding unknowns $c_{ijk}/2^j$ and $c_{i'j'k'}/2^{j'}$ with their sum as a new variable. After all such combinations, we have a Vandermonde system of full rank. In particular, none of the entries are 0 since $\lambda_1\lambda_2\lambda_3 = \det(\widetilde{M_f}) \neq 0$. Therefore, we can solve the linear system and obtain the value of Holant_Ω.

2. Assume $b_1 \neq b_2$. We can permute the Jordan blocks in Λ so that $b_1 = 1$ and $b_2 = 0$, then $\lambda_1 = \lambda_2$, denoted by λ. We only need to consider the assignments to Λ^s that assign

   - $(0, 0)$ i many times,
   - $(1, 1)$ j many times,
   - $(2, 2)$ k many times, and
   - $(0, 1)$ ℓ many times

   since any other assignment contributes a factor of 0. Let $c_{ijk\ell}$ be the sum over all such assignments of the products of evaluations of all signatures in Ω_s except for Λ^s (including the contributions from T and T^{-1}). Then

$$\text{Holant}_\Omega = \sum_{i+j+k=n} \frac{c_{ijk0}}{2^j}$$

   and the value of the Holant on Ω_s, for $s \geqslant 1$, is

$$\begin{aligned}\text{Holant}_{\Omega_s} &= \sum_{i+j+k+\ell=n} \lambda^{(i+j)s} \lambda_3^{ks} \left(s\lambda^{s-1}\right)^\ell \left(\frac{c_{ijk\ell}}{2^{j+\ell}}\right) \\ &= \lambda^{ns} \sum_{i+j+k+\ell=n} \left(\frac{\lambda_3}{\lambda}\right)^{ks} s^\ell \left(\frac{c_{ijk\ell}}{\lambda^\ell 2^{j+\ell}}\right).\end{aligned}$$

If λ_3/λ is a root of unity, then take a t such that $(\lambda_3/\lambda)^t = 1$. Then

$$\text{Holant}_{\Omega_{st}} = \lambda^{nst} \sum_{i+j+k+\ell=n} s^{\ell} \left(\frac{t^{\ell} c_{ijk\ell}}{\lambda^{\ell} 2^{j+\ell}} \right).$$

For $s \geqslant 1$, this gives a coefficient matrix that is Vandermonde. Although this system is not full rank, we can replace all the unknowns $c_{ijk\ell}/2^j$ having $i + j + k = n - \ell$ by their sum to form new unknowns $c'_\ell = \sum_{i+j+k=n-\ell} \frac{c_{ijk\ell}}{2^j}$, where $0 \leqslant \ell \leqslant n$. The new unknown c'_0 is the Holant of Ω that we seek. The resulting Vandermonde system

$$\text{Holant}_{\Omega_{st}} = \lambda^{nst} \sum_{\ell=0}^{n} s^{\ell} \left(\frac{t^{\ell} c'_{\ell}}{\lambda^{\ell} 2^{\ell}} \right)$$

has full rank, so we can solve for the new unknowns and obtain the value of $\text{Holant}_{\Omega} = c'_0$.

If λ_3/λ is not a root of unity, then we replace all the unknowns $c_{ijk\ell}/(\lambda^{\ell} 2^{j+\ell})$ having $i+j = m$ with their sum to form new unknowns $c'_{mk\ell}$, for any $0 \leqslant m, k, \ell$ and $m+k+\ell = n$. The Holant of Ω is now

$$\text{Holant}_{\Omega} = \sum_{m+k=n} c'_{mk0}$$

and the value of the Holant on Ω_s is

$$\begin{aligned}\text{Holant}_{\Omega_s} &= \lambda^{ns} \sum_{i+j+k+\ell=n} \left(\frac{\lambda_3}{\lambda} \right)^{ks} s^{\ell} \left(\frac{c_{ijk\ell}}{\lambda^{\ell} 2^{j+\ell}} \right) \\ &= \lambda^{ns} \sum_{m+k+\ell=n} \left(\frac{\lambda_3}{\lambda} \right)^{ks} s^{\ell} c'_{mk\ell}.\end{aligned}$$

After a suitable reordering of the columns, the matrix of coefficients satisfies the hypothesis of Lemma 3.26. Therefore, the linear system has full rank. We can solve for the unknowns and obtain the value of Holant_{Ω}.

3. Assume $b_1 = b_2 = 1$. In this case, we have $\lambda_1 = \lambda_2 = \lambda_3$, denoted by λ, and we only need to consider the assignments to Λ^s that assign

   - $(0, 0)$ or $(2, 2)$ i many times,

- $(1,1)$ j many times,
- $(0,1)$ k many times,
- $(1,2)$ ℓ many times, and
- $(0,2)$ m many times

since any other assignment contributes a factor of 0. Let $c_{ijk\ell m}$ be the sum over all such assignments of the products of evaluations of all signatures in Ω_s except for Λ^s (including the contributions from T and T^{-1}). Then

$$\text{Holant}_\Omega = \sum_{i+j=n} \frac{c_{ij000}}{2^j}$$

and the value of the Holant on Ω_s, for $s \geqslant 1$, is

$$\begin{aligned}\text{Holant}_{\Omega_s} &= \sum_{i+j+k+\ell+m=n} \lambda^{(i+j)s} \left(s\lambda^{s-1}\right)^{k+\ell} \left(s(s-1)\lambda^{s-2}\right)^m \left(\frac{c_{ijk\ell m}}{2^{j+k+m}}\right) \\ &= \lambda^{ns} \sum_{i+j+k+\ell+m=n} s^{k+\ell+m}(s-1)^m \left(\frac{c_{ijk\ell m}}{\lambda^{k+\ell+2m}2^{j+k+m}}\right).\end{aligned}$$

We replace all the unknowns $c_{ijk\ell m}/(\lambda^{k+\ell+2m}2^{j+k+m})$ having $i+j=p$ and $k+\ell=q$ by their sums to form new unknowns c'_{pqm}, for any $0 \leqslant p, q, m$ and $p+q+m=n$. The Holant of Ω is now c'_{n00}. This new linear system is

$$\text{Holant}_{\Omega_s} = \lambda^{ns} \sum_{p+q+m=n} s^{q+m}(s-1)^m c'_{pqm},$$

but is still rank deficient. We now index the columns by (q, m), where $q \geqslant 0$, $m \geqslant 0$, and $q+m \leqslant n$. Correspondingly, we rename the variables $x_{q,m} = c'_{pqm}$. Note that $p = n-q-m$ is determined by (q, m). Observe that the column indexed by (q, m) is the sum of the columns indexed by $(q-1, m)$ and $(q-2, m+1)$ provided $q-2 \geqslant 0$. Namely, $s^{q+m}(s-1)^m = s^{q-1+m}(s-1)^m + s^{q-2+m+1}(s-1)^{m+1}$. Of course this is only meaningful if $q \geqslant 2$, $m \geqslant 0$ and $q+m \leqslant n$. We write the linear system as

$$\sum_{q\geqslant 0,\ m\geqslant 0,\ q+m\leqslant n} \alpha_{q,m} x_{q,m} = \frac{\text{Holant}_{\Omega_s}}{\lambda^{ns}},$$

where $\alpha_{q,m} = s^{q+m}(s-1)^m$ are the coefficients. Hence $\alpha_{q,m}x_{q,m} = \alpha_{q-1,m}x_{q,m} + \alpha_{q-2,m+1}x_{q,m}$, and we define new variables

$$x_{q-1,m} \leftarrow x_{q,m} + x_{q-1,m}$$
$$x_{q-2,m+1} \leftarrow x_{q,m} + x_{q-2,m+1}$$

from $q = n - m$ down to 2 for every $0 \leqslant m \leqslant n-2$.

Observe that in each update, the newly defined variables have a decreased index value for q. A more crucial observation is that the column indexed by $(0,0)$ is never updated. This is because, in order to be an updated entry, there must be some $q \geqslant 2$ and $m \geqslant 0$ such that $(q-1,m) = (0,0)$ or $(q-2,m+1) = (0,0)$, which is clearly impossible. Hence $x_{0,0} = c'_{n00}$ is still the Holant value on Ω. The $2n+1$ unknowns that remain are

$$x_{0,0},\ x_{1,0},\ x_{0,1},\ x_{1,1},\ x_{0,2},\ x_{1,2},\ \ldots,\ x_{0,n-1},\ x_{1,n-1},\ x_{0,n}$$

and their coefficients in row s are

$$1, s, s(s-1), s^2(s-1), s^2(s-1)^2, \ldots, s^{n-1}(s-1)^{n-1}, s^n(s-1)^{n-1}, s^n(s-1)^n.$$

It is clear that the κ-th entry in this row is a monic polynomial in s of degree κ, where $0 \leqslant \kappa \leqslant 2n$, and thus s^κ is a linear combination of the first κ entries. It follows that the coefficient matrix is a product of the standard Vandermonde matrix multiplied to its right by an upper triangular matrix with all 1's on the diagonal. Therefore, the linear system has full rank. We can solve for these final unknowns and obtain the value of $\text{Holant}_\Omega = x_{0,0} = c'_{n00}$. □

We summarize our progress with the following corollary, which combines Lemmas 3.25 and 3.27.

Corollary 3.28. *Let* f *be an arity 4 signature with complex weights. If* M_f *is redundant and* $\widetilde{M_f}$ *is nonsingular, then* Pl-Holant(f) *is #**P**-hard.*

We also have a consequence of Corolloary 3.28, weaker yet easier to apply, concerning symmetric signatures.

Corollary 3.29. *For a symmetric arity 4 signature* $[f_0, f_1, f_2, f_3, f_4]$ *with complex weights, if there does* not *exist* $a, b, c \in \mathbb{C}$*, not all zero, such that for all* $k \in \{0, 1, 2\}$*,*

$$af_k + bf_{k+1} + cf_{k+2} = 0,$$

then Pl-Holant(f) *is #***P***-hard.*

By Lemma 3.19, for a nonsingular matrix $T \in \mathbb{C}^{2\times 2}$, M_f is redundant and $\widetilde{M_f}$ is nonsingular if and only if $M_{\hat{f}}$ is redundant and $\widetilde{M_{\hat{f}}}$ is nonsingular. Combine this fact with Corollary 3.28 and Lemma 3.19, and we have the following.

Corollary 3.30. *Let* f *be an arity* 4 *signature with complex weights. If there exists a nonsingular matrix* $T \in \mathbb{C}^{2\times 2}$ *such that* $\hat{f} = T^{\otimes 4} f$*, where* $M_{\hat{f}}$ *is redundant and* $\widetilde{M_{\hat{f}}}$ *is nonsingular, then* Holant(f) *is #***P***-hard.*

3.5 A Unary Interpolation Lemma

Before continuing to the main result of this chapter, we prove an interpolation lemma for unary signatures first. It is useful in the following the rest of this chapter as well as in the next.

Lemma 3.31. *Suppose* $M \in \mathbb{C}^{n\times n}$ *and* $s \in \mathbb{C}^{n\times 1}$*. If the following three conditions are satisfied,*

1. $\det(M) \neq 0$*;*
2. s *is not orthogonal to any row eigenvector of* M*;*
3. M *has infinite order modulo a scalar;*

then vectors in the set $S = \{M^k s \mid k \geqslant 0\}$ *are pairwise linearly independent.*

Proof. Since $\det(M) \neq 0$, M is nonsingular and the eigenvalues λ_i of M, for $1 \leqslant i \leqslant n$, are nonzero. Let $M = P^{-1}JP$ be the Jordan decomposition of M and let $p = Ps \in \mathbb{C}^{n\times 1}$. Suppose for a contradiction that vectors in S are not pairwise linearly independent. This means that there exists integers $k > \ell \geqslant 0$ such that $M^k s = \beta M^\ell s$ for some $\beta \neq 0$. Let $t = k - \ell > 0$. Then we have that $P^{-1} J^t P s = M^t s = \beta s$ and hence $J^t p = \beta p$.

Suppose that J contains some nontrivial Jordan block and consider the 2-by-2 submatrix in the bottom right corner of this block. From this portion of J, the two equations given by $J^t p = \beta p$ are $\lambda_i^t p_{i-1} + t\lambda_i^{t-1} p_i = \beta p_{i-1}$ and $\lambda_i^t p_i = \beta p_i$. Since s is not orthogonal to any row eigenvector of M, $p_i \neq 0$. Then these equations imply that $t\lambda_i^{t-1} p_i = 0$, a contradiction.

Otherwise, M is diagonalizable and J contains only trivial Jordan blocks. From $J^t p = \beta p$, we get the equations $\lambda_i^t p_i = \beta p_i$ for $1 \leqslant i \leqslant n$. Since s is not orthogonal to any row eigenvector of M, $p_i \neq 0$ for $1 \leqslant i \leqslant n$. Hence $\lambda_i^t = 1$ for all $1 \leqslant i \leqslant n$. Therefore $M^t = \beta I_n$, which contradicts that fact that M has infinite order modulo a scalar. □

Recall that for a binary signature f, its matrix M_f is defined in the same way as in Definition 3.17:

$$M_f = \begin{bmatrix} f(00) & f(01) \\ f(10) & f(11) \end{bmatrix}.$$

Note that here we do not need to worry about how to order the edges of f.

Lemma 3.32. *Let $\mathcal{F}$ be a set of signatures. If there exists a planar $\mathcal{F}$-gate with signature matrix $M \in \mathbb{C}^{2\times 2}$ and a planar $\mathcal{F}$-gate with signature $s \in \mathbb{C}^{2\times 1}$ satisfying the following conditions,*

1. $\det(M) \neq 0$;
2. $\det([s\ Ms]) \neq 0$;
3. M *has infinite order modulo a scalar;*

then $\text{Pl-Holant}(\mathcal{F} \cup \{[a, b]\}) \leqslant_T \text{Pl-Holant}(\mathcal{F})$ *for any* $a, b \in \mathbb{C}$.

Proof. Let $\Omega = (G, \pi)$ be an instance of $\text{Pl-Holant}(\mathcal{F} \cup \{[a, b]\})$. Let $V' \subseteq V$ be the subset of vertices assigned $[a, b]$ by π and suppose that $|V'| = n$. We construct from Ω a sequence of instances Ω_k of $\text{Pl-Holant}(\mathcal{F})$ indexed by $k \geqslant 1$. We obtain Ω_k from Ω by replacing each occurrence of $[a, b]$ with the unary recursive construction (M, s) in Figure 3.8 containing k copies of the recursive gadget. This unary recursive construction has the signature $[a_k, b_k] = M^k s$.

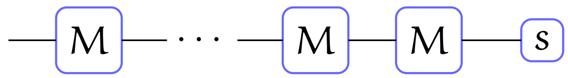


Figure 3.8: Unary recursive construction (M, s).

If s is orthogonal to a row eigenvector β of M, it is easy to verify that Ms is also orthogonal to β. Hence $[s\ Ms]$ is singular. This contradicts to $\det([s\ Ms]) \neq 0$. It implies that s is not orthogonal to any row eigenvector of M. By Lemma 3.31, vectors in the set $S = \{[a_k, b_k] \mid 0 \leqslant k \leqslant n+1\}$ are pairwise linearly independent. In particular, at most one b_k can be 0, so we may assume that $b_k \neq 0$ for $0 \leqslant k \leqslant n$, renaming if necessary.

We stratify the assignments in Ω based on the assignment to $[a, b]$. Let c_ℓ be the sum over all assignments of products of evaluations at all $v \in V \backslash V'$ such that exactly ℓ occurrences of $[a, b]$ have their incident edges assigned 0 (and $n-\ell$ have their incident edges assigned 1). Then

$$\text{Holant}_\Omega = \sum_{0 \leqslant \ell \leqslant n} a^\ell b^{n-\ell} c_\ell$$

and the value of the Holant on Ω_k, for $k \geqslant 1$, is

$$\text{Holant}_{\Omega_k} = \sum_{0 \leqslant \ell \leqslant n} a_k^\ell b_k^{n-\ell} c_\ell = b_k^n \sum_{0 \leqslant \ell \leqslant n} \left(\frac{a_k}{b_k}\right)^\ell c_\ell.$$

The coefficient matrix of this linear system is Vandermonde. Since vectors a_k, b_k in S are pairwise linearly independent, ratios a_k/b_k are distinct (and well-defined since $b_k \neq 0$), which means that the Vandermonde matrix has full rank. Therefore, we can solve the linear system for the unknown c_ℓ's and obtain the value of Holant_Ω. □

3.6 Pl-Holant Dichotomy for a Symmetric Arity 4 Signature

With Corollary 3.28 in hand, the only obstacles remaining to prove a dichotomies for a symmetric arity 4 signature are $\text{Holant}([v, 1, 0, 0, 0])$ and $\text{Pl-Holant}([v, 1, 0, 0, 0])$. On the other hand, if $v = 0$, then $\text{Pl-Holant}([v, 1, 0, 0, 0])$ is tractable, since it counts the number of perfect matchings in a 4-regular planar graph. We will show that otherwise it is #**P**-hard. On the other hand, $\text{Holant}[0, 1, 0, 0, 0]$ is #**P**-hard without the planar assumption.

Lemma 3.33. Holant$([0,1,0,0,0])$ *is #**P**-hard.*

Proof. Using the tetrahedron gadget in Figure 2.9 with $[0,1,0,0,0]$ assigned to each vertex, we get a signature $g = [3,0,1,0,1]$. Since $\det(\widetilde{M_g}) = 4$, we are done by Corollary 3.28. □

We will prove that Pl-Holant$([\nu,1,0,0,0])$ with $\nu \neq 0$ is #**P**-hard by reducing from Pl-Holant$([\nu, 1,0,0])$. These two problems are counting weighted matchings over planar k-regular graphs for $k = 4$ and $k = 3$ respectively. First we want to realize $[1,0,0]$. We will combine the idea of anti-gadgets from [CKW12] with Lemma 3.32.

Lemma 3.34. *For any* $\nu \in \mathbb{C}$ *and signature set* $\mathcal{F}$ *containing* $[\nu,1,0,0,0]$*,*

$$\text{Pl-Holant}(\mathcal{F} \cup \{[1,0,0]\}) \leqslant_T \text{Pl-Holant}(\mathcal{F}).$$

Proof of Lemma 3.34. Consider the gadget construction in Figure 3.9. For $k \geqslant 0$, the signature of N_k is of the form $[a_k, b_k, 0]$, and $N_0 = [\nu, 1, 0]$. Since N_k is symmetric and always ends with 0, we can analyze this construction as if it were a unary recursive construction. Let $s_k = \left[\begin{smallmatrix} a_k \\ b_k \end{smallmatrix}\right]$, so $s_0 = \left[\begin{smallmatrix} \nu \\ 1 \end{smallmatrix}\right]$. It is easy to verify that that $s_k = M^k s_0$, where $M = \left[\begin{smallmatrix} \nu & 2 \\ 1 & 0 \end{smallmatrix}\right]$.

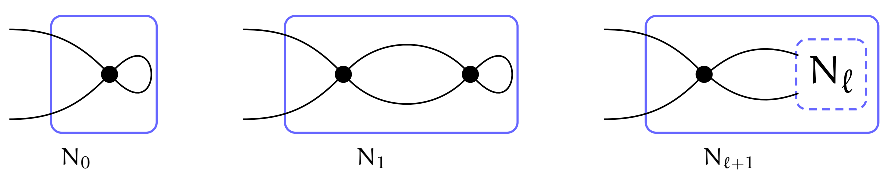


Figure 3.9: Binary recursive construction with a starter to interpolate $[1,0,0]$. Vertices are assigned $[\nu,1,0,0,0]$.

Since $\det(M) = -2$, M is nonsingular. If M has finite order modulo a scalar, then $M^\ell = \beta I_2$ for some positive integer ℓ and some nonzero complex value β. Thus, the signature of $N_{\ell-1}$, which contains the anti-gadget of M, is $M^{\ell-1} s_0 = \beta M^{-1} s_0 = \beta \left[\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right]$. After normalizing, we directly realize $[1,0,0]$.

Otherwise M has infinite order modulo a scalar. It is easy to verify that $\det([s_0 \; M s_0]) = -2$. Hence by the unary interpolation as in Lemma 3.32, we have that

$$\text{Pl-Holant}(\mathcal{F} \cup \{[a,b,0]\}) \leqslant_T \text{Pl-Holant}(\mathcal{F}).$$

for any $a, b \in \mathbb{C}$. The lemma follows from the claim by setting $a = 1$ and $b = 0$. □

Note that towards the end of the proof of Lemma 3.34, Lemma 3.32 does not really directly apply, as we are interpolating a binary signature, instead of a unary one. However, the interpolation is almost identically the same, as the support of the binary signature has dimension two, which is the same as a unary signature.

For the next lemma, we use a well-known and easy generalization of a classic result by Petersen [Pet91]. Petersen's theorem considers 3-regular, bridgeless, simple graphs (i.e. graphs without self-loops or parallel edges) and concludes that there exists a perfect matching. The same conclusion holds even if the graphs are not simple. We provide a proof for completeness.

Theorem 3.35. *Any 3-regular bridgeless graph G has a perfect matching.*

Proof. We may assume that G is connected. If G has a vertex v with a self-loop, then the other edge of v is a bridge since G is 3-regular, which is a contradiction. If there exists some pair of vertices of G joined by exactly three parallel edges, then G has only these two vertices since it is connected and the theorem holds.

In the remaining case, there exists some pair of vertices joined by exactly two parallel edges. We build a new graph G' without any parallel edges. For vertices u and v joined by exactly two parallel edges, we remove these two parallel edges and introduce two new vertices w_1 and w_2. We also introduce the new edges (u, w_1), (u, w_2), (v, w_1), (v, w_2), and (w_1, w_2). Then G' is a 3-regular, bridgeless, simple graph.

By Petersen's theorem, G' has a perfect matching P'. Now we construct a perfect matching P in G using P'. We put any edge in both G and P' into P. If u is matched by a new edge in G', then v must be matched by a new edge in G' as well and we put the edge (u, v) into P. If u and v are not matched by a new edge, then we do not add anything to P. It is easy to see that P is a perfect matching in G. □

We use this result to show the existence of what we call a *planar pairing* for any planar 3-regular graph, which we use in our proof of #**P**-hardness.

Definition 3.36 (Planar pairing)**.** *A* planar pairing *in a graph* $G = (V, E)$ *is a set of edges* $P \subset V \times V$ *such that* P *is a perfect matching in the graph* $(V, V \times V)$*, and the graph* $(V, E \cup P)$ *is planar.*

Obviously, a perfect matching in the original graph is a planar pairing.

Lemma 3.37. *For any planar 3-regular graph* G*, there exists a planar pairing that can be computed in polynomial time.*

Proof. We efficiently find a planar pairing in G by induction on the number of vertices in G. Since G is a 3-regular graph, it must have an even number of vertices. If there are no vertices in G, then there is nothing to do. Suppose that G has $n = 2k$ vertices and that we can efficiently find a planar pairing in graphs containing fewer vertices. If G is not connected, then we can already apply our inductive hypothesis on each connected component of G. The union of planar pairings in each connected component of G is a planar pairing in G, so we are done. Otherwise assume that G is connected.

(a) The neighborhood around u and v in G.
(b) The same neighborhood in H.

Figure 3.10: The neighborhood around u and v both before and after they are removed.

Suppose that G contains a bridge (u, v). Let the three (though not necessarily distinct) neighbors of u be v, u_1, and u_2, and let the three (though not necessarily distinct) neighbors of v be u, v_1, and v_2 (see Figure 3.10a). Furthermore, let H_u be the connected component in $G - \{(u, v)\}$ containing u and let H_v be the connected component in $G - \{(u, v)\}$ containing v. Consider the induced subgraph H'_u of H_u after adding the edge $e_u = (u_1, u_2)$ (which might be a self-loop on $u = u_1 = u_2$) and then removing u. Similarly, consider the induced subgraph H'_v of H_v after adding the edge $e_v = (v_1, v_2)$ (which might be a self-loop on $v = v_1 = v_2$) and then removing v. Both H'_u and H'_v are 3-regular graphs and their disjoint union gives a graph H' with $n - 2 = 2(k - 1)$ vertices (see Figure 3.10b).

By induction on both H'_u and H'_v, we have planar pairings P_u and P_v in H'_u and H'_v respectively. Let H'' be the graph H' including the edges $P_u \cup P_v$. If H'' contains both e_u and e_v, then embed H'' in the plane so that both e_u and e_v are adjacent to the outer face. Otherwise, any planar embedding will do. Then the graph G including the edges $P_u \cup P_v$ is also planar, so $P_u \cup P_v \cup \{(u, v)\}$ is a planar pairing in G.

Otherwise, G is bridgeless. Then by Theorem 3.35, G has a perfect matching, which is also a planar pairing in G. Since a perfect matching can be found in polynomial time by Edmond's blossom algorithm [Edm65], the whole procedure is in polynomial time. □

The approach above to find a planar pairing is reported in [GW13]. Alternatively, an algorithm by Cai and Kowalczyk [CK13] achieves the same goal. It was used to show that counting vertex covers over k-regular graphs is #**P**-hard for even $k \geqslant 4$ (see the proof of Lemma 15 in [CK13]). Their algorithm to find a planar pairing starts by taking a spanning tree and then pairing up the vertices on this tree. The planar pairing idea is special, in the sense that the argument is global. In contrast, most gadget constructions in our hardness proofs are local. Planar pairings permit reductions that are not otherwise possible.

Now we use the planar pairing technique to show the following.

Lemma 3.38. *Let $\nu \neq 0 \in \mathbb{C}$.* Pl-Holant$([\nu, 1, 0, 0, 0])$ *is* #**P**-*hard.*

Proof. We reduce from Pl-Holant$([\nu, 1, 0, 0])$ to Pl-Holant$([\nu, 1, 0, 0, 0], [1, 0, 0])$. Since Pl-Holant$([\nu, 1, 0, 0])$ is #**P**-hard when $\nu \neq 0$ by Theorem 1.14, the reduction implies that Pl-Holant$([\nu, 1, 0, 0, 0])$ is also #**P**-hard when $\nu \neq 0$ by Lemma 3.34.

Let $\Omega = (G, \pi)$ be an instance of Pl-Holant$([\nu, 1, 0, 0])$. Then $G = (V, E)$ is planar and 3-regular. By Lemma 3.37, there exists a planar pairing P in G and it can be found in polynomial time. Then the graph $G' = (V, E \cup P)$ is planar and 4-regular. We assign $[\nu, 1, 0, 0, 0]$ to every vertex in G'. Moreover, we replace each edge in P with a path of length 2 to form a graph G'' and assign $[1, 0, 0] = [1, 0]^{\otimes 2}$ to each of the new vertices. Call this new instance Ω''. Then $\text{Holant}(\Omega; [\nu, 1, 0, 0]) = \text{Holant}(\Omega''; \{[\nu, 1, 0, 0, 0], [1, 0, 0]\})$. □

Note that our proof of Lemma 3.38 reduces Pl-Holant$([\nu, 1, 0, 0])$ to Pl-Holant$([\nu, 1, 0, 0, 0])$ for all $\nu \in \mathbb{C}$. Neither Lemma 3.34 nor Lemma 3.38 ever considers the value of ν. This is consistent because both signatures are in $\mathcal{M}$ when $\nu = 0$, thus tractable, and both signatures are #**P**-hard when ν is different from 0.

Now we are ready to prove our Pl-Holant dichotomy for a symmetric arity 4 signature.

Theorem 3.39. *Let* f *be a non-degenerate, complex-weighted Boolean signature.* Pl-Holant(f) *is* #**P**-*hard unless* f *satisfies one of the following conditions, in which case the problem is computable in polynomial time:*

1. f *is* $\mathcal{A}$- *or* $\mathcal{P}$-*transformable;*

2. f *is vanishing;*

3. f *is* $\mathcal{M}$-*transformable.*

If f *satisfies condition* 1 *or* 2*, then* Holant(f) *is computable in polynomial time without planarity; otherwise* Holant(f) *is* #**P**-*hard.*

Proof. Tractability follows from Lemma 1.7, Lemma 1.9, and Lemma 1.10, as well as the fact that vanishing signatures are tractable.

Suppose $f = [f_0, f_1, f_2, f_3, f_4]$. If there do not exist $a, b, c \in \mathbb{C}$, not all zero, such that for all $k \in \{0, 1, 2\}$, $af_k + bf_{k+1} + cf_{k+2} = 0$, then Pl-Holant$(f)$ is #**P**-hard by Corollary 3.29. Otherwise, such a, b, c exist. If $a = c = 0$, then $b \neq 0$, so $f_1 = f_2 = f_3 = 0$. In this case, $f \in \mathcal{P}$ is a generalized equality signature, so f is $\mathcal{P}$-transformable.

Now suppose a and c are not both 0. If $b^2 - 4ac \neq 0$, then $f_k = \alpha_1^{4-k}\alpha_2^k + \beta_1^{4-k}\beta_2^k$, where $\alpha_1\beta_2 - \alpha_2\beta_1 \neq 0$. A holographic transformation by $\left[\begin{smallmatrix} \alpha_1 & \beta_1 \\ \alpha_2 & \beta_2 \end{smallmatrix}\right]$ transforms f to $=_4$. Apply Theorem 1.15 and we have tha Pl-Holant(f) is #**P**-hard unless f is $\mathcal{A}$-, $\mathcal{P}$-, or $\mathcal{M}$-transformable. Moreover, if f is $\mathcal{M}$-transformable, Holant(f) is #**P**-hard.

The exceptional case is $b^2 - 4ac = 0$. There are two symmetric possibilities. In the first, for any $0 \leqslant k \leqslant 2$, $f_k = ck\alpha^{k-1} + d\alpha^k$, where $c \neq 0$. In the second, for any $0 \leqslant k \leqslant 2$, $f_k = c(4-k)\alpha^{3-k} + d\alpha^{4-k}$, where $c \neq 0$. These two possibilities map between each other under a holographic transformation by $\left[\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right]$, so assume that the first one holds.

If $\alpha = \pm i$, then f is vanishing. Otherwise, a further holographic transformation by $\frac{1}{\sqrt{1+\alpha^2}}$ $\left[\begin{smallmatrix} 1 & \alpha \\ \alpha & -1 \end{smallmatrix}\right]$ transforms f to $\bar{f} = [\nu, 1, 0, 0, 0]$ for some $\nu \in \mathbb{C}$ after normalizing the second entry. (Details are provided after the proof.) If $\nu = 0$, then the problem is counting perfect matchings over planar 4-regular graphs, so $\bar{f} \in \mathcal{M}$ and f is $\mathcal{M}$-transformable. Pl-Holant(f) is tractable but Holant(f) is #**P**-hard by Lemma 3.33. Otherwise, $\nu \neq 0$ and we are done by Lemma 3.38. $\square$

Here we give the details of the *orthogonal* transformation used in the proof of Theorem 3.39. We state the general case for symmetric signatures of arity $n \geqslant 1$, although we only used $n = 4$ in the proof above. Appendix D of [CHL12] has the case of $n = 3$.

We are given a symmetric signature $f = [f_0, \dots, f_n]$ such that $f_k = ck\alpha^{k-1} + d\alpha^k$, where $c \neq 0$, and $\alpha \neq \pm i$. Let $S = \begin{bmatrix} 1 & \frac{d-1}{n} \\ \alpha & c + \frac{d-1}{n}\alpha \end{bmatrix}$. Note that $\det(S) = c \neq 0$. Then f can be expressed as

$$f = S^{\otimes n}[1, 1, 0, \dots, 0],$$

where $[1, 1, 0, \dots, 0]$ should be understood as a dimension 2^n *column* vector, which has 1 in entries with Hamming weight at most one and 0 elsewhere. This identity can be verified by observing that

$$[1, 1, 0, \dots, 0] = [1, 0]^{\otimes n} + \frac{1}{(n-1)!} \operatorname{Sym}_n^{n-1}([1, 0]; [0, 1])$$

and we apply $S^{\otimes n}$ using properties of tensor product, $S^{\otimes n}[1, 0]^{\otimes n} = (S[1, 0])^{\otimes n}$, etc. We consider the value at index $0^{n-k}1^k$, which is the same as the value at any entry of weight k. By considering where the tensor product factor $[0, 1]$ is located among the n possible locations, we get

$$\alpha^k + k\left(c + \frac{d-1}{n}\alpha\right)\alpha^{k-1} + (n-k)\frac{d-1}{n}\alpha^k = ck\alpha^{k-1} + d\alpha^k.$$

Let $T = \frac{1}{\sqrt{1+\alpha^2}}\left[\begin{smallmatrix} 1 & \alpha \\ \alpha & -1 \end{smallmatrix}\right]$, then $T = T^{\mathsf{T}} = T^{-1} \in \mathbf{O}_2(\mathbb{C})$ is orthogonal, and $R = TS = \left[\begin{smallmatrix} u & w \\ 0 & z \end{smallmatrix}\right]$ is upper triangular, where $z, w \in \mathbb{C}$ and $u = \sqrt{1+\alpha^2} \neq 0$. However, $\det(R) = \det(T)\det(S) =$

$(-1)c \neq 0$, so we also have $z \neq 0$. It follows that

$$\begin{aligned}
T^{\otimes n} f &= (TS)^{\otimes n}[1, 1, 0, \ldots, 0] \\
&= R^{\otimes n}[1, 1, 0, \ldots, 0] \\
&= R^{\otimes n}\left([1, 0]^{\otimes n} + \frac{1}{(n-1)!}\operatorname{Sym}_n^{n-1}([1, 0]; [0, 1])\right) \\
&= [u, 0]^{\otimes n} + \frac{1}{(n-1)!}\operatorname{Sym}_n^{n-1}([u, 0]; [w, z]) \\
&= [u^n + nu^{n-1}w, u^{n-1}z, 0, \ldots, 0].
\end{aligned}$$

Since $u^{n-1}z \neq 0$, we can normalize to 1 the entry of Hamming weight 1 by a scalar multiplication. Thus, we have $[v, 1, 0, \ldots, 0]$ for some $v \in \mathbb{C}$.

3.7 Vanishing Signatures Revisited

In this section, we revisit vanishing signatures and show related hardness results. Basically, these hardness results imply that the two tractable cases illustrated in Lemma 3.15 and Lemma 3.16 are essentially maximal, in the sense that adding any other signature yields #**P**-hardness. In planar graphs some exceptional cases do exist, due to $\mathcal{M}$-transformable signatures. Results proven in this section will be useful for the Pl-Holant dichotomy.

Before proving anything, we first show a simple interpolation lemma, which will be handy in future.

Lemma 3.40. *Let $x \in \mathbb{C}$. If $x \neq 0$, then for any set $\mathcal{F}$ containing $[x, 1, 0]$, we have*

$$\operatorname{Holant}(\neq_2 \mid \mathcal{F} \cup \{[v, 1, 0]\}) \leqslant_T \operatorname{Holant}(\neq_2 \mid \mathcal{F}),$$

for any $v \in \mathbb{C}$.

Proof. Consider an instance Ω of $\operatorname{Holant}(\neq_2 \mid \mathcal{F} \cup \{[v, 1, 0]\})$. Suppose that $[v, 1, 0]$ appears n times in Ω. We stratify the assignments in Ω based on its assignments to $[v, 1, 0]$. We only need to consider assignments that give all $[v, 1, 0]$'s Hamming weights 0 and 1 since Hamming weight 2 contributes 0. If there are i many $[v, 1, 0]$'s having Hamming weight 0, then the rest

$n-i$ many have Hamming weight 1. Let c_i denote the summation of the product of evaluations of signatures other than $[\nu, 1, 0]$ in Ω over assignments which give i many $[\nu, 1, 0]$'s Hamming weight 0. We can rewrite the Holant on Ω as

$$\text{Holant}_\Omega = \sum_{i=0}^{n} \nu^i c_i.$$

We construct from Ω a sequence of instances Ω_s of Holant$(\mathcal{F})$ indexed by $s \geqslant 1$. We obtain Ω_s from Ω by replacing each occurrence of $[\nu, 1, 0]$ with a gadget g_s created from s copies of $[x, 1, 0]$, connected sequentially but with $(\neq_2) = [0, 1, 0]$ between each sequential pair. The signature of g_s is $[sx, 1, 0]$, which can be verified by the matrix product

$$\left(\begin{bmatrix} x & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}\right)^{s-1} \begin{bmatrix} x & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & x \\ 0 & 1 \end{bmatrix}^{s-1} \begin{bmatrix} x & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & (s-1)x \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} sx & 1 \\ 1 & 0 \end{bmatrix}.$$

The Holant on Ω_s is

$$\text{Holant}_{\Omega_s} = \sum_{i=0}^{n} (sx)^i c_i.$$

For $s \geqslant 1$, this gives a coefficient matrix that is Vandermonde. Since x is nonzero, sx is distinct for each s. Therefore, the Vandermonde system has full rank. We can solve for the unknowns c_i and obtain the value of Holant$_\Omega$. □

Now consider the mixing of vanishing signatures with unary and binary signatures. For unary signatures, they can be combined with $\mathcal{R}_2^\sigma$, and we show that they cannot be combined with any other vanishing signature.

Lemma 3.41. *Let* $f \in \mathcal{V}^\sigma$ *be a symmetric signature of arity* n *with* $\text{rd}^\sigma(f) = d \geqslant 2$ *where* $\sigma \in \{+, -\}$*. Suppose* $\nu = u^{\otimes m}$ *is a symmetric degenerate signature for some unary signature* u *and some integer* $m \geqslant 1$*. If* u *is not a multiple of* $[1, \sigma i]$*, then* Pl-Holant(f, ν) *is* #**P**-*hard.*

Proof. We consider $\sigma = +$ since the other case is similar. Since $f \in \mathcal{V}^+$, we have $n > 2d \geqslant 4$. Under a holographic transformation by Z, we have

$$\text{Pl-Holant}(f, \nu) \equiv \text{Pl-Holant}\left(\neq_2 \mid \hat{f}, [a, b]^{\otimes m}\right),$$

where $\overline{f} = \left(Z^{-1}\right)^{\otimes n} f$ and $[a,b]^{\otimes m} = \left(Z^{-1}\right)^{\otimes m} \nu$ with $b \neq 0$ since u is not a multiple of $[1,i]$. Moreover, $\overline{f} = [\overline{f}_0, \overline{f}_1, \ldots, \overline{f}_d, 0, \ldots, 0]$ with $\overline{f}_d \neq 0$ by Lemma 3.14.

We get $\overline{f'} = [\overline{f}_{d-2}, \overline{f}_{d-1}, \overline{f}_d, 0, \ldots, 0]$ of arity $n-2d+4$ by $d-2$ self-loops via $\neq_2$ on $\overline{f}$. This is on the right side. With two more self-loops, we get $[1,0]^{\otimes n-2d}$, also on the right.

We claim that we can use $[1,0]^{\otimes n-2d}$ and $[a,b]^{\otimes m}$ to create $[a,b]^{\otimes n-2d}$. Let $t = \gcd(m, n-2d)$. If $n-2d > m$, then we connect $[a,b]^{\otimes m}$ to $[1,0]^{\otimes n-2d}$ via $\neq_2$ to get $[1,0]^{\otimes n-2d-m}$ up to a nonzero factor $b \neq 0$. We repeat this process until we get a tensor power $[1,0]^{\otimes \ell}$ for some $\ell \leqslant m$. We can do a similar construction if $m > n-2d$. Repeat this process, which is a subtractive Euclidean algorithm. Halt upon getting both $[1,0]^{\otimes t}$ and $[a,b]^{\otimes t}$. Then we combine $\frac{n-2d}{t}$ copies of $[a,b]^{\otimes t}$ to get $[a,b]^{\otimes n-2d}$.

Now connecting $[a,b]^{\otimes n-2d}$ back to $\overline{f'}$ via $\neq_2$, gives $\overline{f''} = [\overline{f''}_0, \overline{f''}_1, \overline{f''}_2, 0, 0]$ of arity 4. Moreover, $\overline{f''}_2 = b^{n-2d}\overline{f}_d \neq 0$. Notice that Pl-Holant$(\neq_2 \mid [\overline{f''}_0, \overline{f''}_1, \overline{f''}_2, 0, 0]) \equiv$ Pl-Holant$(\neq_2 \mid [0,0,1,0,0])$, the Eulerian Orientation problem over planar 4-regular graphs, which is #**P**-hard by Theorem 3.21. Thus, Pl-Holant(f, ν) is #**P**-hard. □

Next come binary signatures. We first do some preparation.

Lemma 3.42. *Let* $c, t \in \mathbb{C}$*. If* $ct \neq 0$*, then* Pl-Holant $(\neq_2 \mid [t,1,0,0,0], [c,0,1])$ *is #**P**-hard. Moreover,* Holant $(\neq_2 \mid [0,1,0,0,0], [c,0,1])$ *is #**P**-hard.*

Proof. By connecting two copies of $\neq_2$ to either side of $[c,0,1]$, we get the signature $[1,0,c]$ on the left. Clearly Pl-Holant $([1,0,c] \mid [t,1,0,0,0]) \leqslant_T$ Pl-Holant $(\neq_2 \mid [t,1,0,0,0], [c,0,1])$. Then under a holographic transformation by T^{-1}, where $T = \left[\begin{smallmatrix} 1 & 0 \\ 0 & \sqrt{c} \end{smallmatrix}\right]$, we have

$$\begin{aligned}
\text{Pl-Holant}\,([1,0,c] \mid [t,1,0,0,0]) &\equiv \text{Pl-Holant}\left([1,0,c](T^{-1})^{\otimes 2} \mid T^{\otimes 4}[t,1,0,0,0]\right) \\
&\equiv \text{Pl-Holant}\left([1,0,1] \mid [t,\sqrt{c},0,0,0]\right) \\
&\equiv \text{Pl-Holant}([t,\sqrt{c},0,0,0]).
\end{aligned}$$

The last problem is #**P**-hard by Lemma 3.38 after dividing by $\sqrt{c}$. Clearly above reductions still hold without the planar restriction. The second statement follows from $t = 0$ and Lemma 3.33. □

If $t = 0$, the above problem is tractable in planar graphs. In general, we need to take care of some planar tractable cases. Recall that $Z = \frac{1}{\sqrt{2}} \left[\begin{smallmatrix} 1 & 1 \\ i & -i \end{smallmatrix}\right]$ Define

$$\mathcal{M}_4^+ := \left\{ f \mid \text{arity}(f) = n, f = Z^{\otimes n} \textsc{ExactOne}_n, n \in \mathbb{N} \right\}$$

and

$$\mathcal{M}_4^- := \left\{ f \mid \text{arity}(f) = n, f = Z^{\otimes n} \textsc{AllButOne}_n, n \in \mathbb{N} \right\}.$$

It is easy to see that $\mathcal{M}_4^- = \left\{ \left(Z \left[\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right] \right)^{\otimes n} \textsc{ExactOne}_n, n \in \mathbb{N} \right\} = \left[\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right] \mathcal{M}_4^+$. Let $\mathcal{M}_4 = \mathcal{M}_4^+ \cup \mathcal{M}_4^-$. The reason of the name $\mathcal{M}_4^\pm$ will become clear in Chapter 5, in particular, Definition 5.16. The set $\mathcal{M}_4$ contains all $\mathcal{M}$-transformable signatures in $\mathcal{V}^\pm$.

Also note that if $f \in \mathcal{V}^\pm$ is a symmetric non-degenerate signature, then f has arity at least 3. This is because a unary signature is degenerate, and if a binary symmetric signature f is vanishing, then its vanishing degree is greater than 1, hence at least 2, and therefore f is also degenerate. Nevertheless, we explicitly state this condition $\text{arity}(f) \geqslant 3$ in the following lemmas.

In light of Lemma 3.15, we can allow non-vanishing binary signatures from $\mathcal{R}_2^\sigma$ but not others.

Lemma 3.43. *Let $f \in \mathcal{V}^\sigma$ be a symmetric non-degenerate signature of arity $n \geqslant 3$ for some $\sigma \in \{+, -\}$. Let $h \notin \mathcal{R}_2^\sigma$ be a non-degenerate binary signature. Then* Holant(f, h) *is #**P**-hard. Moreover, if $f \notin \mathcal{M}_4^\sigma$, then* Pl-Holant$(f, h)$ *is #**P**-hard.*

Proof. We consider $\sigma = +$ since the other case is similar. Under a Z transformation,

$$\text{Pl-Holant}(f, h) \equiv \text{Pl-Holant}\left(\neq_2 \mid \overline{f}, \overline{h} \right),$$

$$\text{Holant}(f, h) \equiv \text{Holant}\left(\neq_2 \mid \overline{f}, \overline{h} \right),$$

where $\overline{f} = \left(Z^{-1}\right)^{\otimes n} f$ and $\overline{h} = \left(Z^{-1}\right)^{\otimes 2} h$. Since $h \notin \mathcal{R}_2^+$, we may assume that $\overline{h} = [a, b, 1]$ by Lemma 3.14 with a nonzero entry $\overline{h}_2$. Moreover since h is non-degenerate, so is $\overline{h}$, and $b^2 \neq a$.

We prove the lemma by induction on the arity of f (or equivalently $\overline{f}$). There are two base cases, $n = 3$ and $n = 4$. However, the arity 3 case is easily reduced to the arity 4 case. We show this first, and then show that the lemma holds in the arity 4 case.

Assume $n = 3$. Since $f \in \mathcal{V}^+$, we have $\bar{f} = [t, 1, 0, 0]$ for some $t \neq 0$, by Lemma 3.14 and $f \notin \mathcal{M}_4^+$. Consider the gadget in Figure 3.11. We assign $\bar{f}$ to the circle vertices and $\neq_2$ to the square vertex. Let $\bar{f}'$ be the signature of the resulting gadget. The signature $\bar{f}'$ may not seem symmetric by construction, but it is not hard to verify that indeed $\bar{f}' = [2t, 1, 0, 0, 0]$. The crucial observation is that it takes the same value 0 on inputs 1010 and 1100, where bits are ordered counterclockwise, starting from an arbitrary edge. This finishes our reduction to $n = 4$.

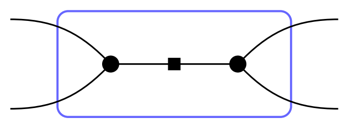


Figure 3.11: Circle vertices are assigned $[t, 1, 0, 0]$ and the square vertex is assigned $\neq_2$.

Now we consider the base case of $n = 4$. Since $f \in \mathcal{V}^+$, we have $\text{vd}^+(f) > 2$ and $\text{rd}^+(f) < 2$. As f is not degenerate, $\text{rd}^+(f) \notin \{-1, 0\}$. It implies that $\text{rd}^+(f) = 1$ and by Lemma 3.14, $\bar{f} = [t, 1, 0, 0, 0]$.

Our next goal is to show that we can realize a signature of the form $[c, 0, 1]$ with $c \neq 0$, namely,

$$\text{Pl-Holant}\left(\neq_2 \mid [t, 1, 0, 0, 0], [c, 0, 1]\right) \leqslant_T \text{Pl-Holant}(f, h);$$

$$\text{Holant}\left(\neq_2 \mid [t, 1, 0, 0, 0], [c, 0, 1]\right) \leqslant_T \text{Holant}(f, h).$$

This finishes our base case because $\text{Holant}\left(\neq_2 \mid [t, 1, 0, 0, 0], [c, 0, 1]\right)$ is #**P**-hard by Lemma 3.42. Moreover, if $f \notin \mathcal{M}_4^+$, then $t \neq 0$. By Lemma 3.42, $\text{Pl-Holant}\left(\neq_2 \mid [t, 1, 0, 0, 0], [c, 0, 1]\right)$ is also #**P**-hard.

If $b = 0$, then $\bar{h}$ is what we want since in this case $a = a - b^2 \neq 0$.

Otherwise $b \neq 0$. By connecting $\bar{h}$ to $\bar{f}$ via $\neq_2$, we get $[t + 2b, 1, 0]$. If $t \neq -2b$, then by Lemma 3.40, we can interpolate any binary signature of the form $[\nu, 1, 0]$. Otherwise $t = -2b$. Then we connect two copies of $\bar{h}$ via $\neq_2$, and get $\bar{h}' = [2ab, a + b^2, 2b]$. By connecting this $\bar{h}'$ to $\bar{f}$ via $\neq_2$, we get $[2(a - b^2), 2b, 0]$, for $t = -2b$. Since $a \neq b^2$ and $b \neq 0$, we can once again interpolate any $[\nu, 1, 0]$ by Lemma 3.40.

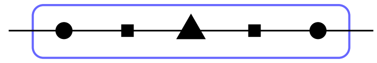


Figure 3.12: A sequence of binary gadgets that forms another binary gadget. The circles are assigned $[v, 1, 0]$, squares are $\neq_2$, and the triangle is $[a, b, 1]$.

Hence, we have the signature $[v, 1, 0]$, where $v \in \mathbb{C}$ is for us to choose. We construct the gadget in Figure 3.12 with the circles assigned $[v, 1, 0]$, the squares assigned $\neq_2$, and the triangle assigned $[a, b, 1]$. The resulting gadget has signature $[a+2bv+v^2, b+v, 1]$, which can be verified by the matrix product

$$\begin{bmatrix} v & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} a & b \\ b & 1 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} v & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} a+2bv+v^2 & b+v \\ b+v & 1 \end{bmatrix}.$$

Setting $v = -b$, we get $[c, 0, 1]$, where $c = a - b^2 \neq 0$. This finishes the base case of $n = 3, 4$.

Now we do the induction step. Assume $n \geqslant 5$. Since f is non-degenerate, $\text{rd}^+(f) \geqslant 1$. If $\text{rd}^+(f) = 1$, then $\bar{f} = [t, 1, 0, \ldots, 0]$ for some $t \neq 0$. We connect $\bar{h}$ to $\bar{f}$ via $\neq_2$, getting $[t + 2b, 1, 0, \ldots, 0]$ of arity $n - 2 \geqslant 3$. If $t + 2b \neq 0$, then we are done by induction hypothesis. Otherwise $t = -2b$, and we connect two $\bar{h}$ together via $\neq_2$. The signature is $\bar{h}' := [2ab, b^2+a, 2b]$. Connect $\bar{h}'$ to $\bar{f}$ via $\neq_2$. We get $[-4b^2 + 2(b^2 + a), 2b, 0, \ldots, 0] = [2(a - b^2), 2b, 0, \ldots, 0]$. If $b = 0$, then $t = 0$. Contradiction. Hence $b \neq 0$, and $a - b^2 \neq 0$ for b is not degenerate. Then we can apply induction hypothesis on $[2(a - b^2), 2b, 0, \ldots, 0]$.

The case left is that $\text{rd}^+(f) = d \geqslant 2$. Then $\bar{f} = [\bar{f}_0, \bar{f}_1, \ldots, \bar{f}_d, 0, \ldots, 0]$ with $\bar{f}_d \neq 0$ by Lemma 3.14. We do a self-loop of $\bar{f}$ via $\neq_2$, getting $\bar{f}'' := [\bar{f}_1, \ldots, \bar{f}_d, 0, \ldots, 0]$ of arity $n - 2 \geqslant 3$. Since $d \geqslant 2$, $\bar{f}''$ is non-degenerate and $f'' = Z^{\otimes(n-2)}\bar{f}'' \in \mathcal{V}^+$. Apply the induction hypothesis and we are done here for $\text{Holant}(f, h)$.

For the planar case, if $f'' \notin \mathcal{M}_4^+$, then apply the induction hypothesis and we are done. Otherwise $d = 2$ and we may assume $\bar{f} = [\bar{f}_0, 0, 1, 0, \ldots, 0]$ since $\bar{f}_2 \neq 0$.

In this case, we connect $\bar{h}$ to $\bar{f}$ via $\neq_2$, getting $\bar{f}''' := [a + \bar{f}_0, 2b, 1, 0, \ldots, 0]$ of arity $n - 2 \geqslant 3$. If $n \geqslant 7$, then we can apply the induction hypothesis. If $n = 6$, then $\bar{f}''' = [a + \bar{f}_0, 2b, 1, 0, 0]$ of arity 4. Notice that $\text{Pl-Holant}\left(\neq_2 \mid [a + \bar{f}_0, 2b, 1, 0, 0]\right)$ is equivalent to $\text{Pl-Holant}\left(\neq_2 \mid [0, 0, 1, 0, 0]\right)$, which is PL-4REG-#EO. Then $\text{Pl-Holant}\left(\neq_2 \mid \bar{f}'''\right)$ is #**P**-hard by Theorem 3.21.

The only case left now is when $n = 5$ and $\bar{f} = [\bar{f}_0, 0, 1, 0, 0, 0]$. We do two self-loops on $\bar{f}$ via $\neq_2$ to get $[1, 0]$. Then connect $[1, 0]$ to $\bar{h}$ via $\neq_2$ and get $[b, 1]$. At last, connect $[b, 1]$ to $\bar{f}$ via $\neq_2$, resulting in $[\bar{f}_0, b, 1, 0, 0]$. Similar to the case above, Pl-Holant $\left(\neq_2 | [\bar{f}_0, b, 1, 0, 0]\right)$ is equivalent to PL-4REG-#EO, and is #**P**-hard by Theorem 3.21. $\square$

If $f \in \mathcal{M}_4^{\pm}$, there is an additional planar tractable case for the binary signature.

Lemma 3.44. *Let* $f \in \mathcal{M}_4^{\sigma}$ *be a symmetric non-degenerate signature with* $\sigma \in \{+, -\}$ *of arity* $k \geqslant 3$*. Suppose* h *is a non-degenerate binary signature such that* $h \notin \mathcal{R}_2^{\sigma}$ *and* h *is not a multiple of* $Z^{\otimes 2}[a, 0, 1]$ *for any* $a \neq 0$*. Then* Pl-Holant(f, h) *is #**P**-hard.*

Proof. We assume $f \in \mathcal{M}_4^{+}$ since the other case is similar. Suppose $h = Z^{\otimes 2}[a, b, c]$ for some $a, b, c \in \mathbb{C}$. Since $h \notin \mathcal{R}_2^{+}$, we have $c \neq 0$, so we assume $c = 1$. Moreover $b \neq 0$. This is because, if $b = 0$ then either h is degenerate or is a multiple of $Z^{\otimes 2}[a, 0, 1]$ for some $a \neq 0$. Either case violates our assumption. Then under a holographic transformation by Z, the problem becomes Pl-Holant $(\neq_2 \mid \text{EXACTONE}_k, [a, b, 1])$. If we connect two copies of EXACTONE$_k$ via $\neq_2$, we get EXACTONE$_{2k-2}$. Hence we may assume that $k \geqslant 5$. Then we connect $[a, b, 1]$ to EXACTONE$_k$ via $\neq_2$, and get $[2b, 1, 0, \ldots, 0]$ of arity $k - 2 \geqslant 3$. Since $b \neq 0$, Pl-Holant(f, h) is #**P**-hard by Lemma 3.43. $\square$

Next we consider mixing signatures from $\mathcal{V}^{+}$ and $\mathcal{V}^{-}$. In general graphs, it is always #**P**-hard. For planar graphs, there is a tractable case when one signature is in $\mathcal{M}_4^{+}$ and the other is in $\mathcal{M}_4^{-}$, since as a set they are $\mathcal{M}$-transformable. We will first show that every other case is #**P**-hard, even for planar graphs. We deal with the case when signatures from both $\mathcal{M}_4^{+}$ and $\mathcal{M}_4^{-}$ show up immediately after.

Lemma 3.45. *Let* $f \in \mathcal{V}^{+}$ *and* $g \in \mathcal{V}^{-}$ *be two symmetric non-degenerate signatures of arities* $\geqslant 3$*. If* $f \notin \mathcal{M}_4^{+}$ *or* $g \notin \mathcal{M}_4^{-}$ *then* Pl-Holant(f, g) *is #**P**-hard.*

Proof. Suppose $\text{rd}^{+}(f) = d$, $\text{rd}^{-}(g) = d'$, $\text{arity}(f) = n$ and $\text{arity}(g) = n'$, then $2d < n$ and $2d' < n'$. Under a holographic transformation by $Z = \left[\begin{smallmatrix} 1 & 1 \\ i & -i \end{smallmatrix}\right]$, we have that

$$\text{Pl-Holant}\,(=_2 \mid f, g) \equiv_T \text{Pl-Holant}\left(\neq_2 \mid \bar{f}, \bar{g}\right),$$

where $\overline{f} := (Z^{-1})^{\otimes n} f = [\overline{f}_0, \ldots, \overline{f}_d, 0, \ldots, 0]$ and $\overline{g} := (Z^{-1})^{\otimes n'} g = [0, \ldots, 0, \overline{g}_{d'}, \ldots, \overline{g}_0]$ due to Lemma 3.14. Moreover $\overline{f}_d \neq 0$ and $\overline{g}_{d'} \neq 0$.

If $d \geqslant 2$, we can do d' many self-loops of $\neq_2$ on $\overline{g}$, getting $\overline{g}' := [0, \ldots, 0, \overline{g}_{d'}]$ of arity $n' - 2d' \geqslant 1$. Thus $g' := Z^{\otimes (n'-2d')} \overline{g}' = [1, -i]^{\otimes (n'-2d')}$ up to a nonzero constant. We apply Lemma 3.41 to derive that Pl-Holant(f, g) is #**P**-hard. If $d' \geqslant 2$, we can similarly get $[1, i]^{\otimes (n-2d)}$ and apply Lemma 3.41. Thus we can assume that $d = d' = 1$.

So up to nonzero constants, we have $\overline{f} = [a, 1, 0, \ldots, 0]$ and $\overline{g} = [0, \ldots, 0, 1, b]$ for some $a, b \in \mathbb{C}$. We can assume that $f \notin \mathcal{M}_4^+$ and $a \neq 0$. The case of $b \neq 0$ is similar. We show that it is always possible to get two such signatures of the same arity $\min\{n, n'\}$. Suppose $n > n'$. We form a loop from $\overline{f}$ via $\neq_2$. It is easy to see that this signature is the degenerate signature $2[1, 0]^{\otimes (n-2)}$. Similarly, we can form a loop from $\overline{g}$ and can get $2[0, 1]^{\otimes (n'-2)}$. Thus we have both $[1, 0]^{\otimes (n-2)}$ and $[0, 1]^{\otimes (n'-2)}$. We can connect all $n' - 2$ edges of the second to the first, connected by $\neq_2$. This gives $[1, 0]^{\otimes (n-n')}$. We can continue subtracting the smaller arity from the larger one. We continue this process in a subtractive version of the Euclidean algorithm, and end up with both $[1, 0]^{\otimes t}$ and $[0, 1]^{\otimes t}$, where $t = \gcd(n - 2, n' - 2) = \gcd(n - n', n' - 2)$. In particular, $t \mid n - n'$ and by taking $\frac{n-n'}{t}$ copies of $[0, 1]^{\otimes t}$, we can get $[0, 1]^{\otimes (n-n')}$. Connecting this back to $\overline{f}$ via $\neq_2$, we get a symmetric signature of arity n' consisting of the first $n' + 1$ entries of $\overline{f}$. A similar proof works when $n' > n$.

Thus we may assume $n = n'$. Connecting $[0, 1]^{\otimes (n-2)}$ to $\overline{f} = [a, 1, 0, \ldots, 0]$ via $\neq_2$ we get $\overline{h} = [a, 1, 0]$. Recall that $a \neq 0$. Translating this back by Z, we have a binary signature $h \notin \mathcal{R}_2^-$ and h is not a multiple of $Z^{\otimes 2}[c, 0, 1]$ for any $c \neq 0$. Since $g \in \mathcal{V}^-$, by Lemma 3.43 or Lemma 3.44, Pl-Holant(g, h) is #**P**-hard. Hence Pl-Holant(f, g) is also #**P**-hard. □

When signatures in both $\mathcal{M}_4^+$ and $\mathcal{M}_4^-$ appear, we show that the only degenerate signatures that mix must also be vanishing.

Lemma 3.46. *Let* $f \in \mathcal{M}_4^+$ *and* $g \in \mathcal{M}_4^-$ *be two non-degenerate signatures of arity* $n \geqslant 3$ *and* $m \geqslant 3$*, respectively. Let* $v = u^{\otimes \ell}$ *be a degenerate signature for some unary signature* u *and some integer* $\ell \geqslant 1$*. If* u *is not a multiple of* $[1, \pm i]$*, then* Pl-Holant(f, g, v) *is* #**P**-*hard.*

Proof. Under a holographic transformation by Z, we have that

$$\text{Pl-Holant}(f, g, v) \equiv \text{Pl-Holant}\left(\neq_2 \mid \textsc{ExactOne}_n, \textsc{AllButOne}_m, [a, b]^{\otimes \ell}\right),$$

where $ab \neq 0$. Notice that v is transformed to $(Z^{-1}u)^{\otimes \ell} = [a, b]^{\otimes \ell}$. We have $ab \neq 0$ since u is not a multiple of $[1, \pm i]$. First we get $[1, 0]^{\otimes n-2}$ by a self-loop via $\neq_2$ on $\textsc{ExactOne}_n$. By the same subtractive Euclidean argument as in Lemma 3.41, we can realize $[a, b]^{\otimes n-2}$ by $[1, 0]^{\otimes n-2}$ and $[a, b]^{\otimes \ell}$. Connecting $[a, b]^{\otimes n-2}$ to $\textsc{ExactOne}_n$ via $\neq_2$ we get a binary signature $h = [(n-2)ab^{n-3}, b^{n-2}, 0]$. After transforming back, we have that

$$\text{Pl-Holant}(g, Z^{\otimes 2}h) \leqslant_T \text{Pl-Holant}(f, g, v).$$

However $Z^{\otimes 2}h \notin \mathcal{R}_2^-$ by Lemma 3.14 and it is not a multiple of $Z^{\otimes 2}[c, 0, 1]$ for any $c \neq 0$. Apply Lemma 3.44, where $(g, Z^{\otimes 2}h)$ plays the role of "(f, h)" in Lemma 3.44 and $\sigma = -$, we conclude that Pl-Holant(f, g, v) is #**P**-hard. □

At last, we show the planar tractable case is #**P**-hard in general graphs. The reduction uses a non-planar construction.

Lemma 3.47. *Let $f \in \mathcal{M}_4^+$ and $g \in \mathcal{M}_4^-$ be two signatures of arity $n \geqslant 3$ and $m \geqslant 3$, respectively. Then* Holant(f, g) *is #**P**-hard.*

Proof. Use the subtractive Euclidean argument in Lemma 3.45, we can realize signatures $f' \in \mathcal{M}_4^+$ and $g \in \mathcal{M}_4^-$ of the same arity $\min\{n, m\}$. Hence we may assume that $m = n$. We do a $Z = \left[\begin{smallmatrix} 1 & 1 \\ i & -i \end{smallmatrix}\right]$ transformation:

$$\text{Holant}(f, g) \equiv \text{Holant}\left(\neq_2 \mid \bar{f}, \bar{g}\right),$$

where $\bar{f} = \textsc{ExactOne}_n$ and $\bar{g} = \left[\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right]^{\otimes n} \textsc{ExactOne}_n = [0, \ldots, 0, 1, 0]$ since $f \in \mathcal{M}_4^+$ and $g \in \mathcal{M}_4^-$.

Our goal is to obtain a signature that satisfies the condition of Corollary 3.30.

The gadget in Figure 3.13a, with $\bar{f}$ assigned to the circle, $\bar{g}$ assigned to the triangle, and $\neq_2$

assigned to squares, has signature h with a signature matrix

$$M_h = \begin{bmatrix} 0 & 0 & 0 & \nu \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix},$$

where $\nu = n - 2$ is positive since $n \geqslant 3$. Although this signature matrix is redundant, its compressed form is singular. Rotating this gadget 90° clockwise and 90° counterclockwise, (recall Figure 3.2) we get signatures h' and h'', respectively, with signature matrices

$$M_{h'} = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & \nu & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix} \qquad \text{and} \qquad M_{h''} = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & \nu & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}.$$

Then we build the gadget in Figure 3.13b, with h' assigned to the circle, h'' assigned to the triangle, and $\neq_2$ assigned to squares. The resulting signature r has a signature matrix

$$M_r = M_{h'} \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix} M_{h''} = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & \nu & \nu^2 + 1 & 0 \\ 0 & 1 & \nu & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}.$$

Note that the effect of the $\neq_2$ signatures is to reverse all four rows of $M_{h''}$ before multiplying it to the right of $M_{h'}$. Although this signature matrix is not redundant, every entry of Hamming weight 2 is nonzero since ν is positive.

(a) The circle is assigned $\bar{f}$, the triangle is assigned $\bar{g}$, and the squares are assigned $\neq_2$.

(b) The circle is assigned h', the triangle is assigned h'', and the squares are assigned $\neq_2$.

Figure 3.13: Gadget constructions used to obtain a hard and redundant arity 4 signature.

Let r' be the signature with its matrix

$$M_{r'} = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & & & 0 \\ 0 & & T & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}, \tag{3.10}$$

where $T = P \left[\begin{smallmatrix} t & 0 \\ 0 & t^{-1} \end{smallmatrix}\right] P^{-1}$, $P = \left[\begin{smallmatrix} 1 & 1 \\ p^{+} & p^{-} \end{smallmatrix}\right]$, and $p^{\pm} = (\nu \pm \sqrt{\nu^2 + 4})/2$, for some $t \in \mathbb{C}$ and $t \neq 0$. We claim that we can use r to interpolate r', for any $t \neq 0$. We use the recursive construction in Figure 3.14.

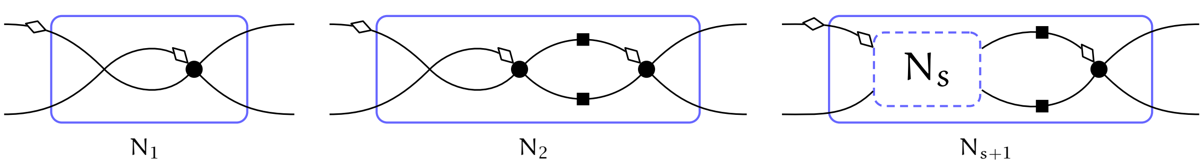


Figure 3.14: Recursive construction to interpolate a signature r' that is only a rotation away from having a redundant signature matrix and nonsingular compressed matrix. The circles are assigned r and the squares are assigned $\neq_2$.

Consider an instance Ω of Holant $(\neq_2 \mid \mathcal{F} \cup \{r'\})$ with $r \in \mathcal{F}$. Suppose that r' appears n times in Ω. We construct from Ω a sequence of instances Ω_s of Holant $(\neq_2 \mid \mathcal{F})$ indexed by $s \geqslant 1$. We obtain Ω_s from Ω by replacing each occurrence of r' with the gadget N_s in Figure 3.14 with r assigned to circles and $\neq_2$ assigned to squares. In Ω_s, the edge corresponding to the ith significant index bit of N_s connects to the same location as the edge corresponding to the ith significant index bit of r' in Ω.

The signature matrix of N_s is the sth power of the matrix obtained from M_r after reversing all rows, and then switching the first and last rows of the final product, namely,

$$\begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & \nu & 0 \\ 0 & \nu & \nu^2+1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}^s = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 1 & \nu & 0 \\ 0 & \nu & \nu^2+1 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & \nu & 0 \\ 0 & \nu & \nu^2+1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}^{s-1}.$$

The twist of the two input edges on the left side for the first copy of M_r switches the middle two rows, which is equivalent to a total reversal of all rows, followed by the switching of the first and last rows. The total reversals of rows for all subsequent $s-1$ copies of M_r are due to the presence of $\neq_2$ signatures.

After such reversals of rows, it is clear that the matrix is a direct sum of block matrices indexed by $\{00,11\} \times \{00,11\}$ and $\{01,10\} \times \{10,01\}$. Furthermore, in the final product, the block indexed by $\{00,11\} \times \{00,11\}$ is $\left[\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right]$. Thus in the gadget N_s, the only entries of M_{N_s} that vary with s are the four entries in the middle. These middle four entries of M_{N_s} form the 2-by-2 matrix $\left[\begin{smallmatrix} 1 & \nu \\ \nu & \nu^2+1 \end{smallmatrix}\right]^s$. Since $\left[\begin{smallmatrix} 1 & \nu \\ \nu & \nu^2+1 \end{smallmatrix}\right] = P \left[\begin{smallmatrix} \lambda_+ & 0 \\ 0 & \lambda_- \end{smallmatrix}\right] P^{-1}$, where $\lambda_\pm = (\nu^2 + 2 \pm \nu\sqrt{\nu^2+4})/2$ are the eigenvalues, we have that

$$\begin{bmatrix} 1 & \nu \\ \nu & \nu^2+1 \end{bmatrix}^s = P \begin{bmatrix} \lambda_+^s & 0 \\ 0 & \lambda_-^s \end{bmatrix} P^{-1}.$$

The determinant of $\left[\begin{smallmatrix} 1 & \nu \\ \nu & \nu^2+1 \end{smallmatrix}\right]$ is $\lambda_+\lambda_- = 1$, so the eigenvalues are nonzero. Since ν is positive, the ratio of the eigenvalues λ_+/λ_- is not a root of unity, so neither λ_+ nor λ_- is a root of unity.

Now we determine the relationship between Holant_Ω and Holant_{Ω_s}. We can view our construction of Ω_s as first replacing $M_{r'}$ with

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \multicolumn{2}{c}{\multirow{2}{*}{P}} & 0 \\ 0 & & & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & t & 0 & 0 \\ 0 & 0 & t^{-1} & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \multicolumn{2}{c}{\multirow{2}{*}{P^{-1}}} & 0 \\ 0 & & & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

which does not change the Holant value, and then replacing the new signature matrix in the middle with the signature matrix

$$\begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & \lambda_+^s & 0 & 0 \\ 0 & 0 & \lambda_-^s & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}.$$

We stratify the assignments in Ω_s based on the assignments to the n occurrences of the signature matrix

$$\begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & t & 0 & 0 \\ 0 & 0 & t^{-1} & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}. \tag{3.11}$$

The inputs to this matrix are from $\{0,1\}^2 \times \{0,1\}^2$, which correspond to the four input bits. Recall the way rows and columns of a signature matrix are ordered from Definition 3.17. Thus, e.g., the entry t corresponds to the cyclic input bit pattern 0110 in counterclockwise order. We only need to consider the assignments that assign

- i many times the bit pattern 0110,
- j many times the bit pattern 1001, and
- k many times the bit patterns 0011 or 1100,

since any other assignment contributes a factor of 0. Let c_{ijk} be the sum over all such assignments of the products of evaluations of all signatures in Ω_s except for the signature in (3.11). Then,

$$\text{Holant}_\Omega = \sum_{i+j+k=n} t^{i-j} c_{ijk}$$

and the value of the Holant on Ω_s, for $s \geqslant 1$, is

$$\text{Holant}_{\Omega_s} = \sum_{i+j+k=n} \lambda_+^{si} \lambda_-^{sj} c_{ijk} = \sum_{i+j+k=n} \lambda_+^{s(i-j)} c_{ijk}.$$

This Vandermonde system does not have full rank. However, we can define for $-n \leqslant \ell \leqslant n$,

$$c'_\ell = \sum_{\substack{i-j=\ell \\ i+j+k=n}} c_{ijk}.$$

Then the Holant of Ω is

$$\text{Holant}_\Omega = \sum_{-n \leqslant \ell \leqslant n} t^\ell c'_\ell$$

and the Holant of Ω_s is

$$\text{Holant}_{\Omega_s} = \sum_{-n \leqslant \ell \leqslant n} \lambda_+^{s\ell} c'_\ell.$$

Now this Vandermonde has full rank because λ_+ is neither 0 nor a root of unity. Therefore, we can solve for the unknowns c'_ℓ and obtain the value of Holant_Ω. This completes our claim that we can interpolate the signature r' in (3.10), for any nonzero $t \in \mathbb{C}$.

Let $t = (\sqrt{\nu^2+8} + \sqrt{\nu^2+4})/2$ so $t^{-1} = (\sqrt{\nu^2+8} - \sqrt{\nu^2+4})/2$. Let $a = (\sqrt{\nu^2+8} - \nu)/2$ and $b = (\sqrt{\nu^2+8} + \nu)/2$, so $ab = 2 \neq 0$. One can verify that

$$P \begin{bmatrix} t & 0 \\ 0 & t^{-1} \end{bmatrix} P^{-1} = \begin{bmatrix} a & 1 \\ 1 & b \end{bmatrix}.$$

Thus, the signature matrix for r' is

$$M_{r'} = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & a & 1 & 0 \\ 0 & 1 & b & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}.$$

After a counterclockwise rotation of 90° on the edges of r', we have a signature r'' with a redundant signature matrix

$$M_{r''} = \begin{bmatrix} 0 & 0 & 0 & a \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ b & 0 & 0 & 0 \end{bmatrix}.$$

Its compressed signature matrix

$$\widetilde{M_{r''}} = \begin{bmatrix} 0 & 0 & a \\ 0 & 2 & 0 \\ b & 0 & 0 \end{bmatrix}$$

is nonsingular. After a holographic transformation by Z^{-1}, where $Z = \frac{1}{\sqrt{2}} \left[\begin{smallmatrix} 1 & 1 \\ i & -i \end{smallmatrix}\right]$, the binary disequality $(\neq_2) = [0, 1, 0]$ is transformed to the binary equality $(=_2) = [1, 0, 1]$. Thus the problem Holant $([0, 1, 0] \mid r'')$ is transformed to Holant $\left(=_2 \mid Z^{\otimes 4} r''\right)$, which is the same as Holant$(Z^{\otimes 4} r'')$. We conclude that this Holant problem is #P-hard by Corollary 3.30. □

To end this section, we show two hardness results related to vanishing signatures. They will be useful in the proof of the Pl-Holant dichotomy. The first is to show that $\mathcal{M}_4$ signatures cannot be combined with $=_n$ in the Z basis.

Lemma 3.48. *Let* $f \in \mathcal{V} \setminus \mathcal{M}_4$ *and* $g = Z^{\otimes n} (=_n)$ *be two non-degenerate signatures with arities* m *and* n *respectively. If* $m, n \geqslant 3$*, then* Pl-Holant(f, g) *is #***P***-hard.*

Proof. We may assume that $f \in \mathcal{V}^+ \setminus \mathcal{M}_4$. The case of $f \in \mathcal{V}^-$ is similar. By Corollary 3.9, we have $\text{rd}^+(f) = d < \frac{m}{2}$. Under a holographic transformation by Z, we have

$$\begin{aligned} \text{Pl-Holant}\,(=_2 \mid f, g) &\equiv \text{Pl-Holant}\left([1, 0, 1] Z^{\otimes 2} \mid Z^{-1}\{f, g\}\right) \\ &\equiv \text{Pl-Holant}\left(\neq_2 \mid \bar{f}, =_n\right), \end{aligned}$$

where $\bar{f} = (Z^{-1})^{\otimes m} f$. By Lemma 3.14, the support of $\bar{f}$ is on the first d entries. As $f \notin \mathcal{M}_4$, we have either $d = 1$ and $\bar{f} = [\bar{f}_0, 1, 0, \ldots, 0]$ with $\bar{f}_0 \neq 0$, or $d \geqslant 2$ and $\bar{f} = [\bar{f}_0, \bar{f}_1, \ldots, \bar{f}_{d-1}, 1, 0, \ldots, 0]$ with $\bar{f}_{d-1} \neq 0$ (and up to a nonzero scalar in either case).

In the first case, a self-loop on $\bar{f}$ via $\neq_2$ gives $[1, 0]^{\otimes m-2}$ on the right side. Let $r = \gcd(n, m-2)$, and let ℓ_1, ℓ_2 be two positive integers such that $\ell_1 n - \ell_2(m-2) = r$. We connect ℓ_1 copies of $=_n$ with ℓ_2 copies of $[1,0]^{\otimes m-2}$ via $\neq_2$'s to get $[0,1]^{\otimes r}$. Since $r \mid m-2$, we can also realize $[0,1]^{\otimes m-2}$ by putting $\frac{m-2}{r}$ copies of $[0,1]^{\otimes r}$ together. Now connect $[0,1]^{\otimes m-2}$ to $\bar{f}$ via $\neq_2$. The resulting signature is $[\bar{f}_0, 1, 0]$. We can also move $=_n$ to the left using n copies of $\neq_2$. Hence, we have that

$$\text{Pl-Holant}(=_n \mid [\bar{f}_0, 1, 0]) \leqslant_T \text{Pl-Holant}(\neq_2 \mid \bar{f}, =_n).$$

The former problem is #**P**-hard by Theorem 1.15 since $\bar{f}_0 \neq 0$, so the latter problem is #**P**-hard as well.

In the second case, we have $m \geqslant 5$ since $2 \leqslant d < \frac{m}{2}$. Furthermore, we may assume that $d = 2$, since otherwise can we do $d-2$ self-loops on $\bar{f}$ via $\neq_2$. With this assumption, we do two self-loops on $\bar{f}$ via $\neq_2$ to get $[1,0]^{\otimes m-4}$ on the right side. By a similar argument as in the previous case, we can construct $[0,1]^{\otimes m-4}$ by using $[1,0]^{\otimes m-4}$ and $=_n$ via $\neq_2$. Now connect $[0,1]^{\otimes m-4}$ back to $\bar{f}$ via $\neq_2$. We get the arity 4 signature $[\bar{f}_0, \bar{f}_1, 1, 0, 0]$. Hence, we have that

$$\text{Pl-Holant}(\neq_2 \mid [\bar{f}_0, \bar{f}_1, 1, 0, 0]) \leqslant_T \text{Pl-Holant}(\neq_2 \mid \bar{f}, =_n).$$

Note that Pl-Holant$(\neq_2 \mid [\bar{f}_0, \bar{f}_1, 1, 0, 0])$ is equivalent to Pl-Holant$(\neq_2 \mid [0,0,1,0,0])$, Pl-4reg-#EO, which is #**P**-hard by Theorem 3.21. Thus Pl-Holant$(\neq_2 \mid \bar{f}, =_n)$ is #**P**-hard as well. □

The second one concerns about signature with self-loops. If the signature with one self-loop is vanishing, then the original one has to be vanishing as well, unless it is #**P**-hard.

Lemma 3.49. *Let f be a non-degenerate symmetric signature of arity $n \geqslant 5$. Let f' be f with a self loop. If f' is non-degenerate and vanishing, then* Pl-Holant(f) *is #**P**-hard unless $\{f, f'\}$ is vanishing, in which case* Pl-Holant(f) *is tractable.*

Proof. Since f' is vanishing, $f' \in \mathcal{V}^{\sigma}$ for some $\sigma \in \{+,-\}$ by Theorem 3.12. For simplicity, assume that $f' \in \mathcal{V}^{+}$. The other case is similar.

Note that f' is of arity $n-2 \geqslant 3$. Suppose $\text{rd}^{+}(f') = d-1$, where $2d < n$ and $d \geqslant 2$ since f'

is non-degenerate. Under the transformation $Z = \frac{1}{\sqrt{2}} \left[\begin{smallmatrix} 1 & 1 \\ i & -i \end{smallmatrix}\right]$, we have that

$$\begin{aligned}\text{Pl-Holant}\left(=_2 \mid f, f'\right) &\equiv_T \text{Pl-Holant}\left([1,0,1]Z^{\otimes 2} \mid (Z^{-1})^{\otimes n} f, (Z^{-1})^{\otimes n} f'\right) \\ &\equiv_T \text{Pl-Holant}\left([0,1,0] \mid \overline{f}, \overline{f'}\right),\end{aligned}$$

where $\overline{f'} = [\overline{f}_1, \ldots, \overline{f}_d, 0, \ldots, 0]$ with $\overline{f}_d \neq 0$ by Lemma 3.14. Note that doing self-loop in the standard basis is the same as connecting to $[0,1,0]$ in the Z basis. Hence we may assume that $\overline{f} = [\overline{f}_0, \overline{f}_1, \ldots, \overline{f}_d, 0, \ldots, 0, c]$, for some $\overline{f}_0$ and c. If $c = 0$, then $\{f, f'\} \subset \mathcal{V}^+$ is vanishing. Hence we may assume that $c \neq 0$. We will show that Pl-Holant(f) is #**P**-hard.

Doing $d-2$ self-loops by $[0,1,0]$ on $\overline{f}$, we get a signature $\overline{h} = [\overline{f}_{d-2}, \overline{f}_{d-1}, \overline{f}_d, 0, \ldots, 0, 0/c]$ of arity $n - 2(d-2) = n - 2d + 4 \geqslant 5$. The last entry of $\overline{h}$ is c when $d = 2$ and is 0 when $d > 2$. As $n > 2d$, we may do two more self loops and get $[\overline{f}_d, 0, \ldots, 0]$ of arity $k = n - 2d \geqslant 1$. This signature is equivalent to $[1,0]^{\otimes k}$. Now connect this signature back to $\overline{f}$ via $[0,1,0]$. It is the same as getting the last $n - k + 1 = 2d + 1$ signature entries of $\overline{f}$ up to a nonzero scalar. We may repeat this operation zero or more times until the arity k' of the resulting signature is less than or equal to k. We claim that this signature has the form $\overline{g} = [0, \ldots, 0, c]$. In other words, the $k' + 1$ entries of $\overline{g}$ consist of the last c and k' many 0's from the signature $\overline{f}$, all appearing after $\overline{f}_d$. This is because there are $n - d - 1$ many 0 entries in the signature $\overline{f}$ after $\overline{f}_d$, and $n - d - 1 \geqslant k \geqslant k'$. Note that $\overline{g} = [0,1]^{\otimes k'}$.

Having both $[1,0]^{\otimes k}$ and $\overline{g} = [0,1]^{\otimes k'}$ in the Z basis, we realize $[0,1]^{\otimes t}$ using the subtractive Euclidean argument as in Lemma 3.41, where $t = \gcd(k, k')$. Then we put $\frac{k}{t}$ many copies of $[0,1]^{\otimes t}$ together to get $[0,1]^k$. Connect $\overline{h}$ with $[0,1]^k$ by $[0,1,0]$. Note that due to $[0,1,0]$ flipping the bits, this gets the prefix of $\overline{h}$ of arity arity$(h) - k$. Recall that arity$(h) = n - 2d + 4$, and hence arity$(h) - k = n - 2d + 4 - (n - 2d) = 4$. The resulting signature has arity 4. Moreover, the signature is $[\overline{f}_{d-2}, \overline{f}_{d-1}, \overline{f}_d, 0, 0]$. The last entry is 0 (and not c), because $k \geqslant 1$ and arity$(\overline{h}) \geqslant 5$.

However, Pl-Holant$([0,1,0] | [\overline{f}_{d-2}, \overline{f}_{d-1}, \overline{f}_d, 0, 0])$ is equivalent to Pl-Holant$([0,1,0] | [0,0,1,0,0])$ when $\overline{f}_d \neq 0$. This is Pl-4reg-#EO and is #**P**-hard by Theorem 3.21. □

Chapter 4

Planar Counting CSP

In this chapter, we will show a dichotomy theorem for planar #CSP problems defined by complex weighted symmetric Boolean functions. This may seem a little detour towards our goal of classifying Holant problems, but in fact it will be a key ingredient to our main planar Holant dichotomy. Recall that for any signature set $\mathcal{F}$, $\widehat{\mathcal{F}}$ denotes $H_2\mathcal{F}$, where $H_2 = \left[\begin{smallmatrix} 1 & 1 \\ 1 & -1 \end{smallmatrix}\right]$. The dichotomy theorem is stated as follows.

Theorem 4.1. *Let $\mathcal{F}$ be any set of symmetric, complex-valued signatures in Boolean variables. Then* Pl-#CSP($\mathcal{F}$) *is #**P**-hard unless $\mathcal{F} \subseteq \mathcal{A}$, $\mathcal{F} \subseteq \mathcal{P}$, or $\mathcal{F} \subseteq \widehat{\mathcal{M}}$, in which case the problem is computable in polynomial time.*

Recall that $\mathcal{EQ}$ denotes the set of all EQUALITY signatures. Our focus is Pl-#CSP($\mathcal{F}$), which is equivalent to Pl-Holant($\mathcal{EQ} \cup \mathcal{F}$), as explained in Section 1.4 and (1.1). We often study this problem in the Hadamard basis, that is, under a holographic transformation by the Hadamard matrix $H_2 = \left[\begin{smallmatrix} 1 & 1 \\ 1 & -1 \end{smallmatrix}\right]$. In the Hadamard basis, the problem to classify becomes Pl-Holant($\widehat{\mathcal{EQ}} \cup \mathcal{F}$).

Results reported in this chapter are mainly from [GW13]. Prior to this work, dichotomy theorems were known for complex weighted Boolean #CSP [DGJ09, BDG+09, CLX14], as well as Boolean Pl-#CSP [CLX10] for real weighted symmetric functions. One important theme of [CLX10] is that all problems that are tractable in planar graphs, but #**P**-hard in general graphs, are captured precisely by holographic algorithms with matchgates (compare Theorem 1.16 to Theorem 4.1). More precisely, the only planar tractable case is $\widehat{\mathcal{M}}$ in the standard basis, or $\mathcal{M}$ in the Hadamard basis. We will show that the theme still holds for complex functions. In contrast, as we will see in Chapter 6, it is not true for Pl-Holant. In fact, $\mathcal{EQ}$ is matchgate realizable only in the Hadamard basis, up to a stabilizer of $\mathcal{M}$, which makes the Hadamard basis the "right" one to work in.

Another reason to work in the Hadamard basis is pinning. Pinning is the very first step to show hardness in many previous dichotomy theorems for Boolean #CSP($\mathcal{F}$) [DGJ09, BDG+09, CLX14]. The goal of pinning is to realize constant functions $[1,0]$ and $[0,1]$ and was always achieved by a *nonplanar* reduction. In the nonplanar setting, $[1,0]$ and $[0,1]$ are contained in each of the maximal tractable sets $\mathcal{A}$ and $\mathcal{P}$ (cf. [CLX14]). Therefore, pinning does not entail any unexpected collapse of complexity classes. However, $\mathcal{EQ}$ with $\{[1,0],[0,1]\}$ are not simultaneously realizable as matchgates. Hence Theorem 4.1 implies that pinning is not possible for Pl-#CSP($\mathcal{F}$), unless #**P** collapses to **P**! Instead, apply the Hadamard transformation and consider Pl-Holant($\widehat{\mathcal{F}} \cup \widehat{\mathcal{EQ}}$). In this Hadamard basis, pinning becomes possible again since $[1,0]$ and $[0,1]$ are included in each maximal tractable set. We prove our pinning result in this Hadamard basis in Section 4.3. Note that as $[1,0] \in \widehat{\mathcal{EQ}}$, we only need to realize $[0,1]$ there.

One important result that we will use extensively is Theorem 1.15. In particular, we will often apply Theorem 1.15 with $\mathcal{G} = \mathcal{EQ}$, in other words, $d = 1$ in Theorem 1.15. This is the special case of Pl-#CSP($\mathcal{F}$) when $\mathcal{F}$ contains a single binary signature. Furthermore, under the holographic transformation by $H_2 = \left[\begin{smallmatrix} 1 & 1 \\ 1 & -1 \end{smallmatrix}\right]$, it is easy to see that conditions $f_0 f_2 = f_1^2$ and $f_0 f_2 = -f_1^2 \wedge f_0 = -f_2$ in Theorem 1.15 are invariant, while conditions $f_1 = 0$ and $f_0 = f_2$ map to each other. Therefore, by an apparent coincidence, the tractability conditions remain the same. To be clear, we restate Theorem 1.15 both before and after a holographic transformation by H with $\mathcal{G} = \mathcal{EQ}$.

Theorem 4.2 (Special case of Theorem 1.15)**.** *For any* $f_0, f_1, f_2 \in \mathbb{C}$*, both* Pl-Holant $([f_0, f_1, f_2] \mid \mathcal{EQ})$ *and* Pl-Holant $\left([f_0, f_1, f_2] \mid \widehat{\mathcal{EQ}}\right)$ *are* #**P***-hard unless one of the following conditions hold, in which case both problems are computable in polynomial time:*

1. $f_0 f_2 = f_1^2$;
2. $f_1 = 0$;
3. $f_0 f_2 = -f_1^2$ *and* $f_0 = -f_2$;

4. $f_0 = f_2$.

As we concern Pl-Holant$(\widehat{\mathcal{EQ}} \cup \mathcal{F})$, we will specialize and summarize tractability Lemmas 1.7, 1.9, and 1.10, as follows. We note that $\widehat{\mathcal{A}} = H_2\mathcal{A} = \mathcal{A}$, and $\widehat{\mathcal{EQ}}$ is matchgate realizable in the standard basis.

Lemma 4.3. *Let $\mathcal{F}$ be a set of symmetric, complex-valued Boolean signatures. Then* Pl-Holant$^c(\mathcal{F} \cup \widehat{\mathcal{EQ}})$ *is tractable if $\mathcal{F} \subseteq \mathcal{A}$, $\mathcal{F} \subseteq \widehat{\mathcal{P}}$, or $\mathcal{F} \subseteq \mathcal{M}$.*

Figure 4.1 is a Venn diagram of the tractable Pl-#CSP signature sets in the Hadamard basis. Each signature may also take an arbitrary constant multiple from $\mathbb{C}$. This figure is particularly useful in Section 4.2, where we consider the complexity of multiple signatures from different tractable sets. The definition of each tractable signature set is given in Section 1.5.

The notation "$f_{\geqslant k}$" is short for signature f with "arity$(f) \geqslant k$". Notice that $\mathcal{M} \cap \widehat{\mathcal{P}} - \mathcal{A}$ is empty.

4.1 Domain Pairing

One important technique that we will use is *domain pairing*. Recall that we have used domain pairing in Chapter 2, in particular, Lemma 2.21. The idea is to pair Boolean variables to simulate a problem on a domain of size four and then reduces a problem in the Boolean domain to it. As explained earlier, we will work in the Hadamard basis instead of the standard basis. The goal is to classify Pl-Holant$(\mathcal{F} \cup \widehat{\mathcal{EQ}})$.

We first prove a simple interpolation lemma for non-degenerate, generalized equality signatures of arity at least 3.

Lemma 4.4. *Let $f = [a, 0, \ldots, 0, b]$ with* arity$(f) \geqslant 3$ *for some $a, b \in \mathbb{C}$. If $ab \neq 0$, then for any set $\mathcal{F}$ containing f,*

$$\text{Pl-Holant}(\mathcal{F} \cup \{=_4\}) \leqslant_T \text{Pl-Holant}(\mathcal{F}).$$

Proof. Since $a \neq 0$, we can normalize the first entry of f to get $[1, 0, \ldots, 0, x]$, where $x \neq 0$. First, we show how to obtain an arity 4 generalized equality signature. If $r = 3$, then we connect two

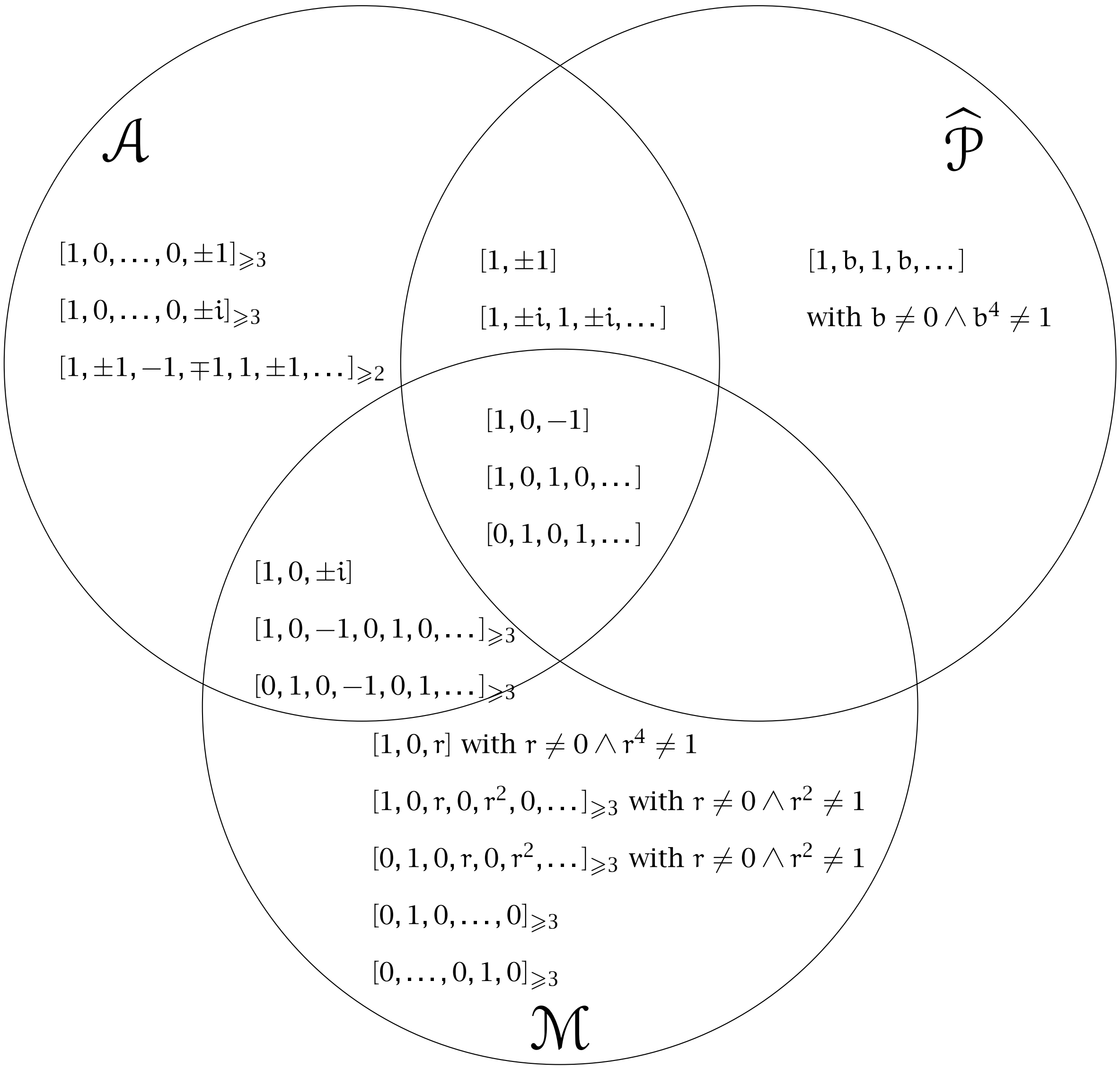


Figure 4.1: Venn diagram of the tractable Pl-#CSP signature sets in the Hadamard basis. Each signature has been normalized for simplicity of presentation. For a signature f, the notation "$f_{\geqslant k}$" is short for "$\text{arity}(f) \geqslant k$".

copies together by a single edge to get an arity 4 signature. For larger arities, we form self-loops until realizing a signature of arity 3 or 4. By this process, we have a signature $g = [1, 0, 0, 0, y]$, where $y \neq 0$. If y is a pth root of unity, then we can directly realize $=_4$ by connecting p copies of g together, two edges at a time as in Figure 3.7. Otherwise, y is not a root of unity and we can interpolate $=_4$ as follows.

Consider an instance Ω of Pl-Holant$(\mathcal{F} \cup \{=_4\})$. Suppose that $=_4$ appears n times in Ω. We stratify the assignments σ's in Ω based on the assignments to $=_4$. We only need to consider

σ's which assign all zeroes or all ones to $=_4$ since otherwise σ contributes 0. Let c_i denote the summation of the product of evaluations of signatures other than $=_4$ in Ω over assignments that give i many $=_4$'s Hamming weight 0 (and $n-i$ many $=_4$'s Hamming weight 4). We can rewrite the Holant on Ω as

$$\text{Holant}_\Omega = \sum_{i=0}^{n} c_i.$$

We construct from Ω a sequence of instances Ω_s of Pl-Holant($\mathcal{F}$) indexed by $s \geqslant 1$. Let g_s be the arity 4 signature of connecting s copies of $[1,0,0,0,y]$, two edges together one pair at a time as in Figure 3.7. We obtain Ω_s from Ω by replacing each occurrence of $=_4$ with g_s. The Holant on Ω_s is

$$\text{Holant}_{\Omega_s} = \sum_{i=0}^{n} (y^s)^i c_i.$$

For $s \geqslant 1$, this gives a coefficient matrix that is Vandermonde. Since y is neither 0 nor a root of unity, y^s is distinct for each s. Therefore, the Vandermonde system has full rank. We can solve for the unknowns c_i and obtain the value of Holant_Ω. □

By a simple parity argument, gadgets constructed with signatures of even arity can only realize other signatures of even arity. In particular, this means that $=_4$ cannot by itself be used to construct $=_3$. The domain pairing argument makes realizing $=_3$ possible using $=_4$ alone. The catch is the domain changes from individual elements to pairs of elements. We prove a generalization of the domain pairing lemma [CLX10, Lemma III.2] for complex weights.

Lemma 4.5. *Let* $a, b, x, y \in \mathbb{C}$. *If* $aby \neq 0$ *and* $x^2 \neq y^2$, *then for any set* $\mathcal{F}$ *of complex-valued symmetric signatures containing* $[x, 0, y, 0]$ *and* $[a, 0, \ldots, 0, b]$ *of arity at least* 3*,* Pl-Holant$(\mathcal{F} \cup \widehat{\mathcal{EQ}})$ *is* #**P***-hard.*

Proof. We reduce from Pl-Holant $([x, y, y] \mid \mathcal{EQ})$ to Pl-Holant$(\mathcal{F} \cup \widehat{\mathcal{EQ}})$. Since Pl-Holant $([x, y, y] \mid \mathcal{EQ})$ is #**P**-hard if $y \neq 0$ and $x^2 \neq y^2$ by Theorem 4.2, this shows that Pl-Holant$(\mathcal{F} \cup \widehat{\mathcal{EQ}})$ is also #**P**-hard.

An instance of Pl-Holant $([x, y, y] \mid \mathcal{EQ})$ is a signature grid Ω with underlying graph $G = (U, V, E)$. In addition to G being bipartite and planar, every vertex in U has degree 2. We replace every vertex in V of degree k (which is assigned $=_k \in \mathcal{EQ}$) with a vertex of degree $2k$, and bundle two adjacent variables to form k bundles of 2 edges each. The k bundles correspond to the k incident edges of the original vertex with degree k. By Lemma 4.4, we have $=_4$, which we use to construct $=_{2k}$ for any k. Then we assign $=_{2k}$ to the new vertices of degree $2k$.

If the inputs to these equality signatures are restricted to $\{(0,0),(1,1)\}$ on each bundle, then these equality signatures take value 1 on $((0,0),\ldots,(0,0))$ and $((1,1),\ldots,(1,1))$ and take value 0 elsewhere. Thus, if we project the domain $[2]\times[2]$ to $\{(0,0),(1,1)\}$, it is the equality signature $=_k$.

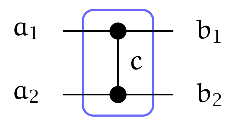


Figure 4.2: Gadget designed for the paired domain. One vertex is assigned $[1,0,1,0]$ and the other is assigned $[x,0,y,0]$.

To simulate $[x,y,y]$, we connect $f=[x,0,y,0]$ to $g=[1,0,1,0]\in\widehat{\mathcal{EQ}}$ by a single edge as shown in Figure 4.2 to form a gadget with signature

$$h(a_1,a_2,b_1,b_2)=\sum_{c=0,1} f(a_1,b_1,c)g(a_2,b_2,c).$$

We replace every (degree 2) vertex in U (which is assigned $[x,y,y]$) by a degree 4 vertex assigned h, where the variables of h are bundled as (a_1,a_2) and (b_1,b_2).

The vertices in this new graph G' are connected as in the original graph G, except that every original edge is replaced by two edges that connect to the same side of the gadget in Figure 4.2. Notice that h is only connected by (a_1,a_2) and (b_1,b_2) to some bundle of two incident edges of an equality signature. Since this equality signature enforces that the value on each bundle is either $(0,0)$ or $(1,1)$, we only need to consider the restriction of h to the domain $\{(0,0),(1,1)\}$. On this domain, h behaves like $[x,y,y]$ as a *symmetric* signature of arity 2. Therefore, the signature grid Ω' with underlying graph G' has the same Holant value as the original signature grid Ω. □

One may notice the apparent similarity between the gadget in Figure 2.4 and the one in Figure 4.2. It is not surprising as both are used for domain pairing arguments, although the details are different.

There are two scenarios leading to Lemma 4.5, the first of which is immediate.

Corollary 4.6. *Let* $a, b, x, y \in \mathbb{C}$*. If* $abxy \neq 0$ *and* $x^4 \neq y^4$*, then for any set* $\mathcal{F}$ *of complex-weighted symmetric signatures containing* $[x, 0, y]$ *and* $[a, 0, \ldots, 0, b]$ *of arity at least* 3*,* Pl-Holant$(\mathcal{F} \cup \widehat{\mathcal{EQ}})$ *is* #**P**-*hard.*

Proof. Connect three copies of $[x, 0, y]$ to $[1, 0, 1, 0]$, with one on each edge. We get $x[x^2, 0, y^2, 0]$. Then apply Lemma 4.5. □

The second scenario that leads to Lemma 4.5 is Lemma 4.8, the proof of which also uses Corollary 4.6. We will apply Corollary 4.6 either directly or after interpolating a unary signature in two possible ways. The next lemma deals with one possibility.

Lemma 4.7. *Suppose* $x \in \mathbb{C}$ *and let* $f = [1, x, 1]$*. If* $x \notin \{0, \pm 1\}$ *and* M_f *has infinite order modulo a scalar, then for any set* $\mathcal{F}$ *of complex-weighted symmetric signatures containing* f *and for any* $a, b \in \mathbb{C}$*, we have*

$$\text{Pl-Holant}(\mathcal{F} \cup \{[a, b]\} \cup \widehat{\mathcal{EQ}}) \leqslant_T \text{Pl-Holant}(\mathcal{F} \cup \widehat{\mathcal{EQ}}).$$

Proof. Consider the unary recursive construction (M_f, s), where $s = \left[\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right] \in \widehat{\mathcal{EQ}}$. The determinant of M_f is $1 - x^2 \neq 0$. The determinant of $[s\ M_f s]$ is $x \neq 0$. By assumption, M_f has infinite order modulo a scalar. Therefore, we can interpolate any unary signature by Lemma 3.32. □

Lemma 4.8. *Let* $a, b \in \mathbb{C}$*. If* $ab \neq 0$ *and* $a^4 \neq b^4$*, then for any set* $\mathcal{F}$ *of complex-weighted symmetric signatures containing* $f = [a, 0, \ldots, 0, b]$ *of arity at least 3,* Pl-Holant$(\mathcal{F} \cup \widehat{\mathcal{EQ}})$ *is* #**P**-*hard.*

Proof. Since $a \neq 0$, we normalize f to $[1, 0, \ldots, 0, x]$, where $x \neq 0$ and $x^4 \neq 1$. If the arity of f is even, then after some number of self-loops, we have $[1, 0, x]$ and are done by Corollary 4.6. Otherwise, the arity of f is odd. After some number of self-loops, we have $g = [1, 0, 0, x]$. If we had the signature $[1, 1]$, then we could connect this to g to get $[1, 0, x]$ and be done by Corollary 4.6. We now show how to interpolate $[1, 1]$ in one of two ways. In either case, we use the signature $[1, x]$, which we obtain via a self-loop on g.

Suppose $\mathfrak{R}(x)$, the real part of x, is not 0. Connecting $[1, x]$ to $[1, 0, 1, 0]$ gives $h = [1, x, 1]$. The eigenvalues of M_h are $\lambda_{\pm} = 1 \pm x$. Since $\mathfrak{R}(x) \neq 0$ if and only if $\left|\frac{\lambda_+}{\lambda_-}\right| \neq 1$, the ratio of the eigenvalues is not a root of unity, so M_h has infinite order modulo a scalar. Therefore, we can interpolate $[1,1]$ by Lemma 4.7.

Otherwise, $\mathfrak{R}(x) = 0$ but x is not a root of unity since $x \neq \pm i$. Connecting $[1,x]$ to g gives $h = [1,0,x^2]$. Clearly $\left(x^2\right)^4 \neq 1$. Hence we apply Corollary 4.6 on h and f, implying that Pl-Holant$(\mathcal{F} \cup \widehat{\mathcal{EQ}})$ is #**P**-hard.. □

4.2 Mixing of Tractable Signatures

In this section, we consider all kinds of cases when tractable signatures of different classes are combined. Basically, each tractable class, $\mathcal{A}$, $\widehat{\mathcal{P}}$, or $\mathcal{M}$, in Lemma 4.3 is maximal. Combining signatures from any two of them gives #**P**-hardness. The Venn diagram in Figure 4.1 is helpful to understand various cases considered in the following lemmas.

The first two lemmas consider the case when one of the signatures is unary.

Lemma 4.9. *Let* $f \in \mathcal{A} - \widehat{\mathcal{P}}$ *be a symmetric signature. Let* $a, b \in \mathbb{C}$ *such that* $ab \neq 0$ *and* $a^4 \neq b^4$*. For any set* $\mathcal{F}$ *of signatures containing* f *and* $[a,b]$*,* Pl-Holant$(\mathcal{F} \cup \widehat{\mathcal{EQ}})$ *is* #**P**-*hard.*

Proof. Up to a nonzero scalar, the possibilities for f are

- $[1,0,\pm i]$;
- $[1,0,\ldots,0,x]$ of arity at least 3 with $x^4 = 1$;
- $[1,\pm 1,-1,\mp 1,1,\pm 1,-1,\mp 1,\ldots]$ of arity at least 2;
- $[1,0,-1,0,1,0,-1,0,\ldots,0 \text{ or } 1 \text{ or } (-1)]$ of arity at least 3;
- $[0,1,0,-1,0,1,0,-1,\ldots,0 \text{ or } 1 \text{ or } (-1)]$ of arity at least 3.

We handle these cases below.

1. Suppose $f = [1,0,\pm i]$. Connecting $[a,b]$ to $[1,0,1,0]$ gives $[a,b,a]$ and connecting two copies of $[1,0,\pm i]$ to $[a,b,a]$, one on each edge, gives $g = [a,\pm ib,-a]$. Since $ab \neq 0$ and $a^4 \neq b^4$, Pl-Holant$\left(g \mid \widehat{\mathcal{EQ}}\right)$ is #**P**-hard by Theorem 4.2, so Pl-Holant$(\mathcal{F} \cup \widehat{\mathcal{EQ}})$ is also #**P**-hard.

2. Suppose $f = [1, 0, \ldots, 0, x]$ of arity at least 3 with $x^4 = 1$. Connecting $[a, b]$ to f gives $g = [a, 0, \ldots, 0, bx]$ of arity at least 2. Note that $(bx)^4 = b^4 \neq a^4$. If the arity of g is exactly 2, then Pl-Holant($\{f, g\} \cup \widehat{\mathcal{EQ}}$) is #**P**-hard by Corollary 4.6, so Pl-Holant($\mathcal{F} \cup \widehat{\mathcal{EQ}}$) is also #**P**-hard. Otherwise, the arity of g is at least 3 and Pl-Holant($\{g\} \cup \widehat{\mathcal{EQ}}$) is #**P**-hard by Lemma 4.8, so Pl-Holant($\mathcal{F} \cup \widehat{\mathcal{EQ}}$) is also #**P**-hard.

3. Suppose $f = [1, \pm 1, -1, \ldots]$ of arity $n \geqslant 2$. Connecting $n-2$ many copies of $[1, 0]$ to f gives $[1, \pm 1, -1]$ of arity exactly 2. Connecting $[a, b]$ to $[1, 0, 1, 0]$ gives $[a, b, a]$ and connecting two copies of $[a, b, a]$ to $[1, \pm 1, -1]$, one on each edge, gives $g = [a^2 \pm 2ab - b^2, \pm(a^2 + b^2), -a^2 \pm 2ab + b^2]$. This is easily verified by

$$\begin{bmatrix} a & b \\ b & a \end{bmatrix} \begin{bmatrix} 1 & \pm 1 \\ \pm 1 & -1 \end{bmatrix} \begin{bmatrix} a & b \\ b & a \end{bmatrix} = \begin{bmatrix} a^2 \pm 2ab - b^2 & \pm(a^2 + b^2) \\ \pm(a^2 + b^2) & -a^2 \pm 2ab + b^2 \end{bmatrix}.$$

Since $ab \neq 0$ and $a^4 \neq b^4$, Pl-Holant$\left(g \mid \widehat{\mathcal{EQ}}\right)$ is #**P**-hard by Theorem 4.2, so Pl-Holant($\mathcal{F} \cup \widehat{\mathcal{EQ}}$) is also #**P**-hard.

4. Suppose $f = [1, 0, -1, 0, \ldots]$ of arity $n \geqslant 3$. Connecting $n - 3$ copies of $[1, 0]$ to f gives $g = [1, 0, -1, 0]$ of arity exactly 3. Connecting $[a, b]$ to g gives $h = [a, -b, -a]$. Since $ab \neq 0$ and $a^4 \neq b^4$, Pl-Holant$\left(h \mid \widehat{\mathcal{EQ}}\right)$ is #**P**-hard by Theorem 4.2, so Pl-Holant($\mathcal{F} \cup \widehat{\mathcal{EQ}}$) is also #**P**-hard.

5. The argument for $f = [0, 1, 0, -1, \ldots]$ is similar to the previous case. □

Lemma 4.10. *Let* $f \in \mathcal{M} - \mathcal{A}$ *be a symmetric signature. If* $ab \neq 0$*, then for any set* $\mathcal{F}$ *of signatures containing* f *and* $[a, b]$*, Pl-Holant(*$\mathcal{F} \cup \widehat{\mathcal{EQ}}$*) is #**P**-hard.*

Proof. Up to a nonzero scalar, the possibilities for f are

- $[1, 0, r]$ with $r \neq 0$ and $r^4 \neq 1$;
- $[1, 0, r, 0, r^2, 0, \ldots]$ of arity at least 3 with $r \neq 0$ and $r^2 \neq 1$;
- $[0, 1, 0, r, 0, r^2, \ldots]$ of arity at least 3 with $r \neq 0$ and $r^2 \neq 1$;
- $[0, 1, 0, \ldots, 0]$ of arity at least 3;

- $[0, \ldots, 0, 1, 0]$ of arity at least 3.

We handle these cases below.

1. Suppose $f = [1, 0, r]$ with $r^4 \neq 1$ and $r \neq 0$. Connecting $[a, b]$ to $[1, 0, 1, 0]$ gives $[a, b, a]$ and connecting two copies of $[1, 0, r]$ to $[a, b, a]$, one on each edge, gives $g = [a, br, ar^2]$. If $a^2 \neq b^2$, then Pl-Holant$\left(g \mid \widehat{\mathcal{EQ}}\right)$ is #**P**-hard by Theorem 4.2, so Pl-Holant$(\mathcal{F} \cup \widehat{\mathcal{EQ}})$ is also #**P**-hard.

   Otherwise, $a^2 = b^2$ and we begin by connecting $[a, b]$ to $[1, 0, r]$ to get $[a, br]$. Then by the same construction, we have $g = [a, br^2, ar^2]$ and Pl-Holant$\left(g \mid \widehat{\mathcal{EQ}}\right)$ is #**P**-hard by Theorem 4.2, so Pl-Holant$(\mathcal{F} \cup \widehat{\mathcal{EQ}})$ is also #**P**-hard.

2. Suppose $f = [1, 0, r, 0, \ldots]$ of arity $n \geqslant 3$ with $r^2 \neq 1$ and $r \neq 0$. Connecting $n-3$ copies of $[1, 0]$ to f gives $g = [1, 0, r, 0]$ of arity exactly 3. Connecting $[a, b]$ to g gives $h = [a, br, a]$. If $a^2 \neq b^2 r$, then Pl-Holant$\left(h \mid \widehat{\mathcal{EQ}}\right)$ is #**P**-hard by Theorem 4.2, so Pl-Holant$(\mathcal{F} \cup \widehat{\mathcal{EQ}})$ is also #**P**-hard.

   Otherwise, $a^2 = b^2 r$ and we begin by connecting $[1, 0]$ and $[a, b]$ to $[1, 0, r, 0]$ to get $[a, br]$. Then by the same construction, we have $g = [a, br^2, ar]$ and Pl-Holant$\left(g \mid \widehat{\mathcal{EQ}}\right)$ is #**P**-hard by Theorem 4.2, so Pl-Holant$(\mathcal{F} \cup \widehat{\mathcal{EQ}})$ is also #**P**-hard.

3. The argument for $f = [0, 1, 0, r, \ldots]$ is similar to the previous case.

4. Suppose $f = [0, 1, 0, \ldots, 0]$ of arity $n \geqslant 3$. Connecting $n - 2$ copies of $[a, b]$ to f gives $g = a^{k-3}[(k-2)b, a, 0]$. Since $ab \neq 0$, Pl-Holant$\left(g \mid \widehat{\mathcal{EQ}}\right)$ is #**P**-hard by Theorem 4.2, so Pl-Holant$(\mathcal{F} \cup \widehat{\mathcal{EQ}})$ is also #**P**-hard.

5. The argument for $f = [0, \ldots, 0, 1, 0]$ is similar to the previous case. □

Now we consider the general case of two signatures from two different tractable sets. The three tractable sets, $\mathcal{A}$, $\widehat{\mathcal{P}}$, and $\mathcal{M}$, give rise to three pairs of tractable sets to consider, each of which is covered in one of the next three lemmas.

Lemma 4.11. *Let* $f \in \mathcal{A} - \widehat{\mathcal{P}}$ *and* $g \in \widehat{\mathcal{P}} - \mathcal{A}$ *be two symmetric signatures. For any set* $\mathcal{F}$ *of signatures containing* f *and* g*,* Pl-Holant$(\mathcal{F} \cup \widehat{\mathcal{EQ}})$ *is #**P**-hard.*

Proof. The only possibility for g is $[a, b, a, b, \dots]$ of arity n, where $ab \neq 0$ and $a^4 \neq b^4$. Connecting $n-1$ copies of $[1,0]$ to g gives $[a,b]$ and we are done by Lemma 4.9. □

Lemma 4.12. *Let $f \in \mathcal{A} - \mathcal{M}$ and $g \in \mathcal{M} - \mathcal{A}$ be two symmetric signatures. For any set $\mathcal{F}$ of signatures containing f and g,* Pl-Holant$(\mathcal{F} \cup \widehat{\mathcal{EQ}})$ *is #**P**-hard.*

Proof. Suppose f has arity n. If f does not contain any 0 entry, then after connecting $n-1$ copies of $[1,0]$ to f, we have a unary signature $[a,b]$ with $ab \neq 0$. Then we are done by Lemma 4.10. Otherwise, f contains a 0 entry. Then $f = [x, 0, \dots, 0, y]$ of arity at least 3 with $xy \neq 0$ (and $x^4 = y^4$). (See Figure 4.1.)

Suppose g has arity k. Up to a nonzero scalar, the possibilities for g are (again, cf. 4.1):

- $[1, 0, r]$ with $r \neq 0$ and $r^4 \neq 1$;
- $[1, 0, r, 0, r^2, 0, \dots]$ with $k \geqslant 3$, $r \neq 0$, and $r^2 \neq 1$;
- $[0, 1, 0, r, 0, r^2, \dots]$ with $k \geqslant 3$, $r \neq 0$, and $r^2 \neq 1$;
- $[0, 1, 0, 0, \dots, 0]$ with $k \geqslant 3$;
- $[0, \dots, 0, 0, 1, 0]$ with $k \geqslant 3$.

We handle these cases below.

1. Suppose $g = [1, 0, r]$ with $r \neq 0$ and $r^4 \neq 1$. Then we are done by Corollary 4.6.
2. Suppose $g = [1, 0, r, 0, \dots]$ with $k \geqslant 3$, $r \neq 0$, and $r^2 \neq 1$. After connecting $k-3$ copies of $[1,0]$ to g, we have $h = [1, 0, r, 0]$ of arity exactly 3. Then Pl-Holant$(\{f, h\} \cup \widehat{\mathcal{EQ}})$ is #**P**-hard by Lemma 4.5, so Pl-Holant$(\mathcal{F} \cup \widehat{\mathcal{EQ}})$ is also #**P**-hard.
3. Suppose $g = [0, 1, 0, r, \dots]$ with $k \geqslant 3$, $r \neq 0$, and $r^2 \neq 1$. After connecting $k-3$ copies of $[1,0]$ to g, we have $h = [0, 1, 0, r]$ of arity exactly 3. Connecting two more copies of $[1,0]$ to h gives $[0,1]$. Then we apply a holographic transformation by $T = \left[\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right]$, so f is transformed to $\hat{f} = [y, 0, \dots, 0, x]$ and h is transformed to $\hat{h} = [r, 0, 1, 0]$. Every even arity signature in $\widehat{\mathcal{EQ}}$ remains unchanged after a holographic transformation by T. By attaching $[0,1]T = [1,0]$ to every even arity signature in $T\widehat{\mathcal{EQ}}$, we obtain all of the odd arity signatures

in $\widehat{\mathcal{EQ}}$ again. Then Pl-Holant($\{\hat{f}, \hat{h}\} \cup \widehat{\mathcal{EQ}}$) is #**P**-hard by Lemma 4.5, so Pl-Holant($\mathcal{F} \cup \widehat{\mathcal{EQ}}$) is also #**P**-hard.

4. Suppose $g = [0, 1, 0, \dots, 0]$ with $k \geqslant 3$. The gadget in Figure 4.3 with g assigned to both vertices has signature $h = [k-1, 0, 1]$. Then Pl-Holant($\{f, h\} \cup \widehat{\mathcal{EQ}}$) is #**P**-hard by Corollary 4.6, so Pl-Holant($\mathcal{F} \cup \widehat{\mathcal{EQ}}$) is also #**P**-hard.

5. The argument for $g = [0, \dots, 0, 1, 0]$ is similar to the previous case. □

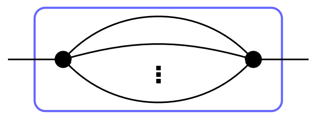


Figure 4.3: The vertices are assigned $g = [0, 1, 0, \dots, 0]$.

Lemma 4.13. *Let* $f \in \mathcal{M} - \widehat{\mathcal{P}}$ *and* $g \in \widehat{\mathcal{P}} - \mathcal{M}$ *be two symmetric signatures such that* $\{f, g\} \not\subseteq \mathcal{A}$. *Then for any set* $\mathcal{F}$ *of signatures containing* f *and* g, Pl-Holant($\mathcal{F} \cup \widehat{\mathcal{EQ}}$) *is* #**P**-*hard.*

Proof. The only possibility for g is $[a, b, a, b, \dots]$, where $ab \neq 0$. Suppose g has arity $n > 0$. Connecting $n-1$ copies of $[1, 0]$ to g gives $h = [a, b]$. If $f \notin \mathcal{A}$, then Pl-Holant($\{f, h\} \cup \widehat{\mathcal{EQ}}$) is #**P**-hard by Lemma 4.10, so Pl-Holant($\mathcal{F} \cup \widehat{\mathcal{EQ}}$) is also #**P**-hard.

Otherwise, $f \in \mathcal{A}$, so $g \notin \mathcal{A}$. Then Pl-Holant($\{f, g\} \cup \widehat{\mathcal{EQ}}$) is #**P**-hard by Lemma 4.11, so Pl-Holant($\mathcal{F} \cup \widehat{\mathcal{EQ}}$) is also #**P**-hard. □

We summarize this section with the following theorem, which says that the tractable signature sets cannot mix. More formally, signatures from different tractable sets, when put together, lead to #**P**-hardness.

Theorem 4.14 (Mixing). *Let* $\mathcal{F}$ *be any set of symmetric, complex-valued signatures in Boolean variables. If* $\mathcal{F} \subseteq \mathcal{A} \cup \widehat{\mathcal{P}} \cup \mathcal{M}$, *then* Pl-Holant($\mathcal{F} \cup \widehat{\mathcal{EQ}}$) *is* #**P**-*hard unless* $\mathcal{F} \subseteq \mathcal{A}$, $\mathcal{F} \subseteq \widehat{\mathcal{P}}$, *or* $\mathcal{F} \subseteq \mathcal{M}$, *in which case* Pl-Holant^c($\mathcal{F} \cup \widehat{\mathcal{EQ}}$) *is tractable.*

Proof. If $\mathcal{F}$ is a subset of $\mathcal{A}$, $\widehat{\mathcal{P}}$, or $\mathcal{M}$, then the tractability is given in Lemma 4.3. Otherwise $\mathcal{F}$ is not a subset of $\mathcal{A}$, $\widehat{\mathcal{P}}$, or $\mathcal{M}$. Then $\mathcal{F}$ contains a signature $g \in (\widehat{\mathcal{P}} \cup \mathcal{M}) - \mathcal{A}$ since $\mathcal{F} \not\subseteq \mathcal{A}$. Suppose $\mathcal{F}$ contains a signature $f \in \mathcal{A} - \widehat{\mathcal{P}} - \mathcal{M}$. If $g \in \widehat{\mathcal{P}} - \mathcal{A}$, then Pl-Holant($\mathcal{F} \cup \widehat{\mathcal{EQ}}$) is #**P**-hard by Lemma 4.11. Otherwise, $g \in \mathcal{M} - \mathcal{A}$ and Pl-Holant($\mathcal{F} \cup \widehat{\mathcal{EQ}}$) is #**P**-hard by Lemma 4.12.

Now assume that $\mathcal{F} \subseteq \widehat{\mathcal{P}} \cup \mathcal{M}$. Since $(\widehat{\mathcal{P}} \cap \mathcal{M}) - \mathcal{A}$ is empty (see Figure 4.1, either $g \in \widehat{\mathcal{P}} - \mathcal{M} - \mathcal{A}$ or $g \in \mathcal{M} - \widehat{\mathcal{P}} - \mathcal{A}$. If $g \in \widehat{\mathcal{P}} - \mathcal{M} - \mathcal{A}$, then there exists a signature $f \in \mathcal{M} - \widehat{\mathcal{P}}$ since $\mathcal{F} \not\subseteq \widehat{\mathcal{P}}$. In which case, Pl-Holant($\mathcal{F} \cup \widehat{\mathcal{EQ}}$) is #**P**-hard by Lemma 4.13. Otherwise, $g \in \mathcal{M} - \widehat{\mathcal{P}} - \mathcal{A}$ and there exists a signature $f \in \widehat{\mathcal{P}} - \mathcal{M}$ since $\mathcal{F} \not\subseteq \mathcal{M}$. In which case, Pl-Holant($\mathcal{F} \cup \widehat{\mathcal{EQ}}$) is #**P**-hard by Lemma 4.13. □

4.3 Pinning for Planar #CSP

The idea of "pinning" is a common reduction technique between counting problems. For the #CSP framework, pinning fixes some variables to specific values of the domain by means of the constant functions [BD07, DGJ09, BDG+09, HL12]. In particular, for counting graph homomorphisms, pinning is used when the input graph is connected and the target graph is disconnected. In this case, pinning a vertex of the input graph to a vertex of the target graph forces all the vertices of the input graph to map to the same connected component of the target graph [DG00a, BG05, GGJT10, Thu10, CCL13]. For the Boolean domain, the constant 0 and constant 1 functions are the signatures $[1, 0]$ and $[0, 1]$ respectively.

From these works, the most relevant pinning lemma for the Pl-#CSP framework is by Dyer, Goldberg, and Jerrum in [DGJ09], where they show how to pin in the #CSP framework. However, the proof of this pinning lemma is intrinsically nonplanar. Cai, Lu, and Xia [CLX10] overcame this difficultly in the proof of their dichotomy theorem for the real-weighted Pl-#CSP framework by first applying a holographic transformation by the Hadamard matrix $H_2 = \left[\begin{smallmatrix} 1 & 1 \\ 1 & -1 \end{smallmatrix}\right]$ and then pinning in this Hadamard basis.[1] We stress that this holographic transformation is necessary. Indeed, if one were able to pin in the standard basis of the Pl-#CSP framework, then P = #P would follow since Pl-#CSP($\widehat{\mathcal{M}}$) is tractable but Pl-#CSP($\widehat{\mathcal{M}} \cup \{[1, 0], [0, 1]\}$) is #**P**-hard by our main dichotomy in Theorem 4.1 (or, more specifically, by Lemma 4.10).

Since Pl-#CSP($\mathcal{F}$) is Turing equivalent to Pl-Holant($\mathcal{F} \cup \mathcal{EQ}$), the expression of Pl-#CSP($\mathcal{F}$) in the Hadamard basis is Pl-Holant($\widehat{\mathcal{F}} \cup \widehat{\mathcal{EQ}}$). Then we already have $[1, 0] \in \widehat{\mathcal{EQ}}$, so pinning in the Hadamard basis of Pl-#CSP($\mathcal{F}$) amounts to obtaining the missing signature $[0, 1]$.

[1] The pinning in [CLX10], which is accomplished in Section IV, is not summarized in a single statement but is implied by the combination of all the results in that section.

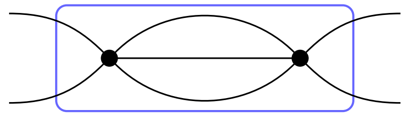


Figure 4.4: The circles are assigned $[a, 0, 0, 0, b, c]$.

The Road to Pinning

We begin the road to pinning with a lemma that assumes the presence of $[0, 0, 1] = [0, 1]^{\otimes 2}$, which is the tensor product of two copies of $[0, 1]$. In our pursuit to realize $[0, 1]$, this may be as close as we can get, such as when every signature has even arity. Another roadblock to realizing $[0, 1]$ is when every signature has even parity. Recall that a signature has even parity if its support is on entries of even Hamming weight. By a simple parity argument, gadgets constructed with signatures of even parity can only realize signatures of even parity. However, if every signature has even parity and $[0, 0, 1]$ is present, then we can already prove a dichotomy, which is Lemma 4.16.

Before proving Lemma 4.16, let us show the following technical lemma first. It will be used in the proof of Lemma 4.16, as well as in Section 4.4.

Lemma 4.15. *Let* $a, b, c \in \mathbb{C}$. *If* $ab \neq 0$, *then* $\operatorname{Pl-Holant}([a, 0, 0, 0, b, c])$ *is* #**P**-*hard.*

Proof. Let f be the signature of the gadget in Figure 4.4 with $[a, 0, 0, 0, b, c]$ assigned to both vertices. The signature matrix of f is

$$M_f = \begin{bmatrix} a^2 & 0 & 0 & 0 \\ 0 & b^2 & b^2 & bc \\ 0 & b^2 & b^2 & bc \\ 0 & bc & bc & 3b^2 + c^2 \end{bmatrix},$$

which is redundant. It is easy to verify that $\det\left(\widetilde{M_f}\right) = 6a^2b^4 \neq 0$. Thus, $\operatorname{Pl-Holant}(f)$ is #**P**-hard by Corollary 3.28, so $\operatorname{Pl-Holant}([a, 0, 0, 0, b, c])$ is also #**P**-hard. □

Lemma 4.16. *Suppose* $\mathcal{F}$ *is a set of symmetric signatures with complex weights containing* $[0, 0, 1]$. *If every signature in* $\mathcal{F}$ *has even parity, then either* $\operatorname{Pl-Holant}(\mathcal{F} \cup \widehat{\mathcal{EQ}})$ *is* #**P**-*hard or* $\mathcal{F}$ *is a subset of* $\mathcal{A}$, $\widehat{\mathcal{P}}$, *or* $\mathcal{M}$, *in which case* $\operatorname{Pl-Holant}^c(\mathcal{F} \cup \widehat{\mathcal{EQ}})$ *is tractable.*

Proof. The tractability is given in Lemma 4.3. If every non-degenerate signature in $\mathcal{F}$ is of arity at most 3, then $\mathcal{F} \subseteq \mathcal{M}$ since all signatures in $\mathcal{F}$ satisfy the (even) parity condition.

Otherwise $\mathcal{F}$ contains some non-degenerate signature of arity at least 4. For every signature $f \in \mathcal{F}$ with $f = [f_0, f_1, \ldots, f_m]$ and $m \geqslant 4$, using $[0, 0, 1]$ and $[1, 0]$, we can obtain all subsignatures of the form $[f_{k-2}, 0, f_k, 0, f_{k+2}]$ for any even k such that $2 \leqslant k \leqslant m - 2$. If any subsignature g of this form satisfies $f_{k-2} f_{k+2} \neq f_k^2$ and $f_k \neq 0$, then Pl-Holant(g) is #**P**-hard by Corollary 3.28, so Pl-Holant$(\mathcal{F} \cup \widehat{\mathcal{EQ}})$ is also #**P**-hard.

Otherwise all subsignatures of signatures in $\mathcal{F}$ of the above form satisfy $f_{k-2} f_{k+2} = f_k^2$ or $f_k = 0$. There are two types of signatures with this property. In the first type, the signature entries of even Hamming weight form a geometric progression. More specifically, the signatures of the first type have the form

$$[\alpha^n, 0, \alpha^{n-1}\beta, 0, \ldots, 0, \alpha\beta^{n-1}, 0, \beta^n] \qquad \text{or} \qquad [\alpha^n, 0, \alpha^{n-1}\beta, 0, \ldots, 0, \alpha\beta^{n-1}, 0, \beta^n, 0]$$

for some $\alpha, \beta \in \mathbb{C}$, which are in $\mathcal{M}$. In the second type, the signatures have arity at least 4 or 5 and are of the form $[x, 0, \ldots, 0, y]$ or $[x, 0, \ldots, 0, y, 0]$ respectively, with $xy \neq 0$ and an odd number of 0's between x and y (since they have even parity). If all of the signatures in $\mathcal{F}$ are of the first type, then $\mathcal{F} \subseteq \mathcal{M}$.

Otherwise $\mathcal{F}$ contains a signature f of the second type. Suppose $f = [x, 0, \ldots, 0, y, 0]$ of arity $n \geqslant 5$ with $xy \neq 0$. Then n is odd as there are odd number of 0's between x and y. With $\frac{n-5}{2}$ many self-loops, we have $g = [x, 0, 0, 0, y, 0]$ of arity exactly 5. Then Pl-Holant(g) is #**P**-hard by Lemma 4.15, so Pl-Holant$(\mathcal{F} \cup \widehat{\mathcal{EQ}})$ is also #**P**-hard.

Otherwise $f = [x, 0, \ldots, 0, y]$ of arity at least 4 with $xy \neq 0$. If $x^4 \neq y^4$, then Pl-Holant$(\mathcal{F} \cup \widehat{\mathcal{EQ}})$ is #**P**-hard by Lemma 4.8. Otherwise $x^4 = y^4$. This puts every signature of the second type in $\mathcal{A}$. Therefore $\mathcal{F} \subseteq \mathcal{A} \cup \mathcal{M}$ and we are done by Theorem 4.14. □

Every other result in the rest of this section states that we are able to pin (under various assumptions on $\mathcal{F}$). Formally speaking, we repeatedly prove that Pl-Holant$^c(\mathcal{F} \cup \widehat{\mathcal{EQ}})$ is #**P**-hard (or in **P**) if and only if Pl-Holant$(\mathcal{F} \cup \widehat{\mathcal{EQ}})$ is #**P**-hard (or in **P**). The difference between these two counting problems is the presence of $[0, 1]$ in Pl-Holant$^c(\mathcal{F} \cup \widehat{\mathcal{EQ}})$. We always prove this statement in one of three ways:

1. we show that $\text{Pl-Holant}^c(\mathcal{F} \cup \widehat{\mathcal{EQ}})$ is tractable (so $\text{Pl-Holant}(\mathcal{F} \cup \widehat{\mathcal{EQ}})$ is as well);
2. or we show that $\text{Pl-Holant}(\mathcal{F} \cup \widehat{\mathcal{EQ}})$ is #**P**-hard (so $\text{Pl-Holant}^c(\mathcal{F} \cup \widehat{\mathcal{EQ}})$ is as well);
3. or we show $\text{Pl-Holant}^c(\mathcal{F} \cup \widehat{\mathcal{EQ}}) \leqslant_T \text{Pl-Holant}(\mathcal{F} \cup \widehat{\mathcal{EQ}})$ by realizing $[0, 1]$ using signatures in $\mathcal{F} \cup \widehat{\mathcal{EQ}}$.

Lemma 4.17. *Let $\mathcal{F}$ be any set of complex-weighted symmetric signatures containing $[0, 0, 1]$. Then $\text{Pl-Holant}^c(\mathcal{F} \cup \widehat{\mathcal{EQ}})$ is #**P**-hard (or in **P**) if and only if $\text{Pl-Holant}(\mathcal{F} \cup \widehat{\mathcal{EQ}})$ is #**P**-hard (or in **P**).*

Proof. If we had a unary signature $[a, b]$ where $b \neq 0$, then connecting $[a, b]$ to $[0, 0, 1]$ gives the signature $[0, b]$, which is $[0, 1]$ after normalization. Thus, in order to reduce $\text{Pl-Holant}^c(\mathcal{F} \cup \widehat{\mathcal{EQ}})$ to $\text{Pl-Holant}(\mathcal{F} \cup \widehat{\mathcal{EQ}})$ by constructing $[0, 1]$, it suffices to construct a unary signature $[a, b]$ with $b \neq 0$.

For every signature $f \in \mathcal{F}$ with $f = [f_0, f_1, \ldots, f_m]$, using $[0, 0, 1]$ and $[1, 0]$, we can obtain all subsignatures of the form $[f_{k-1}, f_k]$ for any odd k such that $1 \leqslant k \leqslant m$. If any subsignature satisfies $f_k \neq 0$, then we can construct $[0, 1]$.

Otherwise all signatures in $\mathcal{F}$ have even parity and we are done by Lemma 4.16. □

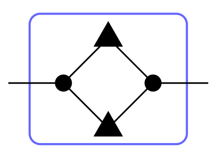


Figure 4.5: The circles are assigned $[1, 0, 1, 0]$ and the triangles are assigned $[1, 0, x]$.

There are two scenarios that lead to Lemma 4.17, which are the focus of the next two lemmas.

Lemma 4.18. *For $x \in \mathbb{C}$, let $\mathcal{F}$ be a set of complex-weighted symmetric signatures containing $[1, 0, x]$ such that $x \notin \{0, \pm 1\}$. Then $\text{Pl-Holant}^c(\mathcal{F} \cup \widehat{\mathcal{EQ}})$ is #**P**-hard (or in **P**) if and only if $\text{Pl-Holant}(\mathcal{F} \cup \widehat{\mathcal{EQ}})$ is #**P**-hard (or in **P**).*

Proof. There are two cases. In either case, we realize $[0, 0, 1]$ and finish by applying Lemma 4.17.

First we claim that the conclusion holds provided $|x| \neq 0, 1$. Combining k copies of $[1, 0, x]$ gives $[1, 0, x^k]$. Since $|x| \notin \{0, 1\}$, x is neither zero nor a root of unity, so we can use polynomial interpolation to realize $[a, 0, b]$ for any $a, b \in \mathbb{C}$, including $[0, 0, 1]$.

Otherwise $|x| = 1$. The gadget in Figure 4.5 has signature $[f_0, f_1, f_2] = [1+x^2, 0, 2x]$. If $x = \pm i$, then we have $[0, 0, \pm 2i]$, which is $[0, 0, 1]$ after normalization.

Otherwise $x \neq \pm i$, so $f_0 \neq 0$. Since $x \neq 0$, $f_2 \neq 0$. Since $x \neq \pm 1$, $|f_0| < 2$. However, $|f_2| = 2$. Therefore, after normalization, the signature $[1, 0, y]$ with $y = \frac{2x}{1+x^2}$ has $|y| > 1$, so it can interpolate $[0, 0, 1]$ by our initial claim since $|y| \notin \{0, 1\}$. □

Lemma 4.19. *Let $\mathcal{F}$ be a set of complex-weighted symmetric signatures containing a signature f that is not identically 0 but $f_0 = 0$. Then* $\text{Pl-Holant}^c(\mathcal{F} \cup \widehat{\mathcal{EQ}})$ *is #**P**-hard (or in **P**) if and only if* $\text{Pl-Holant}(\mathcal{F} \cup \widehat{\mathcal{EQ}})$ *is #**P**-hard (or in **P**).*

Proof. Suppose the first non-zero entry of f is f_i where $i > 0$ and f has arity n. Connecting $n - i$ many copies of $[1, 0]$ to f gives us $[0, 0, \cdots, 0, f_i]$ of arity i. With $\lceil \frac{i-2}{2} \rceil$ many self-loops, we get $[0, f_i]$ or $[0, 0, f_i]$ depending on the parity of i, which is either $[0, 1]$ or $[0, 0, 1]$ after normalization. If it is $[0, 1]$, then we are done directly. If it is $[0, 0, 1]$, then we are done by Lemma 4.17. □

As a significant step toward pinning for any signature set $\mathcal{F}$, we show how to pin given any binary signature. Some cases resist pinning and are excluded.

Lemma 4.20. *Let $\mathcal{F}$ be a set of complex-weighted symmetric signatures containing a binary signature f. Then* $\text{Pl-Holant}^c(\mathcal{F} \cup \widehat{\mathcal{EQ}})$ *is #**P**-hard (or in P) if and only if* $\text{Pl-Holant}(\mathcal{F} \cup \widehat{\mathcal{EQ}})$ *is #**P**-hard (or in P) unless $f \in \{[0, 0, 0], [1, 0, -1], [1, r, r^2], [1, b, 1]\}$, up to a nonzero scalar, for any $b, r \in \mathbb{C}$.*

Proof. Let $f = [f_0, f_1, f_2]$. If $f_0 = 0$ and either $f_1 \neq 0$ or $f_2 \neq 0$, then we are done by Lemma 4.19. Otherwise, $f = [0, 0, 0]$ or $f_0 \neq 0$, in which case we normalize f_0 to 1. If $\text{Pl-Holant}\left(f \mid \widehat{\mathcal{EQ}}\right)$ is #**P**-hard by Theorem 4.2, then $\text{Pl-Holant}(\mathcal{F} \cup \widehat{\mathcal{EQ}})$ is also #**P**-hard. Otherwise, f is one of the tractable cases, which implies that

$$f \in \{[0, 0, 0], [1, r, r^2], [1, 0, x], [1, \pm 1, -1], [1, b, 1]\}.$$

If $f = [1, \pm 1, -1]$, then we connect f to $[1, 0, 1, 0]$ to get $[0, \pm 2]$, which is $[0, 1]$ after normalization. If $f = [1, 0, x]$, then we are done by Lemma 4.18 unless $x \in \{0, \pm 1\}$. The remaining cases are all excluded by assumption, so we are done. □

Pinning in the Hadamard Basis

Before we show how to pin in the Hadamard basis, we handle two simple cases.

Lemma 4.21. *For any set* $\mathcal{F}$ *of complex-weighted symmetric signatures containing* $[1, \pm i]$*, we have* $\text{Pl-Holant}^c(\mathcal{F} \cup \widehat{\mathcal{EQ}}) \leqslant_T \text{Pl-Holant}(\mathcal{F} \cup \widehat{\mathcal{EQ}})$.

Proof. Connect two copies of $[1, \pm i]$ to $[1, 0, 1, 0]$ to get $[0, \pm 2i]$, which is $[0, 1]$ after normalization. □

The next lemma considers the signature $[1, b, 1, b^{-1}]$, which we also encounter in Theorem 4.24, the single signature dichotomy.

Lemma 4.22. *Let* $b \in \mathbb{C}$. *If* $b \notin \{0, \pm 1\}$*, then for any set* $\mathcal{F}$ *of complex-weighted symmetric signatures containing* $f = [1, b, 1, b^{-1}]$, $\text{Pl-Holant}(\mathcal{F} \cup \widehat{\mathcal{EQ}})$ *is #**P**-hard.*

Proof. Connect two copies of $[1, 0]$ to f to get $[1, b]$. Connecting this back to f gives $g = [1 + b^2, 2b, 2]$. Then $\text{Pl-Holant}\left(g \mid \widehat{\mathcal{EQ}}\right)$ is #**P**-hard by Theorem 4.2, so $\text{Pl-Holant}(\mathcal{F} \cup \widehat{\mathcal{EQ}})$ is also #**P**-hard. □

Now we are ready to prove our pinning result.

Theorem 4.23 (Pinning)**.** *Let* $\mathcal{F}$ *be any set of complex-weighted symmetric signatures. Then* $\text{Pl-Holant}^c(\mathcal{F} \cup \widehat{\mathcal{EQ}})$ *is #**P**-hard (or in P) if and only if* $\text{Pl-Holant}(\mathcal{F} \cup \widehat{\mathcal{EQ}})$ *is #**P**-hard (or in P).*

This theorem does not exclude the possibility that either framework can express a problem of intermediate complexity. It merely says that if one framework cannot express a problem of intermediate complexity, then neither can the other. Our goal is to prove a dichotomy for $\text{Pl-Holant}(\mathcal{F} \cup \widehat{\mathcal{EQ}})$. By Theorem 4.23, this is equivalent to proving a dichotomy for $\text{Pl-Holant}^c(\mathcal{F} \cup \widehat{\mathcal{EQ}})$.

Proof of Theorem 4.23. For simplicity, we normalize the first nonzero entry of every signature in $\mathcal{F}$ to 1. If $\mathcal{F}$ contains the degenerate signature $[0, 1]^{\otimes n}$ for some $n \geqslant 1$, then we take self-loops on this signature until we have either $[0, 1]$ or $[0, 0, 1]$ (depending on the parity of n). If we have $[0, 1]$, we are done. Otherwise, we have $[0, 0, 1]$ and are done by Lemma 4.17.

Now assume that any degenerate signature in $\mathcal{F}$ is not of the form $[0,1]^{\otimes n}$. Then we can replace these degenerate signatures in $\mathcal{F}$ by their unary versions using $[1,0]$. This does not change the complexity of the problem. Hence we may assume all degenerate signatures in $\mathcal{F}$ are unary. If $\mathcal{F}$ contains only unary signatures, then $\mathcal{F} \subseteq \widehat{\mathcal{P}}$ and Pl-Holant$^c(\mathcal{F} \cup \widehat{\mathcal{EQ}})$ is tractable by Lemma 4.3.

Otherwise $\mathcal{F}$ contains a non-degenerate signature f of arity $k \geqslant 2$. We connect $k-2$ copies of $[1,0]$ to f until we obtain a signature with arity exactly 2. We call the resulting signature the binary prefix of f. If this binary prefix is not one of the exceptional forms in Lemma 4.20, then we are done, so assume that it is one of the exceptional forms.

We do case analysis according to the exceptional forms in Lemma 4.20. There are five cases below because we split the case of $[1,r,r^2]$ into $[1,0,0]$ and $[1,r,r^2]$ with $r \neq 0$ as two separate cases. In each case, we either show that the conclusion of the theorem holds or that $f \in \mathcal{A} \cup \widehat{\mathcal{P}} \cup \mathcal{M}$, for each non-degenerate $f \in \mathcal{F}$. After the case analysis, we have that $\mathcal{F} \subseteq \mathcal{A} \cup \widehat{\mathcal{P}} \cup \mathcal{M}$. Then we are done by Theorem 4.14.

1. Suppose the binary prefix of f is $[0,0,0]$. Since f is not degenerate, then f is not identically 0, and we are done by Lemma 4.19. Thus, in this case, the theorem holds.

2. Suppose the binary prefix of f is $[1,0,-1]$. If f is not of the form

$$[1,0,-1,0,1,0,-1,0,\ldots,0 \text{ or } 1 \text{ or } (-1)], \tag{4.1}$$

   then after one self-loop, we have a signature of arity at least one with 0 as its first entry but is not identically 0, so we are done by Lemma 4.19.

   Thus, in this case, we may assume f has the form given in (4.1).

3. Suppose the binary prefix of f is $[1,0,0]$. Since f is not degenerate, f is not of the form $[1,0,\ldots,0]$. Suppose the second non-zero entry is $f_i = x \neq 0$ where $i \geqslant 3$. Then after connecting $k-i$ copies of $[1,0]$, where arity$(f) = k$, we have $[1,0,\ldots,0,x]$ of arity i. If $x^4 \neq 1$, then Pl-Holant$(\{f\} \cup \widehat{\mathcal{EQ}})$ is #**P**-hard by Lemma 4.8, so Pl-Holant$(\mathcal{F} \cup \widehat{\mathcal{EQ}})$ is also #**P**-hard.

Otherwise, $x^4 = 1$. If $f = [1, 0, \ldots, 0, x]$ with $x^4 = 1$, then $f \in \mathscr{A}$. Suppose that x is not the last entry in f. Connecting $k - i - 1$ copies of $[1, 0]$ to f, we have $g = [1, 0, \ldots, 0, x, y]$ of arity $i + 1$.

- If i is odd, then doing $\frac{i-3}{2}$ many self-loops, we have $h = [1, 0, 0, x, y]$. The determinant of the compressed signature matrix of h is $-2x^2 \neq 0$. Thus, Holant(h) is #**P**-hard by Corollary 3.28, so Pl-Holant($\mathcal{F} \cup \widehat{\mathcal{EQ}}$) is also #**P**-hard.
- Otherwise, i is even. After $\frac{i-4}{2}$ many self-loops on g, we have $h = [1, 0, 0, 0, x, y]$. Then by Lemma 4.15, Holant(h) is #**P**-hard, so Pl-Holant($\mathcal{F} \cup \widehat{\mathcal{EQ}}$) is also #**P**-hard.

Thus, in this case, we may assume that $f = [1, 0, \ldots, 0, x]$ with $x^4 = 1$.

4. Suppose the binary prefix of f is $[1, r, r^2]$, where $r \neq 0$. Since f is non-degenerate, f is not of the form $[1, r, \ldots, r^n]$. Suppose the first term that breaks the pattern is $f_{m+1} = y \neq r^{m+1}$ with $m \geqslant 2$. Connecting $k - m - 1$ many copies of $[1, 0]$, where arity(f) $= k$, we have $[1, r, \ldots, r^m, y]$. Using $[1, 0]$, we can get $[1, r]$. If $r = \pm i$, then we are done by Lemma 4.21, so assume that $r \neq \pm i$. Then we can attach $[1, r]$ back to the initial signature $k - 3$ times to get $g = [1, r, r^2, x]$ after normalization, where $x \neq r^3$. We connect $[1, r]$ once more to get $h = [1 + r^2, r(1 + r^2), r^2 + rx]$. If h does not have one of the exceptional forms in Lemma 4.20, then we are done, so assume that it does.

   Since the second entry of h is not 0 and $x \neq r^3$, the only possibility is that h has the form $[1, b, 1]$ up to a scalar. This gives $x = r^{-1}$. Note that $r \neq \pm 1$ since $x \neq r^3$. A self-loop on $g = [1, r, r^2, r^{-1}]$ gives $[1 + r^2, r + r^{-1}]$, which is $[1, r^{-1}]$ after normalization. Connecting this back to g gives $h' = [2, 2r, r^2 + r^{-2}]$. We assume that h' has one of the exceptional forms in Lemma 4.20 since we are done otherwise. If h' has the form $[1, r, r^2]$ up to a scalar, then $r^4 = 1$, a contradiction, so it must have the form $[1, b, 1]$ up to a scalar. But then $r^2 = 1$, which is also a contradiction.

   Thus, in this case, the theorem holds.

5. Suppose the binary prefix of f is $[1, b, 1]$. If $b = \pm 1$, then this binary prefix is degenerate and was considered in the previous case, so assume that $b \neq \pm 1$. If f is not of the form $[1, b, 1, b, \ldots]$, then let i be the index of the first entry in f to break the pattern. If i is

even, connecting $k-i$ copies of $[1,0]$ to f, where $k = \text{arity}(f)$, we have $[1,b,1,\ldots,b,y]$ with $y \neq 1$. We do $\frac{i-4}{2}$ more self-loops. After normalization, we get $g = [1,b,1,b,x]$, where $x \neq 1$. The determinant of its compressed signature matrix is $(b^2-1)(1-x) \neq 0$. Thus, $\text{Holant}(g)$ is #**P**-hard by Corollary 3.28, so Pl-Holant$(\mathcal{F} \cup \widehat{\mathcal{EQ}})$ is also #**P**-hard.

Otherwise, i is odd. Then connect $k-i$ many $[1,0]$ to f, and we get $[1,b,1,\ldots,1,y]$ with $y \neq b$. We do $\frac{i-3}{2}$ self-loops. After normalization, we get $[1,b,1,x]$, where $x \neq b$. One more self-loop gives us $g' = [2, b+x]$.

- If $b = 0$, then connecting g' to $[1,0,1,x]$ gives $[2,x,2+x^2]$. We assume that $[2,x,2+x^2]$ has one of the exceptional forms in Lemma 4.20 since we are done otherwise. Because $x \neq 0$, the only possibility is that $[2,x,2+x^2]$ has the form $[1,r,r^2]$ up to a scalar. Then we get $x^2 = -4$, so $g = [2,x] = 2[1,\pm i]$ and we are done by Lemma 4.21.

- Otherwise, $b \neq 0$. Using $[1,0]$, we can get $h = [1,b,1]$. If the signature matrix M_h of h has infinite order modulo a scalar, then we can interpolate $[0,1]$ by Lemma 4.7 since $b \notin \{0,\pm 1\}$, and we are done.

  Hence we may assume that M_h has finite order modulo a scalar. There exists positive integer ℓ and $\beta \neq 0$ such that $M_h^{\ell} = \beta I_2$. Thus after normalization, we can construct the anti-gadget $[1,-b,1]$ by connecting $\ell - 1$ copies of h together. Connecting $[1,0]$ to $[1,-b,1]$ gives $[1,-b]$ and connecting this to $[1,b,1,x]$ gives $[1-b^2, 0, 1-bx]$.

  If $\frac{1-bx}{1-b^2} \notin \{0,\pm 1\}$, then we are done by Lemma 4.18. Otherwise, $y = \frac{1-bx}{1-b^2} \in \{0,\pm 1\}$. For $y = 0$, we get $x = b^{-1}$ and are done by Lemma 4.22 since $b \notin \{0,\pm 1\}$. For $y = 1$, we get $b = x$, a contradiction. For $y = -1$, we get $2 - b^2 - bx = 0$. Then connecting $[1,-b,1]$ to $g = [2, b+x]$ gives $[2-b^2-bx, x-b] = [0, x-b]$, which is $[0,1]$ after normalization.

Thus, in this case, we may assume that $f = [1,b,1,b,\ldots]$.

To summarize, every non-degenerate signature in $\mathcal{F}$ must have one of the following forms:

- $[1,0,-1,0,1,0,-1,0,\ldots,0 \text{ or } 1 \text{ or } (-1)]$, which is in $\mathcal{A} \cap \mathcal{M}$;

- $[1,0,\ldots,0,x]$, where $x^4 = 1$, which is in $\mathcal{A}$;

- $[1, b, 1, b, \dots, 1 \text{ or } b]$, which is in $\widehat{\mathcal{P}}$.

Moreover, as unary signatures are all in $\widehat{\mathcal{P}}$, we have that $\mathcal{F} \subseteq \mathcal{A} \cup \widehat{\mathcal{P}} \cup \mathcal{M}$. We are done by Theorem 4.14. □

4.4 Planar #CSP Dichotomy

In this section, we prove our main dichotomy theorem. We begin with a dichotomy for a single signature.

Theorem 4.24. *Let* f *be a non-degenerate symmetric signature of arity* $n \geqslant 2$ *with complex weights in Boolean variables. Then* Pl-Holant$(\{f\} \cup \widehat{\mathcal{EQ}})$ *is* #**P***-hard unless* $f \in \mathcal{A} \cup \widehat{\mathcal{P}} \cup \mathcal{M}$*, in which case the problem is computable in polynomial time.*

Proof. When $f \in \mathcal{A} \cup \widehat{\mathcal{P}} \cup \mathcal{M}$, the problem is tractable by Lemma 4.3. When $f \notin \mathcal{A} \cup \widehat{\mathcal{P}} \cup \mathcal{M}$, we prove that Pl-Holant$^c(\{f\} \cup \widehat{\mathcal{EQ}})$ is #**P**-hard, which is sufficient because of pinning (Theorem 4.23). Using $[1, 0]$ and $[0, 1]$, we can obtain any subsignature of f.

Notice that once we have $[0, 1]$ and $\widehat{\mathcal{EQ}}$, we can realize every signature in $T\widehat{\mathcal{EQ}}$, where $T = \left[\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right]$. In fact, every even arity signature in $\widehat{\mathcal{EQ}}$ is also in $T\widehat{\mathcal{EQ}}$, and we obtain all the odd arity signatures in $T\widehat{\mathcal{EQ}}$ by attaching $[0, 1]$ to all the even arity signatures in $\widehat{\mathcal{EQ}}$. Therefore, a holographic transformation by T does not change the complexity of the problem. Furthermore, $\mathcal{A} \cup \widehat{\mathcal{P}} \cup \mathcal{M}$ is closed under T. We will use these facts later.

The possibilities for f can be divided into three cases:

- f satisfies the parity condition;
- f does not satisfy the parity condition but does contain a 0 entry;
- f does not contain a 0 entry.

We handle these cases below.

1. Suppose that f satisfies the parity condition. If f has even parity, then we are done by Lemma 4.16. Otherwise, f has odd parity.

- If n, the arity of f, is odd, then under a holographic transformation by $T = \left[\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right]$, f is transformed to $\hat{f}$, which has even parity. Then either Pl-Holant$^c(\{\hat{f}\} \cup \widehat{\mathcal{EQ}})$ is #**P**-hard by Lemma 4.16 (and thus Pl-Holant$^c(\{f\} \cup \widehat{\mathcal{EQ}})$ is also #**P**-hard), or $\hat{f} \in \mathscr{A} \cup \widehat{\mathscr{P}} \cup \mathscr{M}$ (and thus $f \in \mathscr{A} \cup \widehat{\mathscr{P}} \cup \mathscr{M}$).
- Otherwise, n is even. Connect $[0, 1]$ to f to get a signature g with even parity and odd arity. Then either Pl-Holant$^c(\{g\} \cup \widehat{\mathcal{EQ}})$ is #**P**-hard by Lemma 4.16 (and thus Pl-Holant$^c(\{f\} \cup \widehat{\mathcal{EQ}})$ is also #**P**-hard), or $g \in \mathscr{A} \cup \widehat{\mathscr{P}} \cup \mathscr{M}$. In the latter case, it must be that $g \in \mathscr{M}$ since non-degenerate generalized equality signatures cannot have both even parity and odd arity (cf. Figure 4.1). In particular, the even parity entries of g form a geometric progression. Therefore $f \in \mathscr{M}$ since f has odd parity and the same geometric progression among its odd parity entries.

2. Suppose that f does not satisfy the parity condition but contains a 0 entry. Since f does not satisfy the parity condition, there must be at least two nonzero entries separated by an even number of 0 entries. Thus, f contains a subsignature $g = [a, 0, \ldots, 0, b]$ of arity $k = 2\ell + 1 \geqslant 1$, where $ab \neq 0$.

   If $\ell = 0$, then $k = 1$ and we can shift either to the right or to the left and find the 0 entry in f and obtain a binary subsignature h of the form $[a, b, 0]$ or $[0, a, b]$, where $ab \neq 0$. Then Pl-Holant$\left(h \mid \widehat{\mathcal{EQ}}\right)$ is #**P**-hard by Theorem 4.2, so Pl-Holant$(\{f\} \cup \widehat{\mathcal{EQ}})$ is also #**P**-hard.

   Otherwise $\ell \geqslant 1$, so $k \geqslant 3$. If $a^4 \neq b^4$, then Pl-Holant$(\{g\} \cup \widehat{\mathcal{EQ}})$ is #**P**-hard by Lemma 4.8, so Pl-Holant$(\{f\} \cup \widehat{\mathcal{EQ}})$ is also #**P**-hard.

   Otherwise, $a^4 = b^4$, so $g \in \mathscr{A}$. If $f = g$, then we are done, so assume that $f \neq g$, which implies that there is another entry just before a or just after b. If this entry is nonzero, then f has a subsignature h of the form $[c, a, 0]$ or $[0, b, d]$, where $cd \neq 0$. Then Pl-Holant$\left(h \mid \widehat{\mathcal{EQ}}\right)$ is #**P**-hard by Theorem 4.2, so Pl-Holant$(\{f\} \cup \widehat{\mathcal{EQ}})$ is also #**P**-hard.

   Otherwise, this entry is 0 and f has a subsignature h of the form $[0, a, 0, \ldots, 0, b]$ or $[a, 0, \ldots, 0, b, 0]$ of arity at least 4. If the arity of h is even, then after some number of self-loops, we have a signature h' of the form $[0, a, 0, 0, b]$ or $[a, 0, 0, b, 0]$ of arity exactly 4. Then Pl-Holant(h') is #**P**-hard by Corollary 3.28 since $ab \neq 0$, so Pl-Holant$(\{f\} \cup \widehat{\mathcal{EQ}})$ is also

#**P**-hard.

Otherwise, the arity of h is odd. After some number of self-loops, we have a signature h' of the form $[0, a, 0, 0, 0, b]$ or $[a, 0, 0, 0, b, 0]$ of arity exactly 5. Then Pl-Holant(h') is #**P**-hard by Lemma 4.15 since $ab \neq 0$, so Pl-Holant($\{f\} \cup \widehat{\mathcal{EQ}}$) is also #**P**-hard.

3. Suppose f contains no 0 entry. If f has a binary subsignature g such that Pl-Holant $\left(g \mid \widehat{\mathcal{EQ}}\right)$ is #**P**-hard by Theorem 4.2, then Pl-Holant($\{f\} \cup \widehat{\mathcal{EQ}}$) is also #**P**-hard.

   Otherwise every binary subsignature $[a, b, c]$ of f satisfies the conditions of some tractable case in Theorem 4.2. The three possible tractable cases are degenerate with condition $ac = b^2$ (case 1), affine $\mathcal{A}$ with condition $ac = -b^2 \wedge a = -c$ (case 3), and a Hadamard-transformed product type $\widehat{\mathcal{P}}$ with condition $a = c$ (case 4). If every binary subsignature $[a, b, c]$ of f satisfies $ac = b^2$, then f is degenerate, a contradiction. If every binary subsignature $[a, b, c]$ of f satisfies $ac = -b^2 \wedge a = -c$, then $f = [1, \pm 1, -1, \mp 1, 1, \pm 1, -1, \mp 1, \dots] \in \mathcal{A}$ (up to a scalar) and we are done. If every binary subsignature $[a, b, c]$ of f satisfies $a = c$, then $f \in \widehat{\mathcal{P}}$ and we are done.

   Otherwise, there exists two binary subsignatures of f that do not satisfy the same tractable case in Theorem 4.2. Hence f has arity $n \geqslant 3$. Let $h_i = [f_i, f_{i+1}, f_{i+2}]$ for all $0 \leqslant i \leqslant n-2$ be binary subsignatures of f. Suppose there exists an i such that h_i satisfies the affine condition (case 3). We claim that there must exist two successive signatures $h = [a, b, c]$ that is affine and $h' = [b, c, d]$ satisfying either the degenerate or the product-type condition, up to a transformation of $\left[\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right]$. This is because we can start from h_i and search in both directions h_{i-1} and h_{i+1} until we found such h and h'. It is always successful because not all h_i satisfies the affine condition. Let $g = [a, b, c, d]$ be the tenary subsignature of f. Then for either case of h', we have $g = [1, \varepsilon, -1, \varepsilon]$ after normalization, where $\varepsilon^2 = 1$. Connecting two copies of $[0, 1]$ to g gives $[-1, \varepsilon]$. Connecting this back to g gives $g' = [0, -2\varepsilon, 2]$. Then Pl-Holant $\left(g' \mid \widehat{\mathcal{EQ}}\right)$ is #**P**-hard by Theorem 4.2, so Pl-Holant($\{f\} \cup \widehat{\mathcal{EQ}}$) is also #**P**-hard.

   Otherwise, up to the transformation $\left[\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right]$, there exists a ternary subsignature $g = [a, b, c, d]$ of f such that $h = [a, b, c]$ satisfies the product-type condition (but not the degenerate condition) and $h' = [b, c, d]$ satisfies the degenerate condition. Then $g = [1, b, 1, b^{-1}]$ after normalization, where $b^2 \neq 1$. Then $\text{Pl-Holant}\left(g \mid \widehat{\mathcal{EQ}}\right)$ is #**P**-hard by Lemma 4.22, so $\text{Pl-Holant}(\{f\} \cup \widehat{\mathcal{EQ}})$ is also #**P**-hard. □

Now we are ready to prove our main dichotomy theorem.

Theorem 4.25. *Let $\mathcal{F}$ be any set of symmetric, complex-valued signatures in Boolean variables. Then* $\text{Pl-Holant}(\mathcal{F} \cup \widehat{\mathcal{EQ}})$ *is #**P**-hard unless $\mathcal{F} \subseteq \mathcal{A}$, $\mathcal{F} \subseteq \widehat{\mathcal{P}}$, or $\mathcal{F} \subseteq \mathcal{M}$, in which case the problem is computable in polynomial time.*

Proof. The tractability is given in Lemma 4.3. When $\mathcal{F}$ is not a subset of $\mathcal{A}$, $\widehat{\mathcal{P}}$, or $\mathcal{M}$, we prove that $\text{Pl-Holant}^c(\mathcal{F} \cup \widehat{\mathcal{EQ}})$ is #**P**-hard, which is sufficient because of pinning (Theorem 4.23).

For any degenerate signature $f \in \mathcal{F}$, we connect some number of $[1, 0]$ or $[0, 1]$ to f to get its corresponding unary signature. We replace f by this unary signature, which does not change the complexity. Thus, assume that the only degenerate signatures in $\mathcal{F}$ are unary signatures.

If $\mathcal{F} \not\subseteq \mathcal{A} \cup \widehat{\mathcal{P}} \cup \mathcal{M}$, then the problem is #**P**-hard by Theorem 4.24. Otherwise, $\mathcal{F} \subseteq \mathcal{A} \cup \widehat{\mathcal{P}} \cup \mathcal{M}$ and we are done by Theorem 4.14. □

Transforming back to the standard basis, Theorem 4.25 is equivalent to Theorem 4.1.

Chapter 5

A Closer Look at Tractable Signatures

Aside from vanishing signatures, major tractable sets are $\mathscr{A}$-transformable in Lemma 1.7, $\mathscr{P}$-transformable in Lemma 1.9, and $\mathscr{M}$-transformable in Lemma 1.10. In this chapter, we first characterize these three sets, and then show various hardness results related to them. After we finish these hardness results, we will utilize them to show Theorem 5.41, the single signature dichotomy for Holant and Pl-Holant.

Throughout the chapter, we define $\alpha = \frac{1+i}{\sqrt{2}} = \sqrt{i} = e^{\frac{\pi i}{4}}$ and use $\mathbf{O}_2(\mathbb{C})$ to denote the group of 2-by-2 orthogonal matrices over $\mathbb{C}$. While the main results in this section assume that the signatures involved are symmetric, we note that some of the lemmas also hold without this assumption. We use $\mathcal{F}_{123}$ to denote $\mathcal{F}_1 \cup \mathcal{F}_2 \cup \mathcal{F}_3$, where $\mathcal{F}_1$, $\mathcal{F}_2$, and $\mathcal{F}_3$ are defined in (1.4).

Recall that by Definition 1.3, if a set of signatures $\mathcal{F}$ is $\mathcal{C}$-transformable, then there exists a $T \in \mathbf{GL}_2(\mathbb{C})$ such that $[1,0,1]T^{\otimes 2} \in \mathcal{C}$ and $\mathcal{F} \subseteq T\mathcal{C}$. We first consider possible transformations such that $[1,0,1]T^{\otimes 2} \in \mathscr{A}$, $\mathscr{P}$, or $\mathscr{M}$. While there are many binary signatures in $\mathscr{A}$, $\mathscr{P}$, and $\mathscr{M}$, it turns out that it is sufficient to consider the following cases.

Proposition 5.1. *Let* $T \in \mathbb{C}^{2\times 2}$ *be a matrix. Then the following hold:*

1. $[1,0,1]T^{\otimes 2} = [1,0,1]$ *if and only if* $T \in \mathbf{O}_2(\mathbb{C})$*;*
2. $[1,0,1]T^{\otimes 2} = [1,0,i]$ *if and only if there exists an* $H \in \mathbf{O}_2(\mathbb{C})$ *such that* $T = H\left[\begin{smallmatrix} 1 & 0 \\ 0 & \alpha \end{smallmatrix}\right]$*;*
3. $[1,0,1]T^{\otimes 2} = [0,1,0]$ *if and only if there exists an* $H \in \mathbf{O}_2(\mathbb{C})$ *such that* $T = \frac{1}{\sqrt{2}}H\left[\begin{smallmatrix} 1 & 1 \\ i & -i \end{smallmatrix}\right]$.

Proof. Recall (1.3). Case 1 is clear due to

$$[1,0,1]T^{\otimes 2} = [1,0,1] \iff T^{\mathsf{T}} I_2 T = I_2 \iff T^{\mathsf{T}} T = I_2,$$

the last of which is the definition of a (2-by-2) orthogonal matrix. Now we use this case to prove the others.

Let $D_2 = \left[\begin{smallmatrix} 1 & 0 \\ 0 & \alpha \end{smallmatrix}\right]$ and $D_3 = Z = \frac{1}{\sqrt{2}}\left[\begin{smallmatrix} 1 & 1 \\ i & -i \end{smallmatrix}\right]$, let $T_j = HD_j$ (for $j = 2, 3$), where $H \in \mathbf{O}_2(\mathbb{C})$. Then we can directly verify that,

$$[1,0,1]T_j^{\otimes 2} = [1,0,1](HD_j)^{\otimes 2} = [1,0,1]D_j^{\otimes 2} = f_j,$$

where $f_2 = [1,0,i]$ and $f_3 = [0,1,0]$ are the binary signature in case 2 and 3.

On the other hand, suppose that $[1,0,1](T_j)^{\otimes 2} = f_j$, for $j = 2, 3$. Then we have that

$$[1,0,1](T_j D_j^{-1})^{\otimes 2} = f_j (D_j^{-1})^{\otimes 2} = [1,0,1],$$

so $H = T_j D_j^{-1} \in \mathbf{O}_2(\mathbb{C})$ by case 1. Thus $T_j = HD_j$ as desired. □

We also need the following lemma. Its proof is straightforward.

Lemma 5.2. *If a symmetric signature* $f = [f_0, f_1, \ldots, f_n]$ *can be expressed in the form* $f = a[1,\lambda]^{\otimes n} + b[1,\mu]^{\otimes n}$, *for some* $a, b, \lambda, \mu \in \mathbb{C}$, *then the* f_k*'s satisfy the recurrence relation* $f_{k+2} = (\lambda+\mu)f_{k+1} - \lambda\mu f_k$ *for* $0 \leqslant k \leqslant n-2$.

5.1 Characterization of $\mathcal{A}$-transformable Signatures

We start with $\mathcal{A}$-transformable signatures. We introduce the left and right stabilizer groups of $\mathcal{A}$:

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

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

In fact, these two groups are equal and coincide with the group of nonsingular signature matrices of binary affine signatures. Recall (1.2). For a binary signature $f = (f^{00}, f^{01}, f^{10}, f^{11})$, its signature matrix M_f is

$$M_f = \begin{bmatrix} f^{00} & f^{01} \\ f^{10} & f^{11} \end{bmatrix}.$$

Let

$$\mathcal{A}^{2\times 2} = \{M_f \mid f \in \mathcal{A},\ \text{arity}(f) = 2,\ \text{and}\ \det(M_f) \neq 0\}$$

be the set of nonsingular signature matrices of the binary affine signatures. It is straightforward to verify that $\mathcal{A}^{2\times 2}$ is closed under multiplication and inverses. Therefore $\mathcal{A}^{2\times 2}$ forms a group.

Let $D_2 = \left[\begin{smallmatrix} 1 & 0 \\ 0 & i \end{smallmatrix}\right]$ and $H_2 = \frac{1}{\sqrt{2}}\left[\begin{smallmatrix} 1 & 1 \\ 1 & -1 \end{smallmatrix}\right]$. Also let $X = \left[\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right]$ and $Z = \frac{1}{\sqrt{2}}\left[\begin{smallmatrix} 1 & 1 \\ i & -i \end{smallmatrix}\right]$. Note that $Z = D_2 H_2$ and that $D_2^2 Z = \frac{1}{\sqrt{2}}\left[\begin{smallmatrix} 1 & 1 \\ -i & i \end{smallmatrix}\right] = ZX$. Hence $X = Z^{-1} D_2^2 Z$. Furthermore, $D_2, H_2, X, Z \in \text{LStab}(\mathcal{A}) \cap \text{RStab}(\mathcal{A}) \cap \mathcal{A}^{2\times 2}$, as well as all nonzero scalar multiples of these matrices.

Not only are the groups $\text{LStab}(\mathcal{A})$, $\text{RStab}(\mathcal{A})$, and $\mathcal{A}^{2\times 2}$ equal, they are generated by D_2 and H_2 with a nonzero scalar multiple.

Lemma 5.3. $\text{LStab}(\mathcal{A}) = \text{RStab}(\mathcal{A}) = \mathcal{A}^{2\times 2} = \mathbb{C}^* \cdot \langle D_2, H_2 \rangle$.

Proof. Let

$$\mathbf{S} := \{S \in \mathbf{GL}_2(\mathbb{C}) \mid \mathcal{F}_{123} S \subseteq \mathcal{F}_{123}\}$$

be the right stabilizer group of $\mathcal{F}_{123}$. Since $\mathcal{F}_{123} \subset \mathcal{A}$, and symmetric signatures are still symmetric under any transformation, we have that $\text{RStab}(\mathcal{A}) \subseteq \mathbf{S}$. Moreover, as $\mathcal{A}$ is closed under gadget construction, $\mathcal{A}^{2\times 2} \subseteq \text{RStab}(\mathcal{A})$. Hence, $\mathcal{A}^{2\times 2} \subseteq \text{RStab}(\mathcal{A}) \subseteq \mathbf{S}$. Together with the fact that $D_2, H_2 \in \mathcal{A}^{2\times 2}$, we have $\mathbb{C}^* \cdot \langle D_2, H_2 \rangle \subseteq \mathcal{A}^{2\times 2} \subseteq \text{RStab}(\mathcal{A}) \subseteq \mathbf{S}$. To finish the proof, we show that $\mathbf{S} \subseteq \mathbb{C}^* \cdot \langle D_2, H_2 \rangle$. For $\text{LStab}(\mathcal{A})$, the proof is similar.

Let T be an arbitrary element in $\mathbf{S}$. For $f = (=_3)$, we have $fT^{\otimes 3} \in \mathcal{F}_{123}$. Then by the form of $\mathcal{F}_{123}$, for some $M \in \langle D_2, H_2 \rangle$, chosen to be either I, or $H_2^{\mathsf{T}} = H_2$, or $Z^{\mathsf{T}} = H_2 D_2$, we have $f(TM^{-1})^{\otimes 3} \in \mathcal{F}_1$, which is a generalized equality signature. Then either TM^{-1} or $TM^{-1}X$ is a diagonal matrix $T' = \lambda \left[\begin{smallmatrix} 1 & 0 \\ 0 & d \end{smallmatrix}\right]$. Furthermore, by applying T' to $=_4$, we conclude that $(=_4)T'^{\otimes 4} \in \mathcal{F}_1$, since it is in $\mathcal{F}_{123}$ but not in $\mathcal{F}_2 \cup \mathcal{F}_3$ because T' is diagonal. It follows that d is a power of i, and hence $\left[\begin{smallmatrix} 1 & 0 \\ 0 & d \end{smallmatrix}\right]$ is a power of D_2. Thus $T \in \mathbb{C}^* \cdot \langle D_2, H_2 \rangle$. □

Since $\text{LStab}(\mathcal{A}) = \text{RStab}(\mathcal{A})$, we simply write $\text{Stab}(\mathcal{A})$ for this group. Of course each T under which $\mathcal{F}$ is $\mathcal{A}$-transformable is just a particular transformation that can be extended by any element in $\text{Stab}(\mathcal{A})$.

Lemma 5.4. *Let $\mathcal{F}$ be a set of signatures. Then $\mathcal{F}$ is $\mathcal{A}$-transformable under T if and only if $\mathcal{F}$ is $\mathcal{A}$-transformable under any $T' \in T\,\text{Stab}(\mathcal{A})$.*

Proof. Sufficiency is trivial since $I_2 \in \text{Stab}(\mathcal{A})$. If $\mathcal{F}$ is $\mathcal{A}$-transformable under T, then by definition, we have $(=_2)T^{\otimes 2} \in \mathcal{A}$ and $\mathcal{F}' = T^{-1}\mathcal{F} \subseteq \mathcal{A}$. Let $T' = TM \in T\,\text{Stab}(\mathcal{A})$ for any $M \in \text{Stab}(\mathcal{A})$. It then follows that $(=_2)T'^{\otimes 2} = (=_2)T^{\otimes 2}M^{\otimes 2} \in \mathcal{A}M = \mathcal{A}$ and $T'^{-1}\mathcal{F} = M^{-1}\mathcal{F}' \subseteq M^{-1}\mathcal{A} = \mathcal{A}$. Therefore $\mathcal{F}$ is $\mathcal{A}$-transformable under any $T' \in T\,\text{Stab}(\mathcal{A})$. □

After restricting by Proposition 5.1 and normalizing by Lemma 5.4, one only needs to check a small subset of $\mathbf{GL}_2(\mathbb{C})$ to determine if $\mathcal{F}$ is $\mathcal{A}$-transformable.

Lemma 5.5. *Let $\mathcal{F}$ be a set of signatures. Then $\mathcal{F}$ is $\mathcal{A}$-transformable if and only if there exists an $H \in \mathbf{O}_2(\mathbb{C})$ such that $\mathcal{F} \subseteq H\mathcal{A}$ or $\mathcal{F} \subseteq H \left[\begin{smallmatrix} 1 & 0 \\ 0 & \alpha \end{smallmatrix}\right] \mathcal{A}$.*

Proof. Sufficiency is easily verified by checking that $=_2$ is transformed into $\mathcal{A}$ in both cases. In particular, H leaves $=_2$ unchanged.

If $\mathcal{F}$ is $\mathcal{A}$-transformable, then by definition, there exists a matrix T such that $(=_2)T^{\otimes 2} \in \mathcal{A}$ and $T^{-1}\mathcal{F} \subseteq \mathcal{A}$. Since $=_2$ is non-degenerate and symmetric, $(=_2)T^{\otimes 2} \in \mathcal{A}$ is equivalent to $(=_2)T^{\otimes 2} \in \mathcal{F}_{123}$.

Any signature in $\mathcal{F}_{123}$ is expressible as $c(v_1^{\otimes n} + i^t v_2^{\otimes n})$, where $t \in \{0, 1, 2, 3\}$ and (v_1, v_2) is a pair of vectors in the set

$$\left\{ \left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right), \left(\begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \end{bmatrix} \right), \left(\begin{bmatrix} 1 \\ i \end{bmatrix}, \begin{bmatrix} 1 \\ -i \end{bmatrix} \right) \right\}.$$

We use $\text{Stab}(\mathcal{A})$ to further normalize these three sets by Lemma 5.4. In particular, $\mathcal{F}_1 = H_2\mathcal{F}_2$ and $\mathcal{F}_1 = (D_2H_2)^{-1}\mathcal{F}_3$. Furthermore, binary signatures in $\mathcal{F}_1$ are just the four signatures $[1,0,1]$, $[1,0,i]$, $[1,0,-1]$, and $[1,0,-i]$ up to a scalar. We also normalize these four as $[1,0,1] = [1,0,-1]D_2^{\otimes 2}$ and $[1,0,i] = [1,0,-i]D_2^{\otimes 2}$. Hence $\mathcal{F}$ being $\mathcal{A}$-transformable implies that there exists a matrix T such that $(=_2)T^{\otimes 2} \in \{[1,0,1],[1,0,i]\}$ and $T^{-1}\mathcal{F} \subseteq \mathcal{A}$. Now we apply Proposition 5.1.

1. If $(=_2)T^{\otimes 2} = [1,0,1]$, then by case 1 of Proposition 5.1, we have $T \in \mathbf{O}_2(\mathbb{C})$. Therefore $\mathcal{F} \subseteq H\mathcal{A}$ where $H = T \in \mathbf{O}_2(\mathbb{C})$.

2. If $(=_2)T^{\otimes 2} = [1,0,i]$, then by case 2 of Proposition 5.1, there exists an $H \in \mathbf{O}_2(\mathbb{C})$ such that $T = H\left[\begin{smallmatrix}1 & 0\\ 0 & \alpha\end{smallmatrix}\right]$. Therefore $\mathcal{F} \subseteq T\mathcal{A} = H\left[\begin{smallmatrix}1 & 0\\ 0 & \alpha\end{smallmatrix}\right]\mathcal{A}$ where $H \in \mathbf{O}_2(\mathbb{C})$. □

Using these two lemmas, we can characterize all $\mathcal{A}$-transformable signatures. We first define three sets $\mathcal{A}_1$, $\mathcal{A}_2$, and $\mathcal{A}_3$.

Definition 5.6. *A symmetric signature* f *of arity* n *is in* $\mathcal{A}_1$ *if there exists an* $H \in \mathbf{O}_2(\mathbb{C})$ *and a nonzero constant* $c \in \mathbb{C}$ *such that* $f = cH^{\otimes n}\left(\left[\begin{smallmatrix}1\\ 1\end{smallmatrix}\right]^{\otimes n} + \beta\left[\begin{smallmatrix}1\\ -1\end{smallmatrix}\right]^{\otimes n}\right)$, *where* $\beta = \alpha^{tn+2r}$ *for some* $r \in \{0,1,2,3\}$ *and* $t \in \{0,1\}$.

When such an H exists, we say that $f \in \mathcal{A}_1$ with transformation H. If $f \in \mathcal{A}_1$ with I_2, then we say f is in the canonical form of $\mathcal{A}_1$. If f is in the canonical form of $\mathcal{A}_1$, then by Lemma 5.2, for any $0 \leqslant k \leqslant n-2$, we have $f_{k+2} = f_k$ and one of the following holds:

- $f_0 = 0$, or
- $f_1 = 0$, or
- $f_1 = \pm i f_0 \neq 0$, or
- n is odd and $f_1 = \pm(1 \pm \sqrt{2}) i f_0 \neq 0$ (all four sign choices are permissible).

Notice that when n is odd and $t = 1$ in Definition 5.6, it has some complication as described by the factor α^{tn+2r}.

Definition 5.7. *A symmetric signature* f *of arity* n *is in* $\mathcal{A}_2$ *if there exists an* $H \in \mathbf{O}_2(\mathbb{C})$ *and a nonzero constant* $c \in \mathbb{C}$ *such that* $f = cH^{\otimes n}\left(\left[\begin{smallmatrix}1\\ i\end{smallmatrix}\right]^{\otimes n} + \left[\begin{smallmatrix}1\\ -i\end{smallmatrix}\right]^{\otimes n}\right)$.

Similarly, when such an H exists, we say that $f \in \mathcal{A}_2$ with transformation H. If $f \in \mathcal{A}_2$ with I_2, then we say that f is in the canonical form of $\mathcal{A}_2$. If f is in the canonical form of $\mathcal{A}_2$, then by Lemma 5.2, for any $0 \leqslant k \leqslant n-2$, we have $f_{k+2} = -f_k$. Since f is non-degenerate, $f_1 \neq \pm i f_0$ is implied.

It is worth noting that $\{\left[\begin{smallmatrix}1\\i\end{smallmatrix}\right], \left[\begin{smallmatrix}1\\-i\end{smallmatrix}\right]\}$ is setwise invariant up to scalars under any transformation in $\mathbf{O}_2(\mathbb{C})$ up to nonzero constants. In other words, these two vectors are the eigenvectors of orthogonal matrices. Thus for any $H \in \mathbf{O}_2(\mathbb{C})$, we have that $\left[\begin{smallmatrix}1&1\\i&-i\end{smallmatrix}\right]^{-1} H \left[\begin{smallmatrix}1&1\\i&-i\end{smallmatrix}\right] = D$, where D is either diagonal or anti-diagonal. It is also helpful to view this equation as $H\left[\begin{smallmatrix}1&1\\i&-i\end{smallmatrix}\right] = \left[\begin{smallmatrix}1&1\\i&-i\end{smallmatrix}\right] D$.

Using this fact, the following lemma gives a characterization of $\mathcal{A}_2$. It says that any signature in $\mathcal{A}_2$ is essentially in the canonical form.

Lemma 5.8. *Let* f *be a symmetric signature of arity* n*. Then* $f \in \mathcal{A}_2$ *if and only if* $f = c\left(\left[\begin{smallmatrix}1\\i\end{smallmatrix}\right]^{\otimes n} + \beta\left[\begin{smallmatrix}1\\-i\end{smallmatrix}\right]^{\otimes n}\right)$ *for some nonzero constants* $c, \beta \in \mathbb{C}$*.*

Proof. Assume that $f = c\left(\left[\begin{smallmatrix}1\\i\end{smallmatrix}\right]^{\otimes n} + \beta\left[\begin{smallmatrix}1\\-i\end{smallmatrix}\right]^{\otimes n}\right)$ for some $c, \beta \neq 0$. Consider the orthogonal transformation $H = \left[\begin{smallmatrix}a&b\\b&-a\end{smallmatrix}\right]$, where $a = \frac{1}{2}\left(\beta^{\frac{1}{2n}} + \beta^{-\frac{1}{2n}}\right)$ and $b = \frac{1}{2i}\left(\beta^{\frac{1}{2n}} - \beta^{-\frac{1}{2n}}\right)$. We pick a and b in this way so that $a + bi = \beta^{\frac{1}{2n}}$, $a - bi = \beta^{-\frac{1}{2n}}$, and $(a+bi)(a-bi) = a^2 + b^2 = 1$. Also $\left(\frac{a+bi}{a-bi}\right)^n = \beta$. Then

$$\begin{aligned} H^{\otimes n} f &= c\left(\begin{bmatrix} a+bi \\ -ai+b \end{bmatrix}^{\otimes n} + \beta \begin{bmatrix} a-bi \\ ai+b \end{bmatrix}^{\otimes n}\right) \\ &= c\left((a+bi)^n \begin{bmatrix} 1 \\ -i \end{bmatrix}^{\otimes n} + (a-bi)^n \beta \begin{bmatrix} 1 \\ i \end{bmatrix}^{\otimes n}\right) \\ &= c\sqrt{\beta}\left(\begin{bmatrix} 1 \\ -i \end{bmatrix}^{\otimes n} + \begin{bmatrix} 1 \\ i \end{bmatrix}^{\otimes n}\right), \end{aligned}$$

so f can be written as

$$f = c\sqrt{\beta}(H^{-1})^{\otimes n}\left(\begin{bmatrix} 1 \\ i \end{bmatrix}^{\otimes n} + \begin{bmatrix} 1 \\ -i \end{bmatrix}^{\otimes n}\right).$$

Therefore $f \in \mathcal{A}_2$.

On the other hand, the desired form $f = c\left(\left[\begin{smallmatrix}1\\ i\end{smallmatrix}\right]^{\otimes n} + \beta\left[\begin{smallmatrix}1\\ -i\end{smallmatrix}\right]^{\otimes n}\right)$ follows from the fact that $\{\left[\begin{smallmatrix}1\\ i\end{smallmatrix}\right], \left[\begin{smallmatrix}1\\ -i\end{smallmatrix}\right]\}$ is fixed setwise up to nonzero constants under any orthogonal transformation. $\square$

Definition 5.9. *A symmetric signature* f *of arity* n *is in* $\mathcal{A}_3$ *if there exists an* $H \in \mathbf{O}_2(\mathbb{C})$ *and a nonzero constant* $c \in \mathbb{C}$ *such that* $f = cH^{\otimes n}\left(\left[\begin{smallmatrix}1\\ \alpha\end{smallmatrix}\right]^{\otimes n} + i^r\left[\begin{smallmatrix}1\\ -\alpha\end{smallmatrix}\right]^{\otimes n}\right)$ *for some* $r \in \{0, 1, 2, 3\}$.

Again, when such an H exists, we say that $f \in \mathcal{A}_3$ with transformation H. If $f \in \mathcal{A}_3$ with I_2, then we say f is in the canonical form of $\mathcal{A}_3$. If f is in the canonical form of $\mathcal{A}_3$, then by Lemma 5.2, for any $0 \leqslant k \leqslant n-2$, we have that $f_{k+2} = i f_k$ and one of the following holds:

- $f_0 = 0$, or
- $f_1 = 0$, or
- $f_1 = \pm\alpha i f_0 \neq 0$.

Now we characterize $\mathcal{A}$-transformable signatures.

Lemma 5.10. *Let* f *be a non-degenerate symmetric signature. Then* f *is* $\mathcal{A}$*-transformable if and only if* $f \in \mathcal{A}_1 \cup \mathcal{A}_2 \cup \mathcal{A}_3$.

Proof. Assume that f is $\mathcal{A}$-transformable of arity n. By applying Lemma 5.5 to $\{f\}$, there exists an $H \in \mathbf{O}_2(\mathbb{C})$ such that $f \in H\mathcal{A}$ or $f \in H\left[\begin{smallmatrix}1 & 0\\ 0 & \alpha\end{smallmatrix}\right]\mathcal{A}$. This is equivalent to $(H^{-1})^{\otimes n} f \in \mathcal{A}$ or $(H^{-1})^{\otimes n} f \in \left[\begin{smallmatrix}1 & 0\\ 0 & \alpha\end{smallmatrix}\right]\mathcal{A}$. Since f is non-degenerate and symmetric, we can replace $\mathcal{A}$ in the previous expressions with $\mathcal{F}_{123}$. Now we consider all possible cases. Let $\bar{f} = (H^{-1})^{\otimes n} f$.

1. If $\bar{f} \in \mathcal{F}_1$, then $T^{\otimes n}\bar{f}$ is in the canonical form of $\mathcal{A}_1$, where $T = \frac{1}{\sqrt{2}}\left[\begin{smallmatrix}1 & 1\\ 1 & -1\end{smallmatrix}\right] \in \mathbf{O}_2(\mathbb{C})$.
2. If $\bar{f} \in \mathcal{F}_2$, then $\bar{f}$ is already in the canonical form of $\mathcal{A}_1$. Let $T = I_2$ in this case.
3. If $\bar{f} \in \mathcal{F}_3$, then $\bar{f}$ already has the equivalent form of $\mathcal{A}_2$ given by Lemma 5.8. Let $T = I_2$ in this case.
4. If $\bar{f} \in \left[\begin{smallmatrix}1 & 0\\ 0 & \alpha\end{smallmatrix}\right]\mathcal{F}_1$, then $T^{\otimes n}\bar{f}$ is in the canonical form of $\mathcal{A}_1$, where $T = \frac{1}{\sqrt{2}}\left[\begin{smallmatrix}1 & 1\\ 1 & -1\end{smallmatrix}\right] \in \mathbf{O}_2(\mathbb{C})$.
5. If $\bar{f} \in \left[\begin{smallmatrix}1 & 0\\ 0 & \alpha\end{smallmatrix}\right]\mathcal{F}_2$, then $\bar{f}$ is already in the canonical form of $\mathcal{A}_3$. Let $T = I_2$ in this case.

6. If $\bar{f} \in \left[\begin{smallmatrix} 1 & 0 \\ 0 & \alpha \end{smallmatrix}\right] \mathcal{F}_3$, then $\bar{f}$ has the form $\left[\begin{smallmatrix} 1 \\ \alpha^3 \end{smallmatrix}\right]^{\otimes n} + i^r \left[\begin{smallmatrix} 1 \\ -\alpha^3 \end{smallmatrix}\right]^{\otimes n}$, and $T^{\otimes n}\bar{f}$ is in the canonical form of $\mathcal{A}_3$, where $T = \left[\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right] \in \mathbf{O}_2(\mathbb{C})$. To see this,

$$\begin{aligned}
\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}^{\otimes n} \left(\begin{bmatrix} 1 \\ \alpha^3 \end{bmatrix}^{\otimes n} + i^r \begin{bmatrix} 1 \\ -\alpha^3 \end{bmatrix}^{\otimes n} \right) &= \begin{bmatrix} -\alpha^3 \\ 1 \end{bmatrix}^{\otimes n} + i^r \begin{bmatrix} \alpha^3 \\ 1 \end{bmatrix}^{\otimes n} \\
&= \left(-\alpha^3\right)^n \left(\begin{bmatrix} 1 \\ -\frac{1}{\alpha^3} \end{bmatrix}^{\otimes n} + (-1)^n i^r \begin{bmatrix} 1 \\ \frac{1}{\alpha^3} \end{bmatrix}^{\otimes n} \right) \\
&= \left(-\alpha^3\right)^n \left(\begin{bmatrix} 1 \\ \alpha \end{bmatrix}^{\otimes n} + i^{2n+r} \begin{bmatrix} 1 \\ -\alpha \end{bmatrix}^{\otimes n} \right).
\end{aligned}$$

Let $\bar{f}' = T^{\otimes n}\bar{f}$, where $T \in \mathbf{O}_2(\mathbb{C})$ is given in each case. Then $\bar{f}'$ is f after an orthogonal transformation TH^{-1}. As shown above, $\bar{f}'$ is in the canonical form of $\mathcal{A}_1$ or $\mathcal{A}_3$, or is in the equivalent form of $\mathcal{A}_2$ by Lemma 5.8. Hence $f \in \mathcal{A}_1 \cup \mathcal{A}_2 \cup \mathcal{A}_3$.

Conversely, if there exists a matrix $H \in \mathbf{O}_2(\mathbb{C})$ such that $H^{\otimes n} f$ is in one of the canonical forms of $\mathcal{A}_1$, $\mathcal{A}_2$, or $\mathcal{A}_3$, then one can directly check that f is $\mathcal{A}$-transformable by Definition 1.3. In fact, transformations we applied above are all invertible. □

5.2 Characterization of $\mathcal{P}$-transformable Signatures

Turn our attention to $\mathcal{P}$-transformable signatures next. Define the stabilizer group of $\mathcal{P}$:

$$\operatorname{Stab}(\mathcal{P}) := \{T \in \mathbf{GL}_2(\mathbb{C}) \mid T\mathcal{P} \subseteq \mathcal{P}\}.$$

Technically this set is the left stabilizer group of $\mathcal{P}$. However, it is easy to see the left and right stabilizers coincide. Furthermore, $\operatorname{Stab}(\mathcal{P})$ is generated by nonzero scalar multiples of matrices of the form $\left[\begin{smallmatrix} 1 & 0 \\ 0 & \nu \end{smallmatrix}\right]$ for any nonzero $\nu \in \mathbb{C}$ and $X = \left[\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right]$. We start with an analogue of Lemma 5.5.

Lemma 5.11. *Let $\mathcal{F}$ be a set of signatures. Then $\mathcal{F}$ is $\mathcal{P}$-transformable if and only if $\mathcal{F} \subseteq \left[\begin{smallmatrix} 1 & 1 \\ i & -i \end{smallmatrix}\right] \mathcal{P}$ or there exists an $H \in \mathbf{O}_2(\mathbb{C})$ such that $\mathcal{F} \subseteq H\mathcal{P}$.*

Proof. Sufficiency is easily verified by checking that $=_2$ is transformed into $\mathcal{P}$ in both cases. In particular, H leaves $=_2$ unchanged.

If $\mathcal{F}$ is $\mathcal{P}$-transformable, then by definition, there exists a matrix T such that $(=_2)T^{\otimes 2} \in \mathcal{P}$ and $T^{-1}\mathcal{F} \subseteq \mathcal{P}$. The non-degenerate binary signatures in $\mathcal{P}$ are either $[0, 1, 0]$ or of the form $[1, 0, \nu]$, up to a scalar. However, notice that $[1, 0, 1] = [1, 0, \nu] \left[\begin{smallmatrix} 1 & 0 \\ 0 & \nu^{-\frac{1}{2}} \end{smallmatrix}\right]^{\otimes 2}$ and $\left[\begin{smallmatrix} 1 & 0 \\ 0 & \nu^{-\frac{1}{2}} \end{smallmatrix}\right] \in \operatorname{Stab}(\mathcal{P})$. Thus, we only need to consider $[1, 0, 1]$ and $[0, 1, 0]$. Now we apply Proposition 5.1.

1. If $(=_2)T^{\otimes 2} = [1, 0, 1]$, then by case 1 of Proposition 5.1, we have $T \in \mathbf{O}_2(\mathbb{C})$. Therefore $\mathcal{F} \subseteq H\mathcal{P}$ where $H = T \in \mathbf{O}(\mathbb{C})$.

2. If $(=_2)T^{\otimes 2} = [0, 1, 0]$, then by case 3 of Proposition 5.1, there exists an $H \in \mathbf{O}_2(\mathbb{C})$ such that $T = \frac{1}{\sqrt{2}} H \left[\begin{smallmatrix} 1 & 1 \\ i & -i \end{smallmatrix}\right]$. Therefore $\mathcal{F} \subseteq H \left[\begin{smallmatrix} 1 & 1 \\ i & -i \end{smallmatrix}\right] \mathcal{P}$ where $H \in \mathbf{O}_2(\mathbb{C})$. Moreover, it is easy to verify that $H \left[\begin{smallmatrix} 1 & 1 \\ i & -i \end{smallmatrix}\right] = \left[\begin{smallmatrix} 1 & 1 \\ i & -i \end{smallmatrix}\right] D$ or $\left[\begin{smallmatrix} 1 & 1 \\ i & -i \end{smallmatrix}\right] D \left[\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right]$ where D is non-singular and diagonal. Hence $D \in \operatorname{Stab}(\mathcal{P})$ or $D \left[\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right] \in \operatorname{Stab}(\mathcal{P})$, and in either case, $\mathcal{F} \subseteq \left[\begin{smallmatrix} 1 & 1 \\ i & -i \end{smallmatrix}\right] \mathcal{P}$. □

We also define $\mathcal{P}_1$ and $\mathcal{P}_2$ similar to $\mathcal{A}_i$ for $i = 1, 2, 3$.

Definition 5.12. *A symmetric signature* f *of arity* n *is in* $\mathcal{P}_1$ *if there exists* $H \in \mathbf{O}_2(\mathbb{C})$ *and a nonzero constant* $c \in \mathbb{C}$ *such that* $f = cH^{\otimes n} \left(\left[\begin{smallmatrix} 1 \\ 1 \end{smallmatrix}\right]^{\otimes n} + \beta \left[\begin{smallmatrix} 1 \\ -1 \end{smallmatrix}\right]^{\otimes n} \right)$*, where* $\beta \neq 0$*.*

When such an H exists, we say that $f \in \mathcal{P}_1$ with transformation H. If $f \in \mathcal{P}_1$ with I_2, then we say f is in the canonical form of $\mathcal{P}_1$. If f is in the canonical form of $\mathcal{P}_1$, then by Lemma 5.2, for any $0 \leqslant k \leqslant n-2$, we have $f_{k+2} = f_k$. Since f is non-degenerate, $f_1 \neq \pm f_0$ is implied.

It is easy to check that $\mathcal{A}_1 \subset \mathcal{P}_1$. The corresponding definition for $\mathcal{P}_2$ coincides with Definition 5.7 for $\mathcal{A}_2$. In other words, we define $\mathcal{P}_2 = \mathcal{A}_2$.

Now we characterize the $\mathcal{P}$-transformable signatures as we did for the $\mathcal{A}$-transformable signatures in Lemma 5.10.

Lemma 5.13. *Let* f *be a non-degenerate symmetric signature. Then* f *is* $\mathcal{P}$*-transformable if and only if* $f \in \mathcal{P}_1 \cup \mathcal{P}_2$*.*

Proof. Assume that f is $\mathcal{P}$-transformable of arity n. By applying Lemma 5.11 to $\{f\}$, there exists an $H \in \mathbf{O}_2(\mathbb{C})$ such that $f \in H\mathcal{P}$ or $f \in H \left[\begin{smallmatrix} 1 & 1 \\ i & -i \end{smallmatrix}\right] \mathcal{P}$. This is equivalent to $(H^{-1})^{\otimes n} f \in \mathcal{P}$ or $(H^{-1})^{\otimes n} f \in \left[\begin{smallmatrix} 1 & 1 \\ i & -i \end{smallmatrix}\right] \mathcal{P}$. Let $\overline{f} = (H^{-1})^{\otimes n} f$. It is sufficient to show that $\overline{f} \in \mathcal{P}_1$ or $\mathcal{P}_2$.

The symmetric signatures in $\mathcal{P}$ take the form $[0,1,0]$, or $[a,0,\ldots,0,b] = a[1,0]^{\otimes n} + b[0,1]^{\otimes n}$, where $ab \neq 0$ since f is non-degenerate. Now we consider all possible cases.

1. If $\bar{f} = [0,1,0]$, then $\bar{f} = \frac{1}{2i}\left(\left[\begin{smallmatrix}1\\i\end{smallmatrix}\right]^{\otimes 2} - \left[\begin{smallmatrix}1\\-i\end{smallmatrix}\right]^{\otimes 2}\right)$, which is the equivalent form of $\mathcal{P}_2 = \mathcal{A}_2$ given by Lemma 5.8.

2. If $\bar{f} = a\left[\begin{smallmatrix}1\\0\end{smallmatrix}\right]^{\otimes n} + b\left[\begin{smallmatrix}0\\1\end{smallmatrix}\right]^{\otimes n}$, then a further transformation by $\frac{1}{\sqrt{2}}\left[\begin{smallmatrix}1 & 1\\1 & -1\end{smallmatrix}\right] \in \mathbf{O}_2(\mathbb{C})$ puts $\bar{f}$ into the canonical form of $\mathcal{P}_1$.

3. If $\bar{f} = \left[\begin{smallmatrix}1 & 1\\i & -i\end{smallmatrix}\right]^{\otimes 2}[0,1,0]^{\mathsf{T}} = 2[1,0,1] = \left[\begin{smallmatrix}1\\1\end{smallmatrix}\right]^{\otimes 2} + \left[\begin{smallmatrix}1\\-1\end{smallmatrix}\right]^{\otimes 2}$, then $\bar{f}$ is already in the canonical form of $\mathcal{P}_1$.

4. If $\bar{f} = \left[\begin{smallmatrix}1 & 1\\i & -i\end{smallmatrix}\right]^{\otimes n}\left(a\left[\begin{smallmatrix}1\\0\end{smallmatrix}\right]^{\otimes n} + b\left[\begin{smallmatrix}0\\1\end{smallmatrix}\right]^{\otimes n}\right)$, then $\bar{f}$ is already of the equivalent form of $\mathcal{P}_2 = \mathcal{A}_2$ given by Lemma 5.8.

Conversely, if there exists a matrix $H \in \mathbf{O}_2(\mathbb{C})$ such that $H^{\otimes n} f$ is in one of the canonical forms of $\mathcal{P}_1$ or $\mathcal{P}_2$, then one can directly check that f is $\mathcal{P}$-transformable by Definition 1.3. In fact, transformations that we applied above are all invertible. □

Combining Lemma 5.10 and Lemma 5.13, we have a necessary and sufficient condition for a single non-degenerate signature to be $\mathcal{A}$- or $\mathcal{P}$-transformable.

Corollary 5.14. *Let f be a non-degenerate signature. Then f is $\mathcal{A}$- or $\mathcal{P}$-transformable if and only if $f \in \mathcal{P}_1 \cup \mathcal{P}_2 \cup \mathcal{A}_3$.*

Notice that our definitions of $\mathcal{P}_1$, $\mathcal{P}_2$, and $\mathcal{A}_3$ each involve an orthogonal transformation. For any single signature $f \in \mathcal{P}_1 \cup \mathcal{P}_2 \cup \mathcal{A}_3$, $\operatorname{Holant}(f)$ is tractable. However, this does not imply that $\operatorname{Holant}(\mathcal{P}_1)$, $\operatorname{Holant}(\mathcal{P}_2)$, or $\operatorname{Holant}(\mathcal{A}_3)$ is tractable. In fact, $\operatorname{Holant}(\mathcal{P}_2)$ is tractable while $\operatorname{Holant}(\mathcal{P}_1)$ and $\operatorname{Holant}(\mathcal{A}_3)$ are #**P**-hard.

5.3 Characterization of $\mathcal{M}$-transformable Signatures

Next come $\mathcal{M}$-transformable signatures. Define the stabilizer group of $\mathcal{M}$:

$$\operatorname{Stab}(\mathcal{M}) := \{T \in \mathbf{GL}_2(\mathbb{C}) \mid T\mathcal{M} \subseteq \mathcal{M}\}.$$

Technically this set is the left stabilizer group of $\mathcal{M}$. However, it is easy to see the left and right stabilizers coincide. Moreover, $\operatorname{Stab}(\mathcal{M})$ is generated by nonzero scalar multiples of matrices of the form $\left[\begin{smallmatrix}1 & 0\\ 0 & \nu\end{smallmatrix}\right]$ for any nonzero $\nu \in \mathbb{C}$ and $X = \left[\begin{smallmatrix}0 & 1\\ 1 & 0\end{smallmatrix}\right]$. In other words, $\operatorname{Stab}(\mathcal{M}) = \operatorname{Stab}(\mathcal{P})$.

Again, we have an analogue of Lemma 5.5. Notice that binary symmetric signatures in $\mathcal{M}$ are exactly the same as those in $\mathcal{P}$. The statement of Lemma 5.11 also holds for $\mathcal{M}$.

Lemma 5.15. *Let $\mathcal{F}$ be a set of signatures. Then $\mathcal{F}$ is $\mathcal{M}$-transformable if and only if $\mathcal{F} \subseteq \left[\begin{smallmatrix}1 & 1\\ i & -i\end{smallmatrix}\right] \mathcal{M}$ or there exists an $H \in \mathbf{O}_2(\mathbb{C})$ such that $\mathcal{F} \subseteq H\mathcal{M}$.*

We use four sets to characterize the $\mathcal{M}$-transformable signatures. The function $\operatorname{Sym}_n^t(-;-)$ is defined in Definition 1.12.

Definition 5.16. *A symmetric signature f of arity n is in $\mathcal{M}_i$ for $k = 1, 2, 3, 4$, if there exist an $H \in \mathbf{O}_2(\mathbb{C})$ and nonzero constants $c, \gamma \in \mathbb{C}$ such that f has the form (k) as follows:*

$(1):\ cH^{\otimes n}\left(\left[\begin{smallmatrix}1\\ 1\end{smallmatrix}\right]^{\otimes n} \pm i^n \left[\begin{smallmatrix}1\\ -1\end{smallmatrix}\right]^{\otimes n}\right)$;

$(2):\ cH^{\otimes n}\left(\left[\begin{smallmatrix}1\\ \gamma\end{smallmatrix}\right]^{\otimes n} \pm \left[\begin{smallmatrix}1\\ -\gamma\end{smallmatrix}\right]^{\otimes n}\right)$ *for some* $\gamma \neq 0$;

$(3):\ cH^{\otimes n} \operatorname{Sym}_n^{n-1}\left(\left[\begin{smallmatrix}1\\ 0\end{smallmatrix}\right]; \left[\begin{smallmatrix}0\\ 1\end{smallmatrix}\right]\right)$;

$(4):\ cH^{\otimes n} \operatorname{Sym}_n^{n-1}\left(\left[\begin{smallmatrix}1\\ i\end{smallmatrix}\right]; \left[\begin{smallmatrix}1\\ -i\end{smallmatrix}\right]\right)$.

For $k \in \{1, 2, 3, 4\}$, when such an H exists, we say that $f \in \mathcal{M}_k$ with transformation H. If $f \in \mathcal{M}_k$ with I_2, then we say f is in the canonical form of $\mathcal{M}_k$.

Recall that we have defined $\mathcal{M}_4^{\pm}$ in Section 3.7, right before Lemma 3.43. The definition there is consistent with Definition 5.16 as $\textsc{ExactOne}_n = \operatorname{Sym}_n^{n-1}\left(\left[\begin{smallmatrix}1\\ 0\end{smallmatrix}\right]; \left[\begin{smallmatrix}0\\ 1\end{smallmatrix}\right]\right)$.

Notice that $\left\{\left[\begin{smallmatrix}1\\ i\end{smallmatrix}\right], \left[\begin{smallmatrix}1\\ -i\end{smallmatrix}\right]\right\}$ is set-wise invariant under any transformation in $\mathbf{O}_2(\mathbb{C})$ up to nonzero constants. Using this fact, the following lemma gives a characterization of $\mathcal{M}_4$. It says that any signature in $\mathcal{M}_4$ is essentially in its canonical form.

Lemma 5.17. *Let f be a symmetric signature of arity n. Then $f \in \mathcal{M}_4$ if and only if $f = c\operatorname{Sym}_n^{n-1}\left(\left[\begin{smallmatrix}1\\ i\end{smallmatrix}\right]; \left[\begin{smallmatrix}1\\ -i\end{smallmatrix}\right]\right)$ or $f = c\operatorname{Sym}_n^{n-1}\left(\left[\begin{smallmatrix}1\\ -i\end{smallmatrix}\right]; \left[\begin{smallmatrix}1\\ i\end{smallmatrix}\right]\right)$ for some nonzero constant $c \in \mathbb{C}$.*

Proof. Suppose $f \in \mathcal{M}_4$ with the transformation H, that is, $f = cH^{\otimes n} \operatorname{Sym}_n^{n-1}\left(\left[\begin{smallmatrix}1\\ i\end{smallmatrix}\right]; \left[\begin{smallmatrix}1\\ -i\end{smallmatrix}\right]\right)$. If $H \in \mathbf{SO}_2(\mathbb{C})$, then $H = \left[\begin{smallmatrix}a & b\\ -b & a\end{smallmatrix}\right]$ for some $a, b \in \mathbb{C}$ such that $a^2 + b^2 = 1$. In particular, $a \pm bi \neq 0$.

Since $H\left[\begin{smallmatrix}1\\i\end{smallmatrix}\right] = (a+bi)\left[\begin{smallmatrix}1\\i\end{smallmatrix}\right]$ and $H\left[\begin{smallmatrix}1\\-i\end{smallmatrix}\right] = (a-bi)\left[\begin{smallmatrix}1\\-i\end{smallmatrix}\right]$, it follows that $f = c(a+bi)^{n-1}(a-bi)\operatorname{Sym}_n^{n-1}\left(\left[\begin{smallmatrix}1\\i\end{smallmatrix}\right];\left[\begin{smallmatrix}1\\-i\end{smallmatrix}\right]\right)$.

Otherwise, $H \in \mathbf{O}_2(\mathbb{C}) - \mathbf{SO}_2(\mathbb{C})$, so $H = \left[\begin{smallmatrix}a & b\\ b & -a\end{smallmatrix}\right]$ for some $a, b \in \mathbb{C}$ such that $a^2 + b^2 = 1$. Then $f = c(a+bi)(a-bi)^{n-1}\operatorname{Sym}_n^{n-1}\left(\left[\begin{smallmatrix}1\\-i\end{smallmatrix}\right];\left[\begin{smallmatrix}1\\i\end{smallmatrix}\right]\right)$.

To show the other direction, suppose $f = c\operatorname{Sym}_n^{n-1}\left(\left[\begin{smallmatrix}1\\i\end{smallmatrix}\right];\left[\begin{smallmatrix}1\\-i\end{smallmatrix}\right]\right)$ or $f = c\operatorname{Sym}_n^{n-1}\left(\left[\begin{smallmatrix}1\\-i\end{smallmatrix}\right];\left[\begin{smallmatrix}1\\i\end{smallmatrix}\right]\right)$. The first case is already in the canonical form of $\mathcal{M}_4$. In the second case, we pick $H = \left[\begin{smallmatrix}1 & 0\\ 0 & -1\end{smallmatrix}\right] \in \mathbf{O}_2(\mathbb{C})$. Then $H^{\otimes n} f$ is in the canonical form of $\mathcal{M}_4$. □

Next we show that $\mathcal{M}_k$ for $k = 1, 2, 3, 4$ captures all $\mathcal{M}$-transformable signatures.

Lemma 5.18. *Let* f *be a non-degenerate symmetric signature. Then* f *is* $\mathcal{M}$*-transformable if and only if* $f \in \mathcal{M}_1 \cup \mathcal{M}_2 \cup \mathcal{M}_3 \cup \mathcal{M}_4$.

Proof. Assume that f is $\mathcal{M}$-transformable of arity n. By applying Lemma 5.15 to $\{f\}$, we have $f \in \left[\begin{smallmatrix}1 & 1\\ i & -i\end{smallmatrix}\right]\mathcal{M}$ or there exists an $H \in \mathbf{O}_2(\mathbb{C})$ such that $f \in H\mathcal{M}$. Proposition 1.13 lists all symmetric signatures in $\mathcal{M}$. Since we are only interested in non-degenerate signatures, we only consider non-zero a, b, and λ. We divide into two cases.

1. Suppose $f \in \left[\begin{smallmatrix}1 & 1\\ i & -i\end{smallmatrix}\right]\mathcal{M}$.

   - Further suppose $f = \left[\begin{smallmatrix}1 & 1\\ i & -i\end{smallmatrix}\right]^{\otimes n}\left(\left[\begin{smallmatrix}a\\b\end{smallmatrix}\right]^{\otimes n} \pm \left[\begin{smallmatrix}a\\-b\end{smallmatrix}\right]^{\otimes n}\right)$ for some nonzero $a, b \in \mathbb{C}$. Let $T = \frac{1-i}{2}\left[\begin{smallmatrix}u & v\\ v & -u\end{smallmatrix}\right]$, where $u = a + bi$ and $v = i(a - bi)$. Then $f = T^{\otimes n}\left(\left[\begin{smallmatrix}1\\1\end{smallmatrix}\right]^{\otimes n} \pm i^n\left[\begin{smallmatrix}1\\-1\end{smallmatrix}\right]^{\otimes n}\right)$. Since $T \in \mathbf{O}_2(\mathbb{C})$ up to a nonzero factor of $\sqrt{2ab}$, we have $f \in \mathcal{M}_1$.
   - Further suppose $f = \lambda\left[\begin{smallmatrix}1 & 1\\ i & -i\end{smallmatrix}\right]^{\otimes n}\operatorname{Sym}_n^{n-1}\left(\left[\begin{smallmatrix}1\\0\end{smallmatrix}\right];\left[\begin{smallmatrix}0\\1\end{smallmatrix}\right]\right)$ for some nonzero $\lambda \in \mathbb{C}$. Then we have $f = \lambda\operatorname{Sym}_n^{n-1}\left(\left[\begin{smallmatrix}1\\i\end{smallmatrix}\right];\left[\begin{smallmatrix}1\\-i\end{smallmatrix}\right]\right)$, so $f \in \mathcal{M}_4$.
   - Further suppose $f = \lambda\left[\begin{smallmatrix}1 & 1\\ i & -i\end{smallmatrix}\right]^{\otimes n}\operatorname{Sym}_n^{n-1}\left(\left[\begin{smallmatrix}0\\1\end{smallmatrix}\right];\left[\begin{smallmatrix}1\\0\end{smallmatrix}\right]\right)$ for some nonzero $\lambda \in \mathbb{C}$. Then we have $f = \lambda\operatorname{Sym}_n^{n-1}\left(\left[\begin{smallmatrix}1\\-i\end{smallmatrix}\right];\left[\begin{smallmatrix}1\\i\end{smallmatrix}\right]\right)$, so $f \in \mathcal{M}_4$ by Lemma 5.17.

2. Suppose $f \in H\mathcal{M}$.

   - Further suppose $f = H^{\otimes n}\left(\left[\begin{smallmatrix}a\\b\end{smallmatrix}\right]^{\otimes n} \pm \left[\begin{smallmatrix}a\\-b\end{smallmatrix}\right]^{\otimes n}\right)$ for some nonzero $a, b \in \mathbb{C}$. Then we have $f = a^n H^{\otimes n}\left(\left[\begin{smallmatrix}1\\\gamma\end{smallmatrix}\right]^{\otimes n} \pm \left[\begin{smallmatrix}1\\-\gamma\end{smallmatrix}\right]^{\otimes n}\right)$, where $\gamma = \frac{b}{a}$, so $f \in \mathcal{M}_2$.
   - Further suppose $f = \lambda H^{\otimes n}\operatorname{Sym}_n^{n-1}\left(\left[\begin{smallmatrix}1\\0\end{smallmatrix}\right];\left[\begin{smallmatrix}0\\1\end{smallmatrix}\right]\right)$ for some nonzero $\lambda \in \mathbb{C}$. Then $f \in \mathcal{M}_3$.

- Further suppose $f = \lambda H^{\otimes n} \operatorname{Sym}_n^{n-1}(\left[\begin{smallmatrix}0\\1\end{smallmatrix}\right]; \left[\begin{smallmatrix}1\\0\end{smallmatrix}\right])$ for some nonzero $\lambda \in \mathbb{C}$. Let $H' = H \left[\begin{smallmatrix}0&1\\1&0\end{smallmatrix}\right] \in \mathbf{O}_2(\mathbb{C})$. Then we have $f = \lambda H'^{\otimes n} \operatorname{Sym}_n^{n-1}(\left[\begin{smallmatrix}1\\0\end{smallmatrix}\right]; \left[\begin{smallmatrix}0\\1\end{smallmatrix}\right])$, so $f \in \mathcal{M}_3$.

Conversely, if there exists a matrix $H \in \mathbf{O}_2(\mathbb{C})$ such that $H^{\otimes n} f$ is in one of the canonical forms of $\mathcal{M}_1, \mathcal{M}_2, \mathcal{M}_3$, or $\mathcal{M}_4$, then one can directly check that f is $\mathcal{M}$-transformable by Definition 1.3. In fact, transformations that we applied above are all invertible. □

Let $g = [x, y, 0, \ldots, 0, z]$ with arity $n \geqslant 3$, where $xyz \neq 0$. As an example of the theory developed in this section, we show that the signature $Z^{\otimes n} g$ is not in any of the tractable sets. We will use the following lemma in future.

Lemma 5.19. *Let* $g = [x, y, 0, \ldots, 0, z]$ *with arity* $n \geqslant 3$ *and* $xyz \neq 0$*. Then the signature* $Z^{\otimes n} g$ *is neither* $\mathcal{A}$-*,* $\mathcal{P}$-*,* $\mathcal{M}$-*transformable, nor vanishing.*

Proof. By Lemma 3.14 and Theorem 3.12, $Z^{\otimes n} g$ is not vanishing. To show that $Z^{\otimes n} g$ is not $\mathcal{A}$-, $\mathcal{P}$-, $\mathcal{M}$-transformable, we only need to show that $Z^{\otimes n} g \notin \mathcal{P}_1 \cup \mathcal{M}_2 \cup \mathcal{A}_3 \cup \mathcal{M}_3 \cup \mathcal{M}_4$ by Lemma 5.10, 5.13 and 5.18, and the fact that $\mathcal{M}_1 \subset \mathcal{A}_1 \subset \mathcal{P}_1$ and $\mathcal{A}_2 = \mathcal{P}_2 \subset \mathcal{M}_2$. See Figure 5.1. Note that $\mathcal{M}_4 \subset \mathcal{V}$. Since $Z^{\otimes n} g$ is not vanishing, it is not in $\mathcal{M}_4$.

We first show that $Z^{\otimes n} g \notin \mathcal{P}_1 \cup \mathcal{M}_2 \cup \mathcal{A}_3$. We say a signature $f = [f_0, f_1, \ldots, f_n]$ satisfies a second order recurrence of type $\langle a, b, c \rangle$ if $a f_k - b f_{k+1} + c f_{k+2} = 0$ for $1 \leqslant k \leqslant n-2$, for some a, b and c not all zero. Suppose $Z^{\otimes n} g$ is a nonzero constant multiple of $Hf \in \mathcal{P}_1 \cup \mathcal{M}_2 \cup \mathcal{A}_3$ in the forms given in Definitions 5.6, 5.7, 5.9, 5.12, and 5.16, then f, and hence also $(Z^{-1})^{\otimes n} f$, satisfies a second order recurrence by Lemma 5.2. We have $H^{-1} Z = ZD$ or $ZD \left[\begin{smallmatrix}0&1\\1&0\end{smallmatrix}\right]$ for some non-singular diagonal D since $H \in \mathbf{O}_2(\mathbb{C})$. Thus $f = Z^{\otimes n} g'$ for some $g' = [x', y', 0, \ldots, 0, z']$ or $[x', 0, \ldots, 0, y', z']$, with $x'y'z' \neq 0$. We assume the former; the proof is similar for the latter.

However, for $n \geqslant 4$, g' does not satisfy any second order recurrence. For a contradiction suppose g' does. By $x'y'z' \neq 0$, $ay' - b0 + c0 = 0$ gives $a = 0$, $ax' - by' + c0 = 0$ gives $b = 0$, and $a0 - b0 + cz' = 0$ gives $c = 0$; but a, b, c cannot be all zero.

Next suppose $n = 3$, and we show that $g' = (Z^{-1})^{\otimes n} f$ is still impossible.

- For $\mathcal{P}_1$, $f = \left[\begin{smallmatrix}1\\1\end{smallmatrix}\right]^{\otimes 3} + \beta \left[\begin{smallmatrix}1\\-1\end{smallmatrix}\right]^{\otimes 3}$. It is easy to check that $(Z^{-1})^{\otimes n} f$ satisfies a second order recurrence with its two eigenvalues sum to zero. However $g' = [x', y', 0, z']$ has type $\langle y'z', x'z', -y'^2 \rangle$, the sum of its two eigenvalues is $-x'z'/y'^2 \neq 0$.

- For $\mathcal{M}_2$, $f = \left[\begin{smallmatrix}1\\\gamma\end{smallmatrix}\right]^{\otimes 3} \pm \left[\begin{smallmatrix}1\\-\gamma\end{smallmatrix}\right]^{\otimes 3}$. Since $Z^{-1}\left[\begin{smallmatrix}1 & 1\\\gamma & -\gamma\end{smallmatrix}\right]$ has the form $\left[\begin{smallmatrix}u & v\\v & u\end{smallmatrix}\right]$, $(Z^{-1})^{\otimes n}f = \left[\begin{smallmatrix}u\\v\end{smallmatrix}\right]^{\otimes 3} \pm \left[\begin{smallmatrix}v\\u\end{smallmatrix}\right]^{\otimes 3}$. Thus the weight 1 and weight 2 entries of $(Z^{-1})^{\otimes n}f$ are either equal or negative of each other. If $g' = (Z^{-1})^{\otimes n}f$ this would imply $y' = 0$, a contradiction.

- For $\mathcal{A}_3$, $f = \left[\begin{smallmatrix}1\\\alpha\end{smallmatrix}\right]^{\otimes n} + i^r\left[\begin{smallmatrix}1\\-\alpha\end{smallmatrix}\right]^{\otimes n}$. Since $Z^{-1}\left[\begin{smallmatrix}1 & 1\\\alpha & -\alpha\end{smallmatrix}\right] = \left[\begin{smallmatrix}u & v\\v & u\end{smallmatrix}\right]$, with $u = 1-\alpha i$ and $v = 1+\alpha i$, the weight 2 entry of $(Z^{-1})^{\otimes n}f$ is $uv^2 + i^r vu^2 = (uv)(v + i^r u)$. This is nonzero for all r. However the weight 2 entry of $g' = [x', y', 0, z']$ is 0.

It remains to show that $Z^{\otimes n}g \notin \mathcal{M}_3$. If $Z^{\otimes n}g \in \mathcal{M}_3$, then $Z^{\otimes n}g = cHf$ for some $H \in \mathbf{O}_2(\mathbb{C})$ and $f = \mathrm{Sym}_n^{n-1}\left(\left[\begin{smallmatrix}1\\0\end{smallmatrix}\right];\left[\begin{smallmatrix}0\\1\end{smallmatrix}\right]\right)$. Again $f = (cH)^{-1}Z^{\otimes n}g = Z^{\otimes n}g'$ for some g' having the same or its reversal form as g. Then $g' = (Z^{-1})^{\otimes n}f$ is the signature $[n, n-2, \ldots, -(n-2), -n]$. The weight 1 entry and weight $n-1$ entry have the same absolute value. By the form of g' this is a contradiction. □

Combine Lemma 5.19 with Theorem 1.14 and Theorem 3.39. For arity $n = 3$ or 4, Pl-Holant $(Z^{\otimes n}g)$ is #**P**-hard.

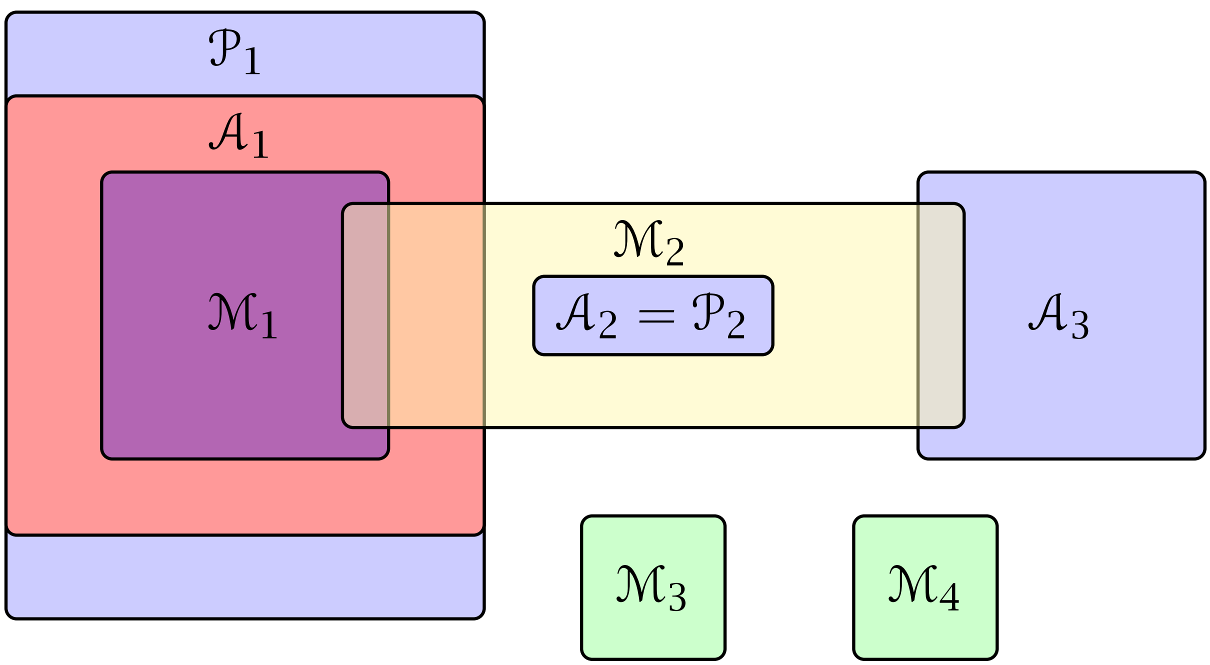


Figure 5.1: Relationships among $\mathcal{A}_1$, $\mathcal{A}_2$, $\mathcal{A}_3$, $\mathcal{P}_1$, $\mathcal{P}_2$, $\mathcal{M}_1$, $\mathcal{M}_2$, $\mathcal{M}_3$, and $\mathcal{M}_4$.

We note that $\mathcal{M}_1 \subset \mathcal{A}_1 \subset \mathcal{P}_1$ and $\mathcal{A}_2 = \mathcal{P}_2 \subset \mathcal{M}_2$. Also note that $\mathcal{P}_1 \cap \mathcal{M}_2 \subseteq \mathcal{A}_1$. See Figure 5.1 for a visual description of the relationships among sets. Combine Corollary 5.14 with Lemma

5.18.

Corollary 5.20. *Let* f *be a non-degenerate signature. Then* f *is* $\mathcal{A}$*-,* $\mathcal{P}$*-, or* $\mathcal{M}$*-transformable if and only if* $f \in \mathcal{P}_1 \cup \mathcal{M}_2 \cup \mathcal{A}_3 \cup \mathcal{M}_3 \cup \mathcal{M}_4$.

To finish this section, we show that signatures in $\mathcal{M}_3$ are not $\mathcal{A}$- or $\mathcal{P}$-transformable.

Lemma 5.21. *Let* $f \in \mathcal{M}_3$ *be a non-degenerate signature of arity* $n \geqslant 3$ *with* $H \in \mathbf{O}_2(\mathbb{C})$*. Then* f *is not* $\mathcal{A}$*- or* $\mathcal{P}$*-transformable. Moreover,* f *is* $\mathcal{M}$*-transformable with only* HD *or* $H\left[\begin{smallmatrix}0 & 1\\ 1 & 0\end{smallmatrix}\right]D$ *for some diagonal matrix* D.

Proof. Suppose $f = [f_0, f_1, \ldots, f_n]$. By Lemma 5.2, Definitions 5.6, 5.7, 5.9, and 5.12, and Corollary 5.14, if f is $\mathcal{A}$- or $\mathcal{P}$-transformable, then f has to satisfy a second order recurrence relation that $af_i + bf_{i+1} + cf_{i+2} = 0$, for $a, b, c \in \mathbb{C}$ such that not all a, b, c are 0 and $b^2 - 4ac \neq 0$. In other words, the second order recurrence relation has to have distinct eigenvalues. Moreover, this property is preserved by holographic transformations (cf. Lemma 6.2 in [CGW14]). However, f is in $\mathcal{M}_3$. Hence $f = H^{\otimes n}\textsc{ExactOne}_n$ for some $H \in \mathbf{O}_2(\mathbb{C})$ up to a nonzero factor. On the other hand, $\textsc{ExactOne}_n$ does not satisfy a second recurrence with distinct eigenvalues if $n \geqslant 3$, a contradiction.

Moreover, notice that the only signatures in $\mathcal{M}$ that do not satisfy such second order recurrence relations are $\textsc{ExactOne}_k$ and $\textsc{AllButOne}_k$ functions. If f is $\mathcal{M}$-transformable, then by Lemma 5.15, either $f = \left[\begin{smallmatrix}1 & 1\\ i & -i\end{smallmatrix}\right]^{\otimes n} g$ for some $g \in \mathcal{M}$, or there exists $T \in \mathbf{O}_2(\mathbb{C})$ such that $f = T^{\otimes n}g$ for some $g \in \mathcal{M}$. Hence $g = \textsc{ExactOne}_n$ or $\textsc{AllButOne}_n$. On the other hand $f = H^{\otimes n}\textsc{ExactOne}_n$ up to a nonzero factor. It is easy to verify that $f = \left[\begin{smallmatrix}1 & 1\\ i & -i\end{smallmatrix}\right]^{\otimes n} g$ is impossible. Therefore $(T^{-1}H)^{\otimes n}\textsc{ExactOne}_n = \textsc{ExactOne}_n$ or $\textsc{AllButOne}_n$ up to a nonzero factor.

Let $J = T^{-1}H = \left[\begin{smallmatrix}x & y\\ z & w\end{smallmatrix}\right]$ and let $h = J^{\otimes n}\textsc{ExactOne}_n$. As $\textsc{ExactOne}_n = \operatorname{Sym}_n^{n-1}(\left[\begin{smallmatrix}1\\ 0\end{smallmatrix}\right]; \left[\begin{smallmatrix}0\\ 1\end{smallmatrix}\right])$, $h = (\left[\begin{smallmatrix}x & y\\ z & w\end{smallmatrix}\right])^{\otimes n}\textsc{ExactOne}_n = \operatorname{Sym}_n^{n-1}(\left[\begin{smallmatrix}x\\ z\end{smallmatrix}\right]; \left[\begin{smallmatrix}y\\ w\end{smallmatrix}\right])$. The first and last entries of h are $x^{n-1}y$ and $z^{n-1}w$. As $h = \textsc{ExactOne}_n$ or $\textsc{AllButOne}_n$, we have that $x^{n-1}y = z^{n-1}w = 0$. It is easy to see that x and z, or y and w cannot be both 0. Then $x = w = 0$ or $y = z = 0$. This implies that $J = D$ or $J = D\left[\begin{smallmatrix}0 & 1\\ 1 & 0\end{smallmatrix}\right]$ for some diagonal matrix D. Thus $T = HJ^{-1} = HD^{-1}$ or $H\left[\begin{smallmatrix}0 & 1\\ 1 & 0\end{smallmatrix}\right]D^{-1}$. □

5.4 Related Hardness Results

Now we give dichotomy theorems when signatures from $\mathcal{P}_1, \mathcal{M}_2 \setminus \mathcal{P}_2, \mathcal{A}_3$, or $\mathcal{M}_3$ appear. A major tool we use is the dichotomy for Pl-#CSP^2, shown in [CFGW15]. Here Pl-#CSP^2 denotes planar #CSP problems where every variable appears a multiple of 2 times. The proof of Theorem 5.22 depends on Theorem 4.1, and is a complicated case analysis, which we will omit.

Theorem 5.22. *Let $\mathcal{F}$ be a set of symmetric signatures. Then* Pl-#CSP$^2(\mathcal{F})$ *is #**P**-hard unless $\mathcal{F}$ satisfies one of the following conditions:*

1. *there exists $T \in \mathcal{T}_8$ such that $\mathcal{F} \subseteq T\mathcal{A}$;*

2. *$\mathcal{F} \subseteq \mathcal{P}$;*

3. *there exists $T \in \mathcal{T}_4$ such that $\mathcal{F} \subseteq T\overline{\mathcal{M}}$.*

In each exceptional case, Pl-#CSP$^2(\mathcal{F})$ *is computable in polynomial time.*

We begin with $\mathcal{P}_1$. Recall that H_2 is the 2-by-2 matrix $\frac{1}{\sqrt{2}}\left[\begin{smallmatrix}1 & 1\\ 1 & -1\end{smallmatrix}\right]$. We do some preparation first.

Lemma 5.23. *Let $a, b \in \mathbb{C}$. If $ab \neq 0$, then for any set $\mathcal{F}$ of complex-weighted signatures containing $[a, 0, \ldots, 0, b]$ of arity $r \geqslant 3$,*

$$\text{Pl-Holant}(\mathcal{F} \cup \{=_4\}) \leqslant_T \text{Pl-Holant}(\mathcal{F}).$$

Proof. Since $a \neq 0$, we can normalize the first entry to get $[1, 0, \ldots, 0, x]$, where $x \neq 0$. First, we show how to obtain an arity 4 generalized equality signature. If $r = 3$, then we connect two copies together by a single edge to get an arity 4 signature. For larger arities, we form self-loops until realizing a signature of arity 3 or 4. By this process, we have a signature $g = [1, 0, 0, 0, y]$, where $y \neq 0$. If y is a pth root of unity, then we can directly realize $=_4$ by connecting p copies of g together, two edges at a time as in Figure 3.7. Otherwise, y is not a root of unity and we can interpolate $=_4$ as follows.

Consider an instance Ω of Pl-Holant$(\mathcal{F} \cup \{=_4\})$. Suppose that $=_4$ appears n times in Ω. We stratify the assignments in Ω based on the assignments to $=_4$. We only need to consider assignments that give all $=_4$'s Hamming weights 0 and 4 since inputs of other Hamming weights contributes 0. If there are i many $=_4$'s having Hamming weight 0, then the rest $n-i$ many have Hamming weight 4. Let c_i denote the summation of the product of evaluations of signatures other than $=_4$ in Ω over assignments which give i many $=_4$'s Hamming weight 0. We can rewrite the Holant on Ω as $\text{Holant}_\Omega = \sum_{i=0}^{n} c_i$.

We construct from Ω a sequence of instances Ω_s of Pl-Holant($\mathcal{F}$) indexed by $s \geqslant 1$. We obtain Ω_s from Ω by replacing each occurrence of $=_4$ with a gadget g_s created from s copies of $[1, 0, 0, 0, y]$, connecting two edges together at a time as in Figure 3.7. The Holant on Ω_s is

$$\text{Holant}_{\Omega_s} = \sum_{i=0}^{n} (y^s)^i c_i.$$

For $s \geqslant 1$, this gives a coefficient matrix that is Vandermonde. Since y is neither 0 nor a root of unity, y^s is distinct for each s. Therefore, the Vandermonde system has full rank. We can solve for the unknowns c_i and obtain the value of Holant_Ω. □

Lemma 5.24. *Let* $f \in \mathcal{P}_1$ *be a non-degenerate signature of arity* $n \geqslant 3$ *with an orthogonal transformation* H *and* $\mathcal{F}$ *be a set of signatures containing* f. *Then* $\text{Pl-\#CSP}^2(H_2H^{-1}\mathcal{F}) \leqslant_T \text{Pl-Holant}(\mathcal{F})$.

Proof. By Definition 5.12, disregarding a nonzero constant factor, f has the following form:

$$f = H^{\otimes n}\left(\begin{bmatrix}1\\1\end{bmatrix}^{\otimes n} + \beta\begin{bmatrix}1\\-1\end{bmatrix}^{\otimes n}\right),$$

where $\beta \neq 0$. We do an orthogonal transformation by H_2H^{-1}. By Theorem 1.2,

$$\text{Pl-Holant}(\mathcal{F}) \equiv \text{Pl-Holant}(H_2H^{-1}\mathcal{F}).$$

Note that $\bar{f} = \left(H_2H^{-1}\right)^{\otimes n} f \in H_2H^{-1}\mathcal{F}$, and $\bar{f} = 2^{n/2}\left(\left[\begin{smallmatrix}1\\0\end{smallmatrix}\right]^{\otimes n} + \left[\begin{smallmatrix}0\\1\end{smallmatrix}\right]^{\otimes n}\right)$. By Lemma 5.23, we can obtain $=_4$, the arity 4 equality signature. With this signature, we can realize any equality signature of even arity. Thus, $\text{Pl-\#CSP}^2(H_2H^{-1}\mathcal{F}) \leqslant_T \text{Pl-Holant}(\mathcal{F})$. □

Combined with Theorem 5.22, Lemma 5.24 implies a dichotomy when $\mathcal{P}_1$ signatures appear.

Corollary 5.25. *Let $\mathcal{F}$ be a set of signatures. Suppose there exists $f \in \mathcal{F}$ which is a non-degenerate signature of arity $n \geqslant 3$ in $\mathscr{P}_1$. Then* Pl-Holant$(\mathcal{F})$ *is #**P**-hard unless $\mathcal{F}$ is $\mathscr{A}$-, $\mathscr{P}$-, or $\mathscr{M}$-transformable, in which case* Pl-Holant$(\mathcal{F})$ *is tractable.*

Proof. Assume that $f \in \mathscr{P}_1$ with $H \in \mathbf{O}_2(\mathbb{C})$. Let $H' = HH_2^{-1} \in \mathbf{O}_2(\mathbb{C})$. By Lemma 5.24 and Theorem 5.22, Pl-Holant$(\mathcal{F})$ is #**P**-hard unless (1) $\mathcal{F} \subseteq H'\mathscr{P}$, or (2) $\mathcal{F} \subseteq H'T\mathscr{A}$, or (3) $\mathcal{F} \subseteq H'T' \left[\begin{smallmatrix} 1 & 1 \\ 1 & -1 \end{smallmatrix}\right] \mathscr{M}$, where $T \in \mathcal{T}_8$ and $T' \in \mathcal{T}_4$. In case (1), $\mathcal{F}$ is $\mathscr{P}$-transformable since $(=_2)H'^{\otimes 2} = (=_2) \in \mathscr{P}$. In case (2), $\mathcal{F}$ is $\mathscr{A}$-transformable since $(=_2)(H'T)^{\otimes 2} = (=_2)T^{\otimes 2} \in \mathscr{A}$. In case (3), $\mathcal{F}$ is $\mathscr{M}$-transformable. If $T' = \left[\begin{smallmatrix} 1 & 0 \\ 0 & \pm 1 \end{smallmatrix}\right]$, then $T' \in \mathbf{O}_2(\mathbb{C})$. So $(=_2)(H'T' \left[\begin{smallmatrix} 1 & 1 \\ 1 & -1 \end{smallmatrix}\right])^{\otimes 2} = (=_2) \in \mathscr{M}$. If $T' = \left[\begin{smallmatrix} 1 & 0 \\ 0 & \pm i \end{smallmatrix}\right]$, then $T' \left[\begin{smallmatrix} 1 & 1 \\ 1 & -1 \end{smallmatrix}\right] = \left[\begin{smallmatrix} 1 & 1 \\ i & -i \end{smallmatrix}\right]$, and $(=_2)(H'T' \left[\begin{smallmatrix} 1 & 1 \\ 1 & -1 \end{smallmatrix}\right])^{\otimes 2} = 2[0, 1, 0] \in \mathscr{M}$. □

In particular, we have the following corollary. It will be useful in Section 5.6 to prove the single signature dichotomy for Pl-Holant.

Corollary 5.26. *Suppose f is a non-degenerate signature of arity $n \geqslant 5$. Let f' be f with a self loop. If $f' \in \mathscr{P}_1$ is non-degenerate, then* Pl-Holant(f) *is #**P**-hard unless f is $\mathscr{A}$-, $\mathscr{P}$-, or $\mathscr{M}$-transformable, in which case* Pl-Holant(f) *is tractable.*

We have similar results for $\mathscr{A}_3$.

Lemma 5.27. *Let $f \in \mathscr{A}_3$ be a non-degenerate signature of arity $n \geqslant 3$ with an orthogonal transformation H and $\mathcal{F}$ be a set of signatures containing f. Let $\alpha = e^{\pi i/4}$ and Y be the 2-by-2 matrix $\left[\begin{smallmatrix} \alpha & 1 \\ -\alpha & 1 \end{smallmatrix}\right]$. Then* Pl-#CSP$^2(YH^{-1}\mathcal{F} \cup \{[1, -i, 1]\}) \leqslant_T$ Pl-Holant$(\mathcal{F})$.

Proof. By Definition 5.9, disregarding a nonzero constant factor, f has the following form:

$$f = H^{\otimes n} \left(\begin{bmatrix} 1 \\ \alpha \end{bmatrix}^{\otimes n} + i^r \begin{bmatrix} 1 \\ -\alpha \end{bmatrix}^{\otimes n} \right),$$

for some integer $r \in [4]$. We do an orthogonal transformation H^{-1} and by Theorem 1.2, Pl-Holant$(\mathcal{F}) \equiv$ Pl-Holant$(H^{-1}\mathcal{F})$. Let $\bar{f} = \left(H^{-1}\right)^{\otimes n} f \in H^{-1}\mathcal{F}$. It is easy to verify that $\bar{f}_{k+2} = i\bar{f}_k$.

A self loop on $\bar{f}$ yields $\bar{f}'$, where $\bar{f}'_k = \bar{f}_k + \bar{f}_{k+2} = (1 + i)f_k$. Thus up to the factor $(1 + i)$, $\bar{f}'$ is just the first $n - 2$ entries of $\bar{f}$. By doing zero or more self loops, we eventually obtain a quaternary signature when n is even or a ternary one when n is odd. There are eight cases depending on $r \in [4]$ and the parity of n. We list all of them, disregarding nonzero factors:

$$[1,0,i,0],\ [0,1,0,i],\ [1,\alpha i,i,-\alpha],\ [1,-\alpha i,i,\alpha] \text{ for odd } n;$$
$$[1,0,i,0,-1],\ [0,1,0,i,0],\ [1,\alpha i,i,-\alpha,-1],\ [1,-\alpha i,i,\alpha,-1] \text{ for even } n.$$

However, for any case, we can realize the signature $[1,0,i]$, due to the following. (In the calculations below, we omit certain nonzero constant factors without explanation.)

- $[1,0,i,0]$: Another self loop gives $[1,0]$. Connect it back to the ternary to get $[1,0,i]$.
- $[0,1,0,i]$: Another self loop gives $[0,1]$. Connect it back to the ternary to get $[1,0,i]$.
- $[1,\alpha i,i,-\alpha]$: Another self loop gives $[1,\alpha i]$. Connect two copies of it to the ternary to get $[1,-\alpha]$. Then connect this back to the ternary to finally get $[1,0,i]$. See Figure 5.2a.
- $[1,-\alpha i,i,\alpha]$: Same construction as the previous case.
- $[1,0,i,0,-1]$: Another self loop gives $[1,0,i]$ directly.
- $[0,1,0,i,0]$: Another self loop gives $[0,1,0]$. Connect it back to the quaternary to get $[1,0,i]$.
- $[1,\alpha i,i,-\alpha,-1]$: Another self loop gives $[1,\alpha i,i]$. Connect two copies of it together to get $[1,-\alpha,-i]$. Connect this back to the quaternary to get $[1,0,i]$. See Figure 5.2b.
- $[1,-\alpha i,i,\alpha,-1]$: Same construction as the previous case.

(a) Vertices assigned $[1,\alpha i,i,-\alpha]$. (b) Vertices assigned $[1,\alpha i,i,-\alpha,-1]$.

Figure 5.2: Constructions to realize $[1,0,i]$.

With $[1,0,i]$ in hand, we can connect three copies of it to get $[1,0,-i]$. Now we construct a bipartite graph, with $H^{-1}\mathcal{F}\cup\{=_2\}$ on the right side and $[1,0,-i]$ on the left, and do a holographic transformation by $Y = \left[\begin{smallmatrix} \alpha & 1 \\ -\alpha & 1 \end{smallmatrix}\right]$ to get

$$\begin{aligned}
&\text{Pl-Holant}\left([1,0,-i] \mid H^{-1}\mathcal{F} \cup \{\bar{f}, =_2\}\right) \\
\equiv_T\ &\text{Pl-Holant}\left([1,0,-i](Y^{-1})^{\otimes 2} \mid YH^{-1}\mathcal{F} \cup \{Y^{\otimes n}\bar{f}, Y^{\otimes 2}(=_2)\}\right) \\
\equiv_T\ &\text{Pl-Holant}\left(\tfrac{1}{2i}[1,0,1] \mid YH^{-1}\mathcal{F} \cup \{[1,0,\ldots,0,i^k],[1,-i,1]\}\right) \\
\equiv_T\ &\text{Pl-Holant}\left(YH^{-1}\mathcal{F} \cup \{[1,0,\ldots,0,i^k],[1,-i,1]\}\right).
\end{aligned}$$

Notice that $\bar{f}$ becomes $g := [1,0,\ldots,0,i^k]$ where $k = r+2n$ (after normalizing the first entry) and $=_2$ on the right becomes $[1,-i,1]$. On the other side, $[1,0,-i]$ becomes $[1,0,1]$. With g, we can construct all EQUALITY signatures of even arity as follows. First connect 4 copies of g together arbitrarily to get an EQUALITY $=_t$ of some arity $t \geqslant 3$. One or more self-loops of $=_t$ gives $=_3$ or $=_4$ eventually. Then $=_4$ is just two $=_3$ connected by one edge. From $=_4$ it is easy to construct any EQUALITY of even arity. Hence, Pl-#CSP$^2(YH^{-1}\mathcal{F} \cup \{[1,-i,1]\}) \leqslant_T$ Pl-Holant$(\mathcal{F})$. □

Corollary 5.28. *Let $\mathcal{F}$ be a set of signatures. Suppose there exists $f \in \mathcal{F}$ which is a non-degenerate signature of arity $n \geqslant 3$ in $\mathcal{A}_3$. Then* Pl-Holant$(\mathcal{F})$ *is #**P**-hard unless $\mathcal{F}$ is $\mathcal{A}$- or $\mathcal{M}$-transformable, in which case* Pl-Holant$(\mathcal{F})$ *is tractable.*

Proof. Assume that $f \in \mathcal{A}_3$ with $H \in \mathbf{O}_2(\mathbb{C})$. By Lemma 5.27, we have Pl-#CSP$^2(YH^{-1}\mathcal{F} \cup \{[1,-i,1]\}) \leqslant_T$ Pl-Holant$(\mathcal{F})$, where $Y = \left[\begin{smallmatrix} \alpha & 1 \\ -\alpha & 1 \end{smallmatrix}\right]$ and $\alpha = e^{\pi i/4}$. Let $g = [1,-i,1]$ and $\mathcal{F}' = YH^{-1}\mathcal{F} \cup \{g\}$.

We apply Theorem 5.22 to Pl-#CSP$^2(\mathcal{F}')$. Then Pl-#CSP$^2(\mathcal{F}')$ (and hence Pl-Holant(F)) is #**P**-hard unless $\mathcal{F}' \subseteq \mathcal{P}$, $\mathcal{F}' \subseteq \left[\begin{smallmatrix} 1 & 0 \\ 0 & i^r \end{smallmatrix}\right]\overline{\mathcal{M}}$ for some integer $0 \leqslant r \leqslant 3$, or $\mathcal{F}' \subseteq \left[\begin{smallmatrix} 1 & 0 \\ 0 & \alpha^r \end{smallmatrix}\right]\mathcal{A}$ for some integer $0 \leqslant r \leqslant 7$ where $\alpha = e^{i\pi/4}$. Notice that $g \notin \mathcal{P}$ and hence the first case is impossible.

Suppose $\mathcal{F}' \subseteq \left[\begin{smallmatrix} 1 & 0 \\ 0 & i^r \end{smallmatrix}\right]\overline{\mathcal{M}}$ for some integer $0 \leqslant r \leqslant 3$. Then as $g \notin \left[\begin{smallmatrix} 1 & 0 \\ 0 & i^r \end{smallmatrix}\right]\overline{\mathcal{M}}$ for $r = 1, 3$, we have that $YH^{-1}\mathcal{F} \subseteq \left[\begin{smallmatrix} 1 & 0 \\ 0 & \pm 1 \end{smallmatrix}\right]\overline{\mathcal{M}}$. Moreover, notice that $\left[\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right]\overline{\mathcal{M}} = \left[\begin{smallmatrix} 1 & 1 \\ 1 & -1 \end{smallmatrix}\right]\left[\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right]\mathcal{M} = \left[\begin{smallmatrix} 1 & 1 \\ 1 & -1 \end{smallmatrix}\right]\mathcal{M} = \overline{\mathcal{M}}$. Hence $YH^{-1}\mathcal{F} \subseteq \overline{\mathcal{M}}$. Rewrite Y as $Y = \left[\begin{smallmatrix} 1 & 1 \\ -1 & 1 \end{smallmatrix}\right]\left[\begin{smallmatrix} \alpha & 0 \\ 0 & 1 \end{smallmatrix}\right]$. We deduce that

$$\begin{aligned}
H^{-1}\mathcal{F} &\subseteq \tfrac{1}{2}\left[\begin{smallmatrix} \alpha^{-1} & 0 \\ 0 & 1 \end{smallmatrix}\right]\left[\begin{smallmatrix} 1 & -1 \\ 1 & 1 \end{smallmatrix}\right]\overline{\mathcal{M}} = \tfrac{1}{2}\left[\begin{smallmatrix} \alpha^{-1} & 0 \\ 0 & 1 \end{smallmatrix}\right]\left[\begin{smallmatrix} 1 & -1 \\ 1 & 1 \end{smallmatrix}\right]\left[\begin{smallmatrix} 1 & 1 \\ 1 & -1 \end{smallmatrix}\right]\mathcal{M} \\
&= \left[\begin{smallmatrix} \alpha^{-1} & 0 \\ 0 & 1 \end{smallmatrix}\right]\left[\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right]\mathcal{M} = \mathcal{M}.
\end{aligned}$$

Hence $\mathcal{F}$ is $\mathcal{M}$-transformable in this case.

The last case is when $\mathcal{F}' \subseteq \left[\begin{smallmatrix}1 & 0\\ 0 & \alpha^r\end{smallmatrix}\right]\mathcal{A}$ for some integer $0 \leqslant r \leqslant 7$. It implies that $r = 0, 2, 4, 6$ as $g \in \left[\begin{smallmatrix}1 & 0\\ 0 & \alpha^r\end{smallmatrix}\right]\mathcal{A}$ and $g \notin \left[\begin{smallmatrix}1 & 0\\ 0 & \alpha\end{smallmatrix}\right]\mathcal{A}$. That is, $\mathcal{F}' \subseteq \left[\begin{smallmatrix}1 & 0\\ 0 & i^l\end{smallmatrix}\right]\mathcal{A}$ for some integer $0 \leqslant l \leqslant 3$. Notice that $\left[\begin{smallmatrix}1 & 0\\ 0 & i^l\end{smallmatrix}\right] \in \operatorname{Stab}(\mathcal{A})$. It implies that $YH^{-1}\mathcal{F} \subseteq \mathcal{A}$. Again, rewriting Y as $Y = \left[\begin{smallmatrix}1 & 1\\ -1 & 1\end{smallmatrix}\right]\left[\begin{smallmatrix}\alpha & 0\\ 0 & 1\end{smallmatrix}\right]$, we have

$$H^{-1}\mathcal{F} \subseteq \tfrac{1}{2}\left[\begin{smallmatrix}\alpha^{-1} & 0\\ 0 & 1\end{smallmatrix}\right]\left[\begin{smallmatrix}1 & -1\\ 1 & 1\end{smallmatrix}\right]\mathcal{A} = \tfrac{1}{2}\left[\begin{smallmatrix}\alpha^{-1} & 0\\ 0 & 1\end{smallmatrix}\right]\mathcal{A}.$$

Therefore $\mathcal{F}$ is $\mathcal{A}$-transformable. This finishes the proof. □

Similarly, we specialize Corollary 5.28 to our need.

Corollary 5.29. *Let* f *be a non-degenerate signature of arity* $n \geqslant 5$. *Let* f' *be* f *with a self loop, and* f' *is non-degenerate and* $f' \in \mathcal{A}_3$. *Then* Pl-Holant(f) *is #***P***-hard unless* f *is* $\mathcal{A}$*- or* $\mathcal{M}$*-transformable, in which case* Pl-Holant(f) *is tractable.*

The next case is when f is in $\mathcal{M}_2$ but not $\mathcal{P}_2$.

Lemma 5.30. *Let* $f \in \mathcal{M}_2 \setminus \mathcal{P}_2$ *be a non-degenerate signature of arity* $n \geqslant 3$ *with an orthogonal transformation* H. *Then* $f = cH^{\otimes n}\left(\left[\begin{smallmatrix}1\\ \gamma\end{smallmatrix}\right]^{\otimes n} \pm \left[\begin{smallmatrix}1\\ -\gamma\end{smallmatrix}\right]^{\otimes n}\right)$ *for some* $c \neq 0$ *and* $\gamma \neq 0, \pm i$.

Let $\mathcal{F}$ *be a set of signatures containing* f, $T = H\left[\begin{smallmatrix}1 & 1\\ \gamma & -\gamma\end{smallmatrix}\right]$, *and* $g = [1+\gamma^2, 1-\gamma^2, 1+\gamma^2]$. *Then,*

$$\text{Pl-\#CSP}^2(T^{-1}\mathcal{F}, g) \leqslant_T \text{Pl-Holant}(\mathcal{F}). \tag{5.1}$$

Proof. The first claim follows from Definition 5.16, Lemma 5.8, and the fact that $\mathcal{A}_2 = \mathcal{P}_2$. In the rest we show (5.1). We will ignore the nonzero factor c.

First assume that $f = H^{\otimes n}\left(\left[\begin{smallmatrix}1\\ \gamma\end{smallmatrix}\right]^{\otimes n} + \left[\begin{smallmatrix}1\\ -\gamma\end{smallmatrix}\right]^{\otimes n}\right)$ with the $+$ sign. We do the transformation T:

$$\begin{aligned}\text{Pl-Holant}(=_2 \mid f, \mathcal{F}) &\equiv_T \text{Pl-Holant}\left([1,0,1]H^{\otimes 2}\left[\begin{smallmatrix}1 & 1\\ \gamma & -\gamma\end{smallmatrix}\right]^{\otimes 2} \;\middle|\; \left(\left[\begin{smallmatrix}1 & 1\\ \gamma & -\gamma\end{smallmatrix}\right]^{-1}\right)^{\otimes n}\left(H^{-1}\right)^{\otimes n} f, T^{-1}\mathcal{F}\right)\\ &\equiv_T \text{Pl-Holant}\left(g \mid =_n, T^{-1}\mathcal{F}\right).\end{aligned}$$

By connecting g to $=_n$, we get $=_{n-2}$ up to a constant factor of $1+\gamma^2 \neq 0$ as $\gamma \neq \pm i$. We repeat this process. If n is even, then we get $=_2$ eventually, which is on the right hand side. If n is odd, then eventually we get $=_3$ and $(=_1) = [1,1]$ on the right. Connecting $[1,1]$ to g we get $2[1,1]$ on the left. Then connecting $[1,1]$ to $=_3$ we get $=_2$ on the right. To summarize, we get that

$$\begin{aligned}\text{Pl-Holant}\left(g \mid =_2, =_n, T^{-1}\mathcal{F}\right) &\leqslant_T \text{Pl-Holant}\left(g \mid =_n, T^{-1}\mathcal{F}\right)\\ &\leqslant_T \text{Pl-Holant}\left(f, \mathcal{F}\right). \qquad (5.2)\end{aligned}$$

Then we show that

$$\text{Pl-Holant}\left(=_2, g \mid =_2, =_n, T^{-1}\mathcal{F}\right) \leqslant_T \text{Pl-Holant}\left(g \mid =_2, =_n, T^{-1}\mathcal{F}\right). \qquad (5.3)$$

Let $N = \left[\begin{smallmatrix} 1+\gamma^2 & 1-\gamma^2 \\ 1-\gamma^2 & 1+\gamma^2 \end{smallmatrix}\right]$ be the signature matrix of g. If there is a positive integer k and a nonzero constant c' such that $N^k = c' I_2$, where I_2 is the 2-by-2 identity matrix, then we may directly implement $=_2$ on the left by connecting k copies of $[1+\gamma^2, 1-\gamma^2, 1+\gamma^2]$ via $=_2$ on the right. It implies (5.3) holds.

Otherwise such k and c' do not exist. The two eigenvalues of N are $\lambda_1 = 2$ and $\lambda_2 = 2\gamma^2$. If $\lambda_1 = \lambda_2$, then $\gamma^2 = 1$ and $N = \left[\begin{smallmatrix} 2 & 0 \\ 0 & 2 \end{smallmatrix}\right]$. Contradiction. Hence $\lambda_1 \neq \lambda_2$, and N is diagonalizable. Let $N = P \left[\begin{smallmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{smallmatrix}\right] P^{-1}$, for some non-singular matrix P. By connecting l many copies of N on the left via $=_2$ on the right, where l is a positive integer, we can implement $N^l = P \left[\begin{smallmatrix} \lambda_1^l & 0 \\ 0 & \lambda_2^l \end{smallmatrix}\right] P^{-1}$. Since N does not have finite order up to a scalar, for any positive integer l, $(\lambda_1/\lambda_2)^l \neq 1$.

Consider an instance Ω of Pl-Holant $\left(=_2, g \mid =_2, =_n, T^{-1}\mathcal{F}\right)$. Suppose that the left $=_2$ appears t times. Let l be a positive integer. We obtain Ω_l from Ω by replacing each occurrence of $=_2$ on the left with N^l.

Since $N^l = P \left[\begin{smallmatrix} \lambda_1^l & 0 \\ 0 & \lambda_2^l \end{smallmatrix}\right] P^{-1}$, we can view our construction of Ω_l as replacing N^l by 3 signatures, with matrix P, $\left[\begin{smallmatrix} \lambda_1^l & 0 \\ 0 & \lambda_2^l \end{smallmatrix}\right]$, and P^{-1}, respectively. This does not change the Holant value,

We stratify the assignments in Ω_l based on the assignments to the t occurrences of the signature whose matrix is the diagonal matrix $\left[\begin{smallmatrix} \lambda_1^l & 0 \\ 0 & \lambda_2^l \end{smallmatrix}\right]$. Suppose there are i many times it was assigned 00 with function value λ_1^l, and j times 11 with function value λ_2^l. Clearly $i + j = t$ if the assignment has a nonzero evaluation. Let c_{ij} be the sum over all such assignments of the products of evaluations of all signatures (including the signatures corresponding to matrices

P and P^{-1}) in Ω_l except for this diagonal one. Then

$$\begin{aligned}\text{Holant}_{\Omega_l} &= \sum_{i+j=t} \left(\lambda_1^l\right)^i \left(\lambda_2^l\right)^j c_{ij} \\ &= \lambda_2^{lt} \sum_{0 \leqslant i \leqslant t} \left(\left(\frac{\lambda_1}{\lambda_2} \right)^l \right)^i c_{i,t-i}.\end{aligned}$$

By an oracle of Pl-Holant $\left(g \mid =_2, =_n, T^{-1}\mathcal{F}\right)$, we can get Holant_{Ω_l} for any $1 \leqslant l \leqslant t+1$. Recall that for any positive integer l, $(\lambda_1/\lambda_2)^l \neq 1$. This implies that for any two distinct integers $i, j \geqslant 0$, $(\lambda_1/\lambda_2)^i \neq (\lambda_1/\lambda_2)^j$. Therefore we get a non-singular Vandermonde system. We can solve all c_{ij} for $i+j=t$ given Holant_{Ω_l} for all $1 \leqslant l \leqslant t+1$. Then notice that $\sum_{i+j=t} c_{ij}$ is the Holant value of Ω_l by replacing both λ_1^l and λ_2^l with 1, which is the instance Ω as $PI_2P^{-1} = I_2$. Therefore we may compute Holant_Ω via $t+1$ many oracle calls to Pl-Holant $\left(g \mid =_2, =_n, T^{-1}\mathcal{F}\right)$. This finishes the reduction in (5.3).

In the left hand side of (5.3) we have $=_2$ on both sides. Therefore we may lift the bipartite restriction. Combining it with (5.2), we get

$$\text{Pl-Holant}\left(=_n, g, T^{-1}\mathcal{F}\right) \leqslant_T \text{Pl-Holant}\,(f, \mathcal{F}).$$

Notice that given an equality of arity $n \geqslant 3$, we can always construct all equalities of even arity in a planar way, regardless of the parity of n. Therefore, we have Pl-#CSP$^2(T^{-1}\mathcal{F}, g) \leqslant_T$ Pl-Holant $(\mathcal{F})$.

To prove (5.1), there is another case that $f = H^{\otimes n}\left(\left[\begin{smallmatrix}1\\ \gamma\end{smallmatrix}\right]^{\otimes n} - \left[\begin{smallmatrix}1\\ -\gamma\end{smallmatrix}\right]^{\otimes n}\right)$, with the $-$ sign. Again we do the transformation T, where $(T^{-1})^{\otimes} f = [1, 0, \ldots, 0, -1]$ has arity n:

$$\text{Pl-Holant}\,(=_2 \mid f, \mathcal{F}) \equiv_T \text{Pl-Holant}\left(g \mid [1, 0, \ldots, 0, -1], T^{-1}\mathcal{F}\right).$$

We then do the same construction as in the previous case of connecting g to $[1, 0, \ldots, 0, -1]$ repeatedly. Depending on the parity of n, we have two cases.

1. If n is odd, then eventually we get $[1, 0, 0, -1]$ and $[1, -1]$ on the right as $\gamma \neq \pm i$, and therefore $2\gamma^2[1, -1]$, i.e., $[1, -1]$ on the left as $\gamma \neq 0$. Then connecting $[1, -1]$ to $[1, 0, 0, -1]$

we get $=_2$ on the right. Thus, for odd n,

$$\begin{aligned}\text{Pl-Holant}\left(g \mid =_2, [1,0,\ldots,0,-1], T^{-1}\mathcal{F}\right) &\leqslant_T \text{Pl-Holant}\left(g \mid [1,0,\ldots,0,-1], T^{-1}\mathcal{F}\right)\\ &\leqslant_T \text{Pl-Holant}\left(f, \mathcal{F}\right).\end{aligned}$$

Notice that our previous binary interpolation proof only relies on g and $=_2$. Hence we get

$$\begin{aligned}\text{Pl-Holant}\left(g \mid =_2, [1,0,\ldots,0,-1], T^{-1}\mathcal{F}\right) &\geqslant_T \text{Pl-Holant}\left(=_2, g \mid =_2, [1,0,\ldots,0,-1], T^{-1}\mathcal{F}\right)\\ &\equiv_T \text{Pl-Holant}([1,0,\ldots,0,-1], g, T^{-1}\mathcal{F}).\end{aligned}$$

Moreover it is straightforward to construct all even equalities from $[1,0,\ldots,0,-1]$ in the normal Pl-Holant setting as $n \geqslant 5$. Combining everything together gives us

$$\text{Pl-\#CSP}^2(g, T^{-1}\mathcal{F}) \leqslant_T \text{Pl-Holant}\left(\mathcal{F}\right).$$

2. Otherwise n is even. By the same construction of connecting g to $[1,0,\ldots,0,-1]$ repeatedly, we get $[1,0,0,0,-1]$ and $[1,0,-1]$ on the right eventually. Then we connect two copies of g via $[1,0,-1]$, resulting in $\left[\begin{smallmatrix}1+\gamma^2 & 1-\gamma^2\\ 1-\gamma^2 & 1+\gamma^2\end{smallmatrix}\right]\left[\begin{smallmatrix}1 & 0\\ 0 & -1\end{smallmatrix}\right]\left[\begin{smallmatrix}1+\gamma^2 & 1-\gamma^2\\ 1-\gamma^2 & 1+\gamma^2\end{smallmatrix}\right] = 4\gamma^2\left[\begin{smallmatrix}1 & 0\\ 0 & -1\end{smallmatrix}\right]$ on the left. Then connect $[1,0,-1]$ to $[1,0,0,0,-1]$ to get $[1,0,1]$ on the right. At last we connect two $[1,0,-1]$'s on the left via $[1,0,1]$ on the right to get $[1,0,1]$ on the left. Then it reduces to the previous case. □

Corollary 5.31. *Let $\mathcal{F}$ be a set of signatures. Suppose there exists $f \in \mathcal{F}$ which is a non-degenerate signature of arity $n \geqslant 3$ in $\mathcal{M}_2 \setminus \mathcal{P}_2$. Then* Pl-Holant$(\mathcal{F})$ *is #**P**-hard unless $\mathcal{F}$ is $\mathcal{A}$-, $\mathcal{P}$-, or $\mathcal{M}$-transformable, in which case* Pl-Holant$(\mathcal{F})$ *is tractable.*

Proof. By Lemma 5.30, we have (5.1). We apply Theorem 5.22 to Pl-#CSP$^2(T^{-1}\mathcal{F}, g)$. Then we have that Pl-#CSP$^2(T^{-1}\mathcal{F}, g)$ (and hence Pl-Holant$(\mathcal{F})$) is #**P**-hard unless $T^{-1}\mathcal{F} \cup \{g\} \subseteq \mathcal{P}$, or $T^{-1}\mathcal{F} \cup \{g\} \subseteq \left[\begin{smallmatrix}1 & 0\\ 0 & i^r\end{smallmatrix}\right]\widehat{\mathcal{M}}$ for some integer $0 \leqslant r \leqslant 3$, or $T^{-1}\mathcal{F} \cup \{g\} \subseteq \left[\begin{smallmatrix}1 & 0\\ 0 & \alpha^r\end{smallmatrix}\right]\mathcal{A}$ for some integer $0 \leqslant r \leqslant 7$ where $\alpha = e^{i\pi/4}$. We discuss each case.

1. The first case is that $T^{-1}\mathcal{F} \cup \{g\} \subseteq \mathcal{P}$. Recall that $\gamma \neq 0$ or $\pm i$, it can be verified that $g \notin \mathcal{P}$

unless $\gamma^2 = 1$. Hence $\gamma = \pm 1$. In either case we have that $\left[\begin{smallmatrix} 1 & 1 \\ \gamma & -\gamma \end{smallmatrix}\right]$ is an orthogonal matrix up to a nonzero scalar, and hence so is T. It implies that $\mathcal{F}$ is $\mathcal{P}$-transformable.

2. Next suppose $T^{-1}\mathcal{F} \cup \{g\} \subseteq \left[\begin{smallmatrix} 1 & 0 \\ 0 & i^r \end{smallmatrix}\right] \overline{\mathcal{M}}$ for some integer $0 \leqslant r \leqslant 3$. If $\gamma = \pm 1$, then T is an orthogonal matrix as $\left[\begin{smallmatrix} 1 & 1 \\ \gamma & -\gamma \end{smallmatrix}\right]$ is, up to a factor of $\frac{1}{\sqrt{2}}$. Hence $\mathcal{F}$ is $\mathcal{M}$-transformable, as $\mathcal{F} \subseteq T \left[\begin{smallmatrix} 1 & 0 \\ 0 & i^r \end{smallmatrix}\right] \left[\begin{smallmatrix} 1 & 1 \\ 1 & -1 \end{smallmatrix}\right] \mathcal{M}$ and $(=_2)\left(T \left[\begin{smallmatrix} 1 & 0 \\ 0 & i^r \end{smallmatrix}\right] \left[\begin{smallmatrix} 1 & 1 \\ 1 & -1 \end{smallmatrix}\right]\right)^{\otimes 2}$ is either $[1, 0, 1]$ when $r = 0, 2$, or $[0, 1, 0]$ when $r = 1, 3$, up to a nonzero factor.

   Otherwise $\gamma^2 \neq 1$ and it is straightforward to verify that $g \notin \left[\begin{smallmatrix} 1 & 0 \\ 0 & i^r \end{smallmatrix}\right] \overline{\mathcal{M}}$ for $r = 1, 3$. Hence we may assume that $T^{-1}\mathcal{F} \subseteq \left[\begin{smallmatrix} 1 & 0 \\ 0 & \pm 1 \end{smallmatrix}\right] \overline{\mathcal{M}}$. Moreover, $\left[\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right] \overline{\mathcal{M}} = \left[\begin{smallmatrix} 1 & 1 \\ 1 & -1 \end{smallmatrix}\right] \left[\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right] \mathcal{M} = \left[\begin{smallmatrix} 1 & 1 \\ 1 & -1 \end{smallmatrix}\right] \mathcal{M} = \overline{\mathcal{M}}$. Then $T^{-1}\mathcal{F} \subseteq \overline{\mathcal{M}}$. As $T^{-1} = \left[\begin{smallmatrix} 1 & 1 \\ \gamma & -\gamma \end{smallmatrix}\right]^{-1} H^{-1}$, it implies that

$$\begin{aligned} H^{-1}\mathcal{F} \subseteq \left[\begin{smallmatrix} 1 & 1 \\ \gamma & -\gamma \end{smallmatrix}\right] \overline{\mathcal{M}} &= \left[\begin{smallmatrix} 1 & 0 \\ 0 & \gamma \end{smallmatrix}\right] \left[\begin{smallmatrix} 1 & 1 \\ 1 & -1 \end{smallmatrix}\right] \left[\begin{smallmatrix} 1 & 1 \\ 1 & -1 \end{smallmatrix}\right] \mathcal{M} \\ &= \left[\begin{smallmatrix} 1 & 0 \\ 0 & \gamma \end{smallmatrix}\right] \mathcal{M} = \mathcal{M}. \end{aligned}$$

   Hence $\mathcal{F} \subseteq H\mathcal{M}$ and $\mathcal{F}$ is $\mathcal{M}$-transformable.

3. In the last case, $T^{-1}\mathcal{F} \cup \{g\} \subseteq \left[\begin{smallmatrix} 1 & 0 \\ 0 & \alpha^r \end{smallmatrix}\right] \mathcal{A}$ for some integer $0 \leqslant r \leqslant 7$. If $\gamma = \pm 1$, then T is an orthogonal matrix as $\left[\begin{smallmatrix} 1 & 1 \\ \gamma & -\gamma \end{smallmatrix}\right]$ is, up to a factor of $\frac{1}{\sqrt{2}}$. Hence $\mathcal{F}$ is $\mathcal{A}$-transformable, as $\mathcal{F} \subseteq T \left[\begin{smallmatrix} 1 & 0 \\ 0 & \alpha^r \end{smallmatrix}\right] \mathcal{A}$ and $(=_2)\left(T \left[\begin{smallmatrix} 1 & 0 \\ 0 & \alpha^r \end{smallmatrix}\right]\right)^{\otimes 2}$ is $[1, 0, i^r] \in \mathcal{A}$, up to a nonzero factor.

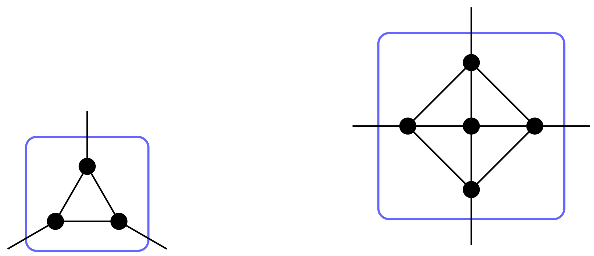
   Otherwise $\gamma^2 \neq 1$ and $g \notin \left[\begin{smallmatrix} 1 & 0 \\ 0 & \alpha^r \end{smallmatrix}\right] \mathcal{A}$ for any integer $r = 1, 3, 5, 7$. Hence $T^{-1}\mathcal{F} \cup \{g\} \subseteq \mathcal{A}$ as $\left[\begin{smallmatrix} 1 & 0 \\ 0 & i^r \end{smallmatrix}\right] \mathcal{A} = \mathcal{A}$ for any integer $0 \leqslant r \leqslant 3$. If $\frac{1+\gamma^2}{1-\gamma^2} \neq \pm i$, then one can check that $g \notin \mathcal{A}$. A contradiction. Otherwise $\frac{1+\gamma^2}{1-\gamma^2} = \pm i$. It implies that $\gamma = \alpha^l$ for some integer $l = 1, 3, 5, 7$. We may assume $l = 1$ as other cases are similar. In this case it is possible that $T^{-1}\mathcal{F} \cup \{g\} \subseteq \mathcal{A}$. As $T^{-1} = \left[\begin{smallmatrix} 1 & 1 \\ \gamma & -\gamma \end{smallmatrix}\right]^{-1} H^{-1} = \left[\begin{smallmatrix} 1 & 1 \\ \alpha & -\alpha \end{smallmatrix}\right]^{-1} H^{-1}$, it implies that

$$H^{-1}\mathcal{F} \subseteq \left[\begin{smallmatrix} 1 & 1 \\ \alpha & -\alpha \end{smallmatrix}\right] \mathcal{A} = \left[\begin{smallmatrix} 1 & 0 \\ 0 & \alpha \end{smallmatrix}\right] \left[\begin{smallmatrix} 1 & 1 \\ 1 & -1 \end{smallmatrix}\right] \mathcal{A} = \left[\begin{smallmatrix} 1 & 0 \\ 0 & \alpha \end{smallmatrix}\right] \mathcal{A}.$$

   Hence, $\mathcal{F}$ is $\mathcal{A}$-transformable by Lemma 5.5. This finishes the proof. □

Corollary 5.31 leads to the following specialization.

Corollary 5.32. *Let* f *be a non-degenerate signature of arity* $n \geqslant 5$*. Let* f' *be* f *with a self loop, and* f' *is non-degenerate and* $f' \in \mathcal{M}_2 \setminus \mathcal{P}_2$*. Then* Pl-Holant$(f)$ *is* #**P***-hard unless* f *is* $\mathcal{A}$-, $\mathcal{P}$-, *or* $\mathcal{M}$*-transformable, in which case* Pl-Holant(f) *is tractable.*



(a) Triangle gadget (b) Planar tetrahedron gadget

Figure 5.3: Two gadgets used to create a signature in $\mathcal{M}_2 \setminus \mathcal{P}_2$.

We can reduce the case of $f \in \mathcal{M}_3$ to the previous case.

Lemma 5.33. *Let* $\mathcal{F}$ *be a set of signatures. Suppose there exists* $f \in \mathcal{F}$ *which is a non-degenerate signature of arity* $n \geqslant 3$ *in* $\mathcal{M}_3$ *with* $H \in \mathbf{O}_2(\mathbb{C})$*. Then* Pl-Holant$(\mathcal{F})$ *is* #**P***-hard unless* $\mathcal{F} \subseteq H\mathcal{M}$*, in which case* $\mathcal{F}$ *is* $\mathcal{M}$*-transformable and* Pl-Holant$(\mathcal{F})$ *is tractable.*

Proof. We first claim that Pl-Holant$(\mathcal{F})$ is #**P**-hard unless $\mathcal{F}$ is $\mathcal{A}$-, $\mathcal{P}$-, or $\mathcal{M}$-transformable.

By the definition of $\mathcal{M}_3$, we may assume that $f = \textsc{ExactOne}_n$ is of arity n after an orthogonal transformation H. After zero or more self loops, we can further assume that either $f = \textsc{ExactOne}_3$ or $f = \textsc{ExactOne}_4$ depending on the parity of n.

Suppose $f = \textsc{ExactOne}_3$. Consider the gadget in Figure 5.3a. We assign f to all vertices. The signature of the resulting gadget is $g = [0, 1, 0, 1]$, which is in $\mathcal{M}_2$ and not in $\mathcal{P}_2 = \mathcal{A}_2$ by Lemma 5.8. Thus, the claim follows from Corollary 5.31.

Otherwise, $f = \textsc{ExactOne}_4$. Consider the gadget in Figure 5.3b. We assign f to all vertices. Note that this is a matchgate. The signature of the resulting gadget is $[0, 2, 0, 1, 0]$, which is in $\mathcal{M}_2$ and not in $\mathcal{P}_2 = \mathcal{A}_2$ by Lemma 5.8. Thus, the claim follows from Corollary 5.31.

However, as $f \in \mathcal{F}$ and $f \in \mathcal{M}_3$, $\mathcal{F}$ cannot be $\mathcal{A}$- or $\mathcal{P}$-transformable by Lemma 5.21. Also by Lemma 5.21, if $\mathcal{F}$ is $\mathcal{M}$-transformable, then $\mathcal{F} \subseteq HD\mathcal{M}$ or $H \left[\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right] D\mathcal{M}$ for some diagonal matrix D. Notice that $D \in \mathrm{Stab}(\mathcal{M})$ and $\left[\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right] D \in \mathrm{Stab}(\mathcal{M})$. It implies that $\mathcal{F} \subseteq H\mathcal{M}$. $\square$

Once again, we specialize Lemma 5.33 to our needs.

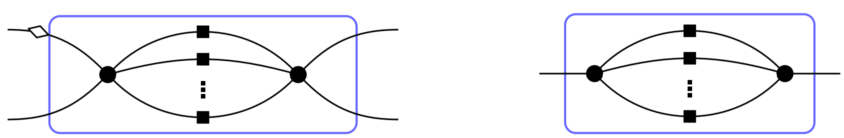
Corollary 5.34. *Let* f *be a non-degenerate signature of arity* $n \geqslant 5$*. Let* f' *be* f *with a self loop, and* f' *is non-degenerate and* $f' \in \mathcal{M}_3$*. Then* Pl-Holant(f) *is #***P***-hard unless* f *is* $\mathcal{M}$*-transformable, in which case* Pl-Holant(f) *is tractable.*

5.5 Hardness results for the Inductive Step

To finish this chapter, we will prove Theorem 5.41, which is the single signature dichotomy for Holant and Pl-Holant problems. We prove Theorem 5.41 by induction on the arity. It relies on the following key lemma. The important assumption here is that f' is non-degenerate.

Lemma 5.35. *Suppose* f *is a non-degenerate signature of arity* $n \geqslant 5$*. Let* f' *be* f *with a self loop. If* $f' \in \mathcal{P}_1 \cup \mathcal{M}_2 \cup \mathcal{A}_3 \cup \mathcal{M}_3 \cup \mathcal{V}$ *is non-degenerate, then* Pl-Holant(f) *is #***P***-hard unless* $f \in \mathcal{P}_1 \cup \mathcal{M}_2 \cup \mathcal{A}_3 \cup \mathcal{M}_3 \cup \mathcal{V}$*.*

Lemma 5.35 depends on several results, each of which handles a different case. In fact, the proof of Lemma 5.35 is a straightforward combination of Lemma 3.49 (for $\mathcal{V}$) from Section 3.7, Corollary 5.26 (for $\mathcal{P}_1$), Corollary 5.29 (for $\mathcal{A}_3$), Corollary 5.32 (for $\mathcal{M}_2 \setminus \mathcal{P}_2$), and Corollary 5.34 (for $\mathcal{M}_3$) from Section 5.4, as well as Corollary 5.37 (for $\mathcal{P}_2$), which will be proved shortly.



(a) An arity-4 construction

(b) A binary construction

Figure 5.4: Two gadgets in the Z basis. In the normal basis, circles are assigned f and squares are assigned $=_2$. In the Z basis, circles are assigned $\overline{f}$ and squares are assigned $\neq_2$.

Lemma 5.36. *Let* f *be a non-degenerate signature of arity* $n \geqslant 5$*. If* $f = Z^{\otimes n}[a, 1, 0, \ldots, 0, 1, b]$ *for some* $a, b \in \mathbb{C}$*, where the number of* 0*'s is* $n-3$*. Then* Pl-Holant(f) *is #***P***-hard.*

Proof. First we use the gadget in Figure 5.4a, where we put f on both circles, and squares are $=_2$. Let the resulting signature be $h = Z^{\otimes 4}\overline{h}$. It is easier to calculate $\overline{h}$, that is, h in the Z basis.

Indeed, $\overline{h}$ is not symmetric, but $\overline{h}$ has the following matrix representation as $n \geqslant 5$:

$$M_{\overline{h}} = \begin{bmatrix} 0 & a & a & ab+(n-2) \\ a & 2 & 2 & b \\ a & 2 & 2 & b \\ ab+(n-2) & b & b & 0 \end{bmatrix}.$$

Notice that this matrix is redundant, and $\det(\widetilde{M_{\overline{h}}}) = -4(n-2)(ab+n-2)$. If $ab \neq 2-n$, then by Corollary 3.30, Pl-Holant(h) is #**P**-hard, and so is Pl-Holant(f). Hence in the following we assume that $ab = 2-n$.

Let f' be f with a self loop. Then apply the Z transformation as follows:

$$\text{Pl-Holant}\left(=_2 \mid f, f'\right) \equiv_T \text{Pl-Holant}\left([0,1,0] \mid \overline{f}, \overline{f'}\right),$$

where $\overline{f'} = [1,0,\ldots,0,1]$ and $\overline{f} = [a,1,0,\ldots,0,1,b]$ for some $a, b \in \mathbb{C}$. We get this expression of $\overline{f'}$ because doing a self loop commutes with the operation of holographic transformations.

We connect $\overline{f'}$ to $\overline{f}$ via $[0,1,0]$, getting $[a,2,b]$. Then we connect $[a,2,b]$ to $\overline{f}$ via $[0,1,0]$ again, getting $\overline{g} = [ab+4, b, 0, \ldots, 0, a, ab+4]$ of arity $n-2$.

If $n \geqslant 7$, then we use the gadget in Figure 5.4a again, where we put g on both vertices this time. We get some signature h', which in Z basis has the following matrix representation as $n-2 \geqslant 5$:

$$M_{\overline{h'}} = \begin{bmatrix} 0 & a(ab+4) & a(ab+4) & (n-4)ab+(ab+4)^2 \\ a(ab+4) & 2ab & 2ab & b(ab+4) \\ a(ab+4) & 2ab & 2ab & b(ab+4) \\ (n-4)ab+(ab+4)^2 & b(ab+4) & b(ab+4) & 0 \end{bmatrix}.$$

Once again this matrix is redundant. It can be simplified as $ab = 2-n$. The compressed matrix is

$$\widetilde{M_{\overline{h'}}} = \begin{bmatrix} 0 & -2(n-6)a & -6n+28 \\ -(n-6)a & 8-4n & -(n-6)b \\ -6n+28 & -2(n-6)b & 0 \end{bmatrix}.$$

It is easy to compute that $\det(\widetilde{M_{\overline{h'}}}) = -8(3n-14)(ab(n-6)^2 - 6n^2 + 40n - 56) = 8(n-4)(n-2)^2(3n-14)$. Since $n \geqslant 7$, $\det(\widetilde{M_{\overline{h'}}}) > 0$. Then by Corollary 3.30, Pl-Holant(h') is #**P**-hard, and so is Pl-Holant(f).

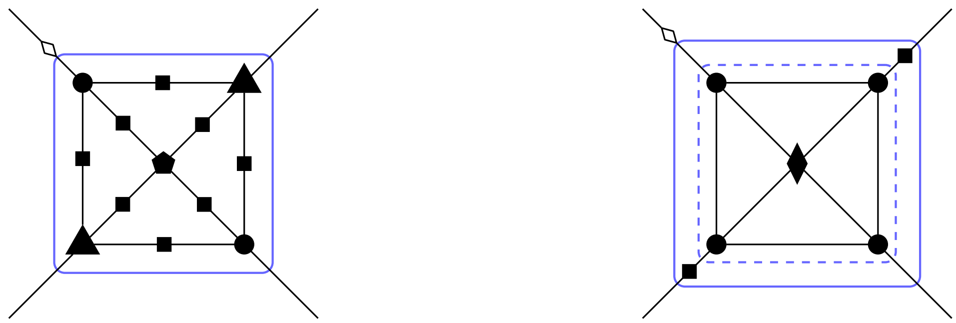
The remaining cases are $n = 6$ and $n = 5$. When $n = 6$, $ab = 2 - n = -4$. Moreover, $\overline{g}$ has arity 4 and $\overline{g} = [ab+4, b, 0, a, ab+4] = [0, b, 0, a, 0]$. We do one more self loop on g via $[0, 1, 0]$ in the Z basis, resulting in $\overline{g'} = [b, 0, a]$. Connecting $\overline{g'}$ to $\overline{f}$ via $[0, 1, 0]$, we get $\overline{g_1} = [a^2, a, 0, b, b^2]$. Hence $\det(\widetilde{M_{\overline{g_1}}}) = -4a^2b^2 = -64 \neq 0$. Then we are done by Corollary 3.30.

At last, $n = 5$ and $ab = 2 - n = -3$. We also have $\overline{g} = [ab + 4, b, a, ab + 4] = [1, b, a, 1]$. One more self loop on g via $[0, 1, 0]$ in the Z basis results in $\overline{g''} = [b, a]$. Connecting $\overline{g''}$ to $\overline{f}$ via $[0, 1, 0]$, we get $\overline{g_2} = [a^2 + b, a, 0, b, b^2 + a]$. Hence $\det(\widetilde{M_{\overline{g_2}}}) = -2(a^3 + 2a^2b^2 + b^3) = -2(a^3 + b^3 + 18)$. If $a^3 + b^3 + 18 \neq 0$, then we are done by Corollary 3.30. Otherwise $a^3 + b^3 = -18$, and we construct a binary signature $[a, 0, b]$ by doing a self-loop on $\overline{g_2}$ in the Z basis. Then we construct another unary signature by connecting $\overline{g''} = [b, a]$ to $[a, 0, b]$ via $[0, 1, 0]$, which gives $[a^2, b^2]$. Connecting $[a^2, b^2]$ to $\overline{f}$ via $[0, 1, 0]$, we have another arity-4 signature $\overline{g_3} = [ab^2 + a^2, b^2, 0, a^2, a^2b + b^2]$. We compute $\det(\widetilde{M_{\overline{g_3}}}) = -2(a^6 + a^5b^2 + a^2b^5 + b^6) = -2(a^6 + b^6 - 162)$. If $a^6 + b^6 - 162 \neq 0$, again we are done by Corollary 3.30. Otherwise $a^6 + b^6 = 162$. Together with $a^3 + b^3 = -18$ and $ab = -3$, there is no solution of a and b. This finishes the proof. □

This lemma essentially handles the case of $f' \in \mathcal{P}_2$ due to the following corollary.

Corollary 5.37. *Suppose* f *be a non-degenerate signature of arity* $n \geqslant 5$*. Let* f' *be* f *with a self loop. If* $f' \in \mathcal{P}_2$ *is non-degenerate, then* Pl-Holant(f) *is #***P***-hard.*

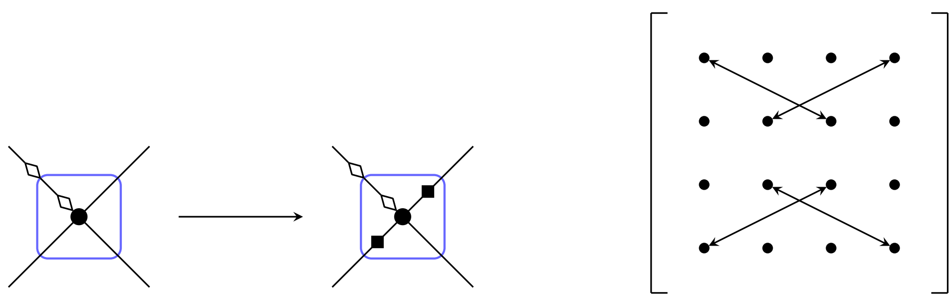
Proof. Since $f' \in \mathcal{P}_2$, we have that $f' = Z^{\otimes n-2}[1, 0, \ldots, 0, 1]$ up to an orthogonal transformation H. Since H does not change the complexity, we may assume we are under this transformation. Then f is of the form $Z^{\otimes n}[a, 1, 0, \ldots, 0, 1, b]$. The claim follows by Lemma 5.36. □



(a) $(\neq_2 \mid [0,1,0,0,0], [0,0,0,1,0], \overline{g})$-gate on right side

(b) Simpler construction with the same signature

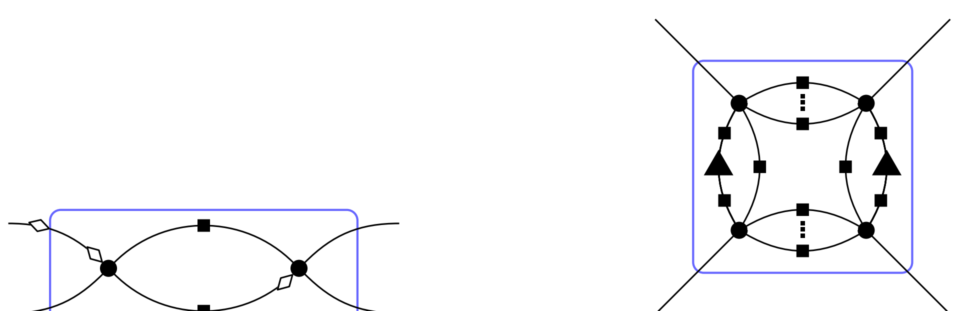
Figure 5.5: Two gadgets with the same signature used in Lemma 5.39.



(a) Negating the second and fourth inputs

(b) Movement of even Hamming weight entries

Figure 5.6: The movement of the even Hamming weight entries in the signature matrix of a quaternary signature under the negation of the second and fourth inputs (i.e. the square vertices are assigned $[0, 1, 0]$).



(a) Gadget with a useful signature matrix

(b) Gadget that realizes a partial crossover

Figure 5.7: Two quaternary gadgets used in the proof of Lemma 5.39 and 5.40.

Lemma 5.35 does not solve the case when f' is degenerate. In general, when f' is degenerate, the inductive step is straightforward unless f' is also vanishing. Lemma 5.38 and 5.40 are the two missing pieces to this end.

Lemma 5.38. *Let* $a, b \in \mathbb{C}$. *Suppose* f *is a signature of the form* $Z^{\otimes n}[a, 1, 0, \dots, 0, b]$ *with arity* $n \geqslant 3$. *If* $ab \neq 0$*, then* Pl-Holant(f) *is #**P**-hard.*

Proof. We prove by induction on n. For $n = 3$ or 4, it follows from Lemma 5.19, Theorem 1.14, and Theorem 3.39 that Pl-Holant(f) is #**P**-hard.

Now assume $n \geqslant 5$. Under a holographic transformation by $Z = \left[\begin{smallmatrix} 1 & 1 \\ i & -i \end{smallmatrix}\right]$, we have

$$\begin{aligned} \text{Pl-Holant}\,(=_2 \mid f) &\equiv_T \text{Pl-Holant}\left([1,0,1]Z^{\otimes 2} \mid (Z^{-1})^{\otimes n} f\right) \\ &\equiv_T \text{Pl-Holant}\left([0,1,0] \mid \overline{f}\right), \end{aligned}$$

where $\overline{f} = [a, 1, 0, \dots, 0, b]$. Now consider the gadget in Figure 5.4b with $\overline{f}$ assigned to both circles, and $[0, 1, 0]$ both squares. This gadget has the binary signature $\overline{g_1} = [0, ab, 2b]$, which is equivalent to $[0, a, 2]$ since $b \neq 0$. Translating back by Z to the original setting, this signature is $g_1 = [a + 1, -i, a - 1]$. This can be verified as

$$\begin{bmatrix} 1 & 1 \\ i & -i \end{bmatrix} \begin{bmatrix} 0 & a \\ a & 2 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ i & -i \end{bmatrix}^{\mathsf{T}} = 2 \begin{bmatrix} a+1 & -i \\ -i & a-1 \end{bmatrix}.$$

By the form of $\overline{g_1} = [0, ab, 2b]$ and $b \neq 0$, it follows from Lemma 3.14 that $g_1 \notin \mathcal{R}_2^{+}$. Moreover, since $a \neq 0$, g_1 is non-degenerate.

Doing a self loop on f yields $f' = Z^{\otimes n-2}[1, 0, \dots, 0]$. Connecting f' back to f, we get a binary signature $g_2 = Z^{\otimes 2}[0, 0, b]$. Once again we connect g_2 to f, the resulting signature is $h = Z^{\otimes n-2}[a, 1, 0, \dots, 0]$ of arity $n - 2 \geqslant 3$ up to the constant factor of $b \neq 0$.

Notice that h is non-degenerate and $h \in \mathcal{V}^{+}$. By Lemma 3.43, Pl-Holant(h, g_1) is #**P**-hard, hence Pl-Holant(f) is also #**P**-hard. □

The next case is similar to Lemma 5.38 but $a = 0$. We need the following technical lemma.

Lemma 5.39. *Let $\overline{g}$ be the arity 4 signature whose matrix is*

$$M_{\overline{g}} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}. \tag{5.4}$$

Then Pl-Holant $(\neq_2 \mid [0,1,0,0,0], [0,0,0,1,0], \overline{g})$ *is #**P**-hard.*

Proof. Consider the gadget in Figure 5.5a. We assign $[0,0,0,1,0]$ to the triangle vertices, $[0,1,0,0,0]$ to the circle vertices, $\overline{g}$ to the pentagon vertex, and $[0,1,0]$ to the square vertices. Let $\overline{h}$ be the signature of this gadget. By adding two more disequality signatures and then grouping appropriately, it is clear that the gadget in Figure 5.5b has the same signature of the gadget in Figure 5.5a, where the circle vertices are still assigned $[0,1,0,0,0]$, the square vertices are still assigned $[0,1,0]$, and the diamond vertex is assigned $=_4$. To compute the signature $\overline{h}$, first compute the signature $\overline{h}'$ of the inner gadget enclosed by the dashed line, which has signature matrix

$$M_{\overline{h}'} = \begin{bmatrix} 3 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix}.$$ Then by Figure 5.6, the signature matrix of $\overline{h}$ is $$M_{\overline{h}} = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 1 & 3 & 0 \\ 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}.$$

One more gadget before we finish the proof using interpolation. Consider the gadget in Figure 5.7a. We assign $\overline{h}$ to the circle vertices and $[0,1,0]$ to the square vertices. The signature of the resulting gadget is $\overline{r}$ with signature matrix $M_{\overline{r}}$ (see Figure 3.2 for the signature of a rotated copy of $\overline{h}$ that appears as the second circle vertex in Figure 5.7a), where

$$M_{\overline{r}} = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 1 & 3 & 0 \\ 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix} \left(\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \otimes \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \right) \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 3 & 1 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 6 & 4 & 0 \\ 0 & 4 & 2 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}.$$

Consider an instance Ω of Pl-Holant $(\neq_2 \mid \mathcal{F} \cup \{\bar{r}'\})$ with $\bar{r} \in \mathcal{F}$, where the signature matrix of $\bar{r}'$ is

$$M_{\bar{r}'} = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 3 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}.$$

Suppose that $\bar{r}'$ appears n times in Ω. We construct from Ω a sequence of instances Ω_s of Pl-Holant $(\neq_2 \mid \mathcal{F})$ indexed by $s \geqslant 1$. We obtain Ω_s from Ω by replacing each occurrence of $\bar{r}'$ with the gadget N_s in Figure 3.7 with $\bar{r}$ assigned to the circle vertices and $[0, 1, 0]$ assigned to the square vertices. In Ω_s, the edge corresponding to the ith significant index bit of N_s connects to the same location as the edge corresponding to the ith significant index bit of $\bar{r}'$ in Ω.

We can express the signature matrix of N_s as

$$M_{N_s} = X(XM_{\bar{r}})^s = XP \operatorname{diag}\left(1, 4 + 2\sqrt{3}, 4 - 2\sqrt{3}, 1\right)^s P^{-1},$$

where

$$X = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix} \qquad \text{and} \qquad P = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & \sqrt{3} & -\sqrt{3} & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Since $M_{\bar{r}'} = XP \operatorname{diag}\left(1, 1 + \sqrt{3}, 1 - \sqrt{3}, 1\right) P^{-1}$, we can view our construction of Ω_s as first replacing $M_{\bar{r}'}$ with $XP \operatorname{diag}\left(1, 1 + \sqrt{3}, 1 - \sqrt{3}, 1\right) P^{-1}$, which does not change the Holant value, and then replacing the diagonal matrix with the diagonal matrix $\operatorname{diag}\left(1, 4 + 2\sqrt{3}, 4 - 2\sqrt{3}, 1\right)^s$.

We stratify the assignments in Ω based on the assignments to the n occurrences of the signature whose signature matrix is the diagonal matrix

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 + \sqrt{3} & 0 & 0 \\ 0 & 0 & 1 - \sqrt{3} & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \tag{5.5}$$

We only need to consider the assignments that assign

- i many times the bit patterns 0000 or 1111,
- j many times the bit pattern 0110, and
- k many times the bit pattern 1001,

since any other assignment contributes a factor of 0. Let c_{ijk} be the sum over all such assignments of the products of evaluations of all signatures (including the signatures corresponding to the signature matrices X, P, and P^{-1}) in Ω except for signature corresponding to the signature matrix in (5.5). Then

$$\text{Holant}_{\Omega} = \sum_{i+j+k=n} \left(1+\sqrt{3}\right)^{j} \left(1-\sqrt{3}\right)^{k} c_{ijk}$$

and the value of the Holant on Ω_s, for $s \geqslant 1$, is

$$\text{Holant}_{\Omega_s} = \sum_{i+j+k=n} \left(\left(4+2\sqrt{3}\right)^{j} \left(4-2\sqrt{3}\right)^{k}\right)^{s} c_{ijk} = \sum_{i+j+k=n} \left(\left(4+2\sqrt{3}\right)^{j-k} 4^{k}\right)^{s} c_{ijk}.$$

We argue that this Vandermonde system has full rank, which is to say that

$$\left(4+2\sqrt{3}\right)^{j-k} 4^{k} \neq \left(4+2\sqrt{3}\right)^{j'-k'} 4^{k'}$$

unless $(j, k) = (j', k')$. Suppose otherwise. Then we have that $\left(4+2\sqrt{3}\right)^{j-k-(j'-k')} 4^{k-k'} = 1$. Since any nonzero integer power of $4 + 2\sqrt{3}$ is not rational, we must have $j - k = j' - k'$. Moreover, $4^{k-k'} = 1$, and hence $k = k'$ and $j = j'$.

Therefore, we can solve for the unknown c_{ijk}'s and obtain the value of Holant_{Ω}. Then after a counterclockwise rotation of $\bar{r}'$ (c.f. Figure 3.2), we are done by Corollary 3.30. □

With Lemma 5.39 at hand, we continue to prove Lemma 5.40.

Lemma 5.40. *Let* $b \in \mathbb{C}$. *Suppose* f *is a signature of the form* $Z^{\otimes n}[0, 1, 0, \ldots, 0, b]$ *with arity* $n \geqslant 4$. *If* $b \neq 0$, *then* Pl-Holant(f) *is #***P***-hard.*

Remark For $n = 3$, $Z^{\otimes 3}[0, 1, 0, b]$ is tractable, as it is $\mathcal{M}$-transformable.

Proof. If $n = 4$, then we are done by Corollary 3.30. Thus, assume that $n \geqslant 5$.

Under a holographic transformation by $Z = \left[\begin{smallmatrix} 1 & 1 \\ i & -i \end{smallmatrix}\right]$, we have

$$\begin{aligned}\text{Pl-Holant}\,(=_2 \mid f) &\equiv_T \text{Pl-Holant}\left([1,0,1]Z^{\otimes 2} \mid (Z^{-1})^{\otimes n} f\right)\\ &\equiv_T \text{Pl-Holant}\left([0,1,0] \mid \bar{f}\right),\end{aligned}$$

where $\bar{f} = [0, 1, 0, \ldots, 0, b]$. We show how to construct the following three signatures: $[0, 0, 0, 1, 0]$, $[0, 1, 0, 0, 0]$, and $\bar{g}$, where $\bar{g}$ is defined by (5.4). Then we are done by Lemma 5.39.

Consider the gadget in Figure 5.4a. We assign $\bar{f}$ to the circle vertices and $[0, 1, 0]$ to the square vertices. The signature of the resulting gadget is $[0, 0, 0, 1, 0]$ up to a nonzero factor of b.

Taking a $[0, 1, 0]$ self loop on $[0, 0, 0, 1, 0]$ gives $[0, 0, 1] = [0, 1]^{\otimes 2}$. We connect this back to $\bar{f}$ through $[0, 1, 0]$ until the arity of the resulting signature is either 4 or 5, depending on the parity of n. If n is even, then we have $[0, 1, 0, 0, 0]$ as desired. Otherwise, n is odd and we have $[0, 1, 0, 0, 0, b/0]$, where the last entry is b if $n = 5$ and 0 if $n > 5$. Connection $[0, 1]^{\otimes 2}$ through $[0, 1, 0]$ to $\bar{f}$ twice more gives $[0, 1]$. We connect this through $[0, 1, 0]$ to $[0, 1, 0, 0, 0, b/0]$ to get $[0, 1, 0, 0, 0]$ as desired.

Taking a $[0, 1, 0]$ self loop on $[0, 1, 0, 0, 0]$ gives $[1, 0, 0] = [1, 0]^{\otimes 2}$. Now consider the gadget in Figure 5.7b. We assign $\bar{f}$ to the circle vertices, $[1, 0]^{\otimes 2}$ to the triangle vertices, and $[0, 1, 0]$ to the square vertices. Up to a factor of b^2, the signature of the resulting gadget is $\bar{g}$ with signature matrix $M_{\bar{g}}$ given in (5.4). To see this, first replace the two copies of the signatures $[1, 0]^{\otimes 2}$ assigned to the triangle vertices with two copies of $[1, 0]$ each. Then notice that $\bar{f}$ simplifies to a weighted equality signature when connected to $[1, 0]$ through $[0, 1, 0]$. □

5.6 Single Signature Dichotomy

Now we are ready to prove Theorem 5.41. By Corollary 5.20, f is $\mathcal{A}$-, $\mathcal{P}$-, or $\mathcal{M}$-transformable if and only if $f \in \mathcal{P}_1 \cup \mathcal{M}_2 \cup \mathcal{A}_3 \cup \mathcal{M}_3 \cup \mathcal{M}_4$. Recall that $\mathcal{M}_4 \subset \mathcal{V}$. We list $\mathcal{M}_4$ here for merely conceptual reasons.

Theorem 5.41. *If* f *is a non-degenerate symmetric signature of arity* $n \geqslant 3$ *with complex weights in Boolean variables, then* Pl-Holant(f) *is #**P**-hard unless* $f \in \mathcal{P}_1 \cup \mathcal{M}_2 \cup \mathcal{A}_3 \cup \mathcal{M}_3 \cup \mathcal{M}_4 \cup \mathcal{V}$*, in which case the problem is computable in polynomial time.*

Holant(f) *is #**P**-hard unless* $f \in \mathcal{P}_1 \cup \mathcal{P}_2 \cup \mathcal{A}_3 \cup \mathcal{V}$*, in which case the problem is computable in polynomial time.*

Proof. Tractability for both parts follow from Lemmas 1.7, 1.9, 1.10, Corollaries 5.14, 5.20, and Theorem 3.12. We prove the first claim by induction on n. The base cases of $n = 3$ and $n = 4$ are proved in Theorem 1.14 and Theorem 3.39. Now assume $n \geqslant 5$.

With the signature f, we form a self loop to get a signature f' of arity at least 3. In general we use prime to denote the signature with a self loop. We consider separately whether or not f' is degenerate.

- Suppose $f' = [a, b]^{\otimes(n-2)}$ is degenerate. Then there are three cases to consider.

  1. If $a = b = 0$, then f' is the all zero signature. For f, this means $f_{k+2} = -f_k$ for $0 \leqslant k \leqslant n-2$, so $f \in \mathcal{P}_2$ by Lemma 5.8, and therefore Pl-Holant(f) is tractable.
  2. If $a^2 + b^2 \neq 0$, then f' is nonzero and $[a, b]$ is not a constant multiple of either $[1, i]$ or $[1, -i]$. We may normalize so that $a^2 + b^2 = 1$. Then the orthogonal transformation $\left[\begin{smallmatrix} a & b \\ -b & a \end{smallmatrix}\right]$ transforms the column vector $[a, b]$ to $[1, 0]$. Let $\bar{f}$ be the transformed signature from f, and $\overline{f'} = [1, 0]^{\otimes(n-2)}$ the transformed signature from f'.

     Since an orthogonal transformation keeps $=_2$ invariant, this transformation commutes with the operation of taking a self loop, i.e., $\overline{f'} = (\bar{f})'$. Here $(\bar{f})'$ is the function obtained from $\bar{f}$ by taking a self loop. As $(\bar{f})' = [1, 0]^{\otimes(n-2)}$, we have $\bar{f}_0 + \bar{f}_2 = 1$ and for every integer $1 \leqslant k \leqslant n-2$, we have $\bar{f}_k = -\bar{f}_{k+2}$. With one or more self loops on $(\bar{f})'$, we eventually obtain either $[1, 0]$ when n is odd or $[1, 0, 0]$ when n is even. In either case, we connect $[1, 0]$ or $[1, 0, 0]$ to $\bar{f}$ until we get an arity 4 signature, which is $\bar{g} = [\bar{f}_0, \bar{f}_1, \bar{f}_2, -\bar{f}_1, -\bar{f}_2]$. This is possible because that the parity matches and the arity of $\bar{f}$ is at least 5. We show that Pl-Holant$(\bar{g})$ is #**P**-hard. To see this, we first compute $\det(\widetilde{M_g}) = -2(\bar{f}_0 + \bar{f}_2)(\bar{f}_1^2 + \bar{f}_2^2) = -2(\bar{f}_1^2 + \bar{f}_2^2)$, since $\bar{f}_0 + \bar{f}_2 = 1$. Therefore if $\bar{f}_1^2 + \bar{f}_2^2 \neq 0$, Pl-Holant$(\bar{g})$ is #**P**-hard by Corollary 3.28. Otherwise $\bar{f}_1^2 + \bar{f}_2^2 = 0$, and we assume $\overline{f}_2 = i\overline{f}_1$ since the other case is similar. Since f is non-degenerate, $\overline{f}$ is non-degenerate, which implies $\overline{f}_2 \neq 0$. We can rewrite $\overline{g}$ as $[1,0]^{\otimes 4} - \overline{f}_2[1,i]^{\otimes 4}$. Under the holographic transformation by $T = \left[\begin{smallmatrix} 1 & (-\overline{f}_2)^{1/4} \\ 0 & i(-\overline{f}_2)^{1/4} \end{smallmatrix}\right]$, we have

$$\begin{aligned}\text{Pl-Holant}\,(=_2 \mid \overline{g}) &\equiv_T \text{Pl-Holant}\left([1,0,1]T^{\otimes 2} \mid (T^{-1})^{\otimes 4}\overline{g}\right) \\ &\equiv_T \text{Pl-Holant}\left(\overline{h} \mid =_4\right),\end{aligned}$$

where

$$\overline{h} = [1,0,1]T^{\otimes 2} = [1, (-\overline{f}_2)^{1/4}, 0]$$

and $\overline{g}$ is transformed by T^{-1} into the arity 4 equality $=_4$, since

$$T^{\otimes 4}\left(\begin{bmatrix}1\\0\end{bmatrix}^{\otimes 4} + \begin{bmatrix}0\\1\end{bmatrix}^{\otimes 4}\right) = \begin{bmatrix}1\\0\end{bmatrix}^{\otimes 4} - \overline{f}_2\begin{bmatrix}1\\i\end{bmatrix}^{\otimes 4} = \overline{g}.$$

By Theorem 1.15, Pl-Holant $\left(\overline{h} \mid =_4\right)$ is #**P**-hard as $\overline{f}_2 \neq 0$.

3. If $a^2 + b^2 = 0$ but $(a,b) \neq (0,0)$, then $[a,b]$ is a nonzero multiple of $[1, \pm i]$. Ignoring the constant multiple, we have $f' = [1,i]^{\otimes(n-2)}$ or $[1,-i]^{\otimes(n-2)}$. We consider the first case since the other case is similar.

   Under the holographic transformation $Z = \frac{1}{\sqrt{2}}\left[\begin{smallmatrix} 1 & 1 \\ i & -i \end{smallmatrix}\right]$, we have that

$$\begin{aligned}\text{Pl-Holant}\left(=_2 \mid f, f'\right) &\equiv_T \text{Pl-Holant}\left([1,0,1]Z^{\otimes 2} \mid (Z^{-1})^{\otimes n}f, (Z^{-1})^{\otimes n-2}f'\right) \\ &\equiv_T \text{Pl-Holant}\left([0,1,0] \mid \overline{f}, \overline{f'}\right),\end{aligned}$$

   where $\overline{f'} := (Z^{-1})^{\otimes n-2}f' = [1,0]^{\otimes n-2}$ and $\overline{f} := (Z^{-1})^{\otimes n}f$. Since $\overline{f}$ connecting with $[0,1,0]$ gives $\overline{f'}$, $\overline{f}$ must take the form $[a,1,0,\ldots,0,b]$ with some $a, b \in \mathbb{C}$. Depending on whether $a = 0$ or not, we apply Lemma 5.40 or Lemma 5.38 and Pl-Holant(f) is #**P**-hard.

- Suppose f' is non-degenerate. By inductive hypothesis, Pl-Holant(f) is #**P**-hard, unless $f' \in \mathscr{P}_1 \cup \mathscr{M}_2 \cup \mathscr{A}_3 \cup \mathscr{M}_3 \cup \mathscr{M}_4 \cup \mathscr{V}$. Note that f' has arity $n-2 \geqslant 3$, and every signature in $\mathcal{M}_4$ of arity at least 3 is also in $\mathcal{V}$. Hence the exceptional case is equivalent to $f' \in \mathcal{P}_1 \cup \mathcal{M}_2 \cup \mathcal{A}_3 \cup \mathcal{M}_3 \cup \mathcal{V}$. In this case, we apply Lemma 5.35 to f' and f. Hence Pl-Holant(f) is #**P**-hard, unless $f \in \mathcal{P}_1 \cup \mathcal{M}_2 \cup \mathcal{A}_3 \cup \mathcal{M}_3 \cup \mathcal{V}$. The exceptional cases imply that f is $\mathcal{A}$- or $\mathcal{P}$- or $\mathcal{M}$-transformable or vanishing, and Pl-Holant(f) is tractable.

This finishes the proof for Pl-Holant(f).

For Holant(f), suppose $f \notin \mathcal{P}_1 \cup \mathcal{P}_2 \cup \mathcal{A}_3 \cup \mathcal{V}$. If $f \notin \mathcal{P}_1 \cup \mathcal{M}_2 \cup \mathcal{A}_3 \cup \mathcal{M}_3 \cup \mathcal{M}_4 \cup \mathcal{V}$, then Pl-Holant$(f)$ is #**P**-hard, and so is Holant(f). As $\mathcal{M}_4 \subset \mathcal{V}$, we only need to show that Holant(f) is #**P**-hard, if $f \in \mathcal{M}_2 \setminus (\mathcal{P}_1 \cup \mathcal{P}_2 \cup \mathcal{A}_3)$ or $f \in \mathcal{M}_3$ (cf. Figure 5.1).

If $f \in \mathcal{M}_2 \setminus (\mathcal{P}_1 \cup \mathcal{P}_2 \cup \mathcal{A}_3)$, then by Corollary 5.14, f is not $\mathcal{A}$- or $\mathcal{P}$-transformable. It can be verified that reductions in Lemma 5.30 does not rely on planarity. Hence we have $\#\mathrm{CSP}^2(g) \leqslant_T$ Holant(f) by (5.1), where $f = H^{\otimes n} \left(\left[\begin{smallmatrix} 1 \\ \gamma \end{smallmatrix}\right]^{\otimes n} \pm \left[\begin{smallmatrix} 1 \\ -\gamma \end{smallmatrix}\right]^{\otimes n} \right)$ for some $\gamma \neq 0, \pm i$, and $g = [1+\gamma^2, 1-\gamma^2, 1+\gamma^2]$. Since g is binary, we can apply Theorem 1.15, or more conceptually, Theorem 1.15'. Hence Holant(f) is #**P**-hard unless g satisfies one of the exceptional conditions in Theorem 1.15. However it can be verified that all exceptional conditions imply that f is $\mathcal{A}$- or $\mathcal{P}$-transformable. Hence Holant(f) is always #**P**-hard.

Lastly, if $f \in \mathcal{M}_3$, then by Definition 5.16, it is easy to see that

$$\text{Holant}(f) \equiv \text{Holant}(\textsc{ExactOne}_n).$$

With zero or more self-loops on $\textsc{ExactOne}_n$, we get $\textsc{ExactOne}_3$ or $\textsc{ExactOne}_4$ eventually, depending on the parity of n. Either case is #**P**-hard by Theorem 1.14 or Theorem 3.39, combined with Lemma 5.21. □

Chapter 6

Dichotomy for Holant Problems

One way to look at Holant problems is #CSP where every variable appears exactly twice. As we have seen so far, $\#\mathrm{CSP}^2$, where every variable appears even number of times, plays a key role in our single signature dichotomy, Theorem 5.41. In order to classify the complexity of Holant problems, in both [HL12] (for real weights) and [CGW13] (for complex weights), a dichotomy for $\#\mathrm{CSP}^d$ played a similar role. We state the $\#\mathrm{CSP}^d$ dichotomy as follows.

Theorem 6.1 (Theorem IV.1 in [HL12])**.** *Let* $\mathcal{T}_k = \left\{ \left[\begin{smallmatrix} 1 & 0 \\ 0 & \omega \end{smallmatrix}\right] \in \mathbb{C}^{2\times 2} \mid \omega^k = 1 \right\}$, $d \geqslant 1$ *be an integer, and* $\mathcal{F}$ *be any set of symmetric, complex-valued signatures in Boolean variables. Then* $\#\mathrm{CSP}^d(\mathcal{F})$ *is* #**P***-hard unless there exists a* $T \in \mathcal{T}_{4d}$ *such that* $T\mathcal{F} \subseteq \mathcal{P}$ *or* $T\mathcal{F} \subseteq \mathcal{A}$*, in which case the problem is computable in polynomial time.*

Recall that we use $\mathcal{EQ}_d$ to denote the set $\{=_{kd} \mid k \in \mathbb{N}^+\}$. Then $\#\mathrm{CSP}^d(\mathcal{F}) \equiv \mathrm{Holant}(\mathcal{EQ}_d \mid \mathcal{F})$. If we restrict the Boolean #CSP dichotomy, Theorem 1.16, to symmetric functions, then it is a special case ($d = 1$) of Theorem 6.1.

Given that $\mathcal{M}$-transformable signatures are the only newly tractable planar case for symmetric Boolean #CSP (cf. Theorem 1.16 and Theorem 4.1), one would conjecture that this is also the case generalizing Theorem 6.1 to planar graphs. Surprisingly, the putative form of a Pl-$\#\mathrm{CSP}^d$ dichotomy does not hold. For example, Pl-$\#\mathrm{CSP}^d([0,1,0,0])$ is #**P**-hard when $d \leqslant 4$, but tractable when $d \geqslant 5$, while the set $\{[0,1,0,0]\} \cup \mathcal{EQ}_d$ is not $\mathcal{M}$-transformable. We will show this in Section 6.1.

Let $f \in \mathcal{P}_2$ be a symmetric signature, that is, $f = Z^{\otimes d}(=_d)$. Then for any signature set $\mathcal{F}$, we do a transformation of $Z = \frac{1}{\sqrt{2}}\left[\begin{smallmatrix}1 & 1\\ i & -i\end{smallmatrix}\right]$,

$$\begin{aligned}\operatorname{Holant}(\{f\} \cup \mathcal{F}) &\equiv \operatorname{Holant}\left(=_2 Z^{\otimes 2} \mid Z^{-1}(\{f\} \cup \mathcal{F})\right)\\ &\equiv \operatorname{Holant}\left(\neq_2 \mid \{=_d\} \cup Z^{-1}\mathcal{F}\right).\end{aligned}$$

If $d \geqslant 3$, then with $\neq_2$ on the left and $=_d$ on the right, we can realize all signatures in $\mathcal{EQ}_d$ on the left, as follows. We construct $=_{kd}$ inductively on k. For $k = 1$, we just need to attach $\neq_2$ on every edge of $=_d$, which effectively moves $=_d$ to the left. Suppose we have $=_{kd}$, and we want to construct $=_{(k+1)d}$. We can connect $=_{kd}$ and two copies of $=_d$'s on the left to one $=_d$ on the right, using up all edges of the $=_d$ on the right. Since $d \geqslant 3$, we can make sure that the resulting signature is connected. The order of the connection does not matter. The resulting signature is an EQUALITY, and its arity is $kd + 2d - d = (k+1)d$, which is what we want. Hence, we have that

$$\#\mathrm{CSP}^d(Z^{-1}\mathcal{F}) \leqslant_T \operatorname{Holant}(\{f\} \cup \mathcal{F}).$$

Moreover, clearly the construction above can be done in a planar way, implying that,

$$\text{Pl-}\#\mathrm{CSP}^d(Z^{-1}\mathcal{F}) \leqslant_T \text{Pl-Holant}(\{f\} \cup \mathcal{F}).$$

If $\#\mathrm{CSP}^d(Z^{-1}\mathcal{F})$ is #**P**-hard, then $\operatorname{Holant}(\{f\} \cup \mathcal{F})$ is #**P**-hard as well. It is easy to verify that all tractable cases in Theorem 6.1 also make $\operatorname{Holant}(\{f\} \cup \mathcal{F})$ tractable. Hence, in a sense, classifying $\#\mathrm{CSP}^d(\mathcal{F})$ is necessary to classify Holant problems.

On the other hand, in the planar setting, due to the newly tractable cases that we will see shortly in Section 6.1, proving a dichotomy for Pl-#CSP^d seems overly complicated. In fact, we will take a different route to achieve the dichotomy for Pl-Holant than that in [HL12, CGW13]. Only necessary results related to these new planar tractable cases are derived. We show the dichotomy for Holant afterwards, which does use Theorem 6.1.

6.1 Another Planar Tractable Case

Given a set $\mathcal{F}$ of symmetric signatures, by Theorem 5.41, Pl-Holant($\mathcal{F}$) is #**P**-hard unless every single non-degenerate signature f of arity at least 3 in $\mathcal{F}$ is in $\mathcal{P}_1 \cup \mathcal{M}_2 \cup \mathcal{A}_3 \cup \mathcal{M}_3 \cup \mathcal{M}_4 \cup \mathcal{V}$. We have already proved that the desired full dichotomy holds if $\mathcal{F}$ contains such an f in $\mathcal{P}_1$, $\mathcal{A}_3$, $\mathcal{M}_2 \setminus \mathcal{P}_2$, or $\mathcal{M}_3$ due to Corollary 5.25, Corollary 5.28, Corollary 5.31, or Lemma 5.33, respectively.

The remaining cases are when all non-degenerate signatures of arity at least 3 in $\mathcal{F}$ are contained in $\mathcal{P}_2 \cup \mathcal{M}_4 \cup \mathcal{V}$. In this section, we consider the mixing of $\mathcal{P}_2$ and $\mathcal{M}_4$. For this, we do a holographic transformation by Z. Then the problem becomes Pl-Holant $(\neq_2 \mid =_k, \text{EXACTONE}_d)$ with various arities k and d. Recall that EXACTONE_d denotes the exact one function $[0, 1, 0, \ldots, 0]$ of arity d. These are the signatures for PERFECT MATCHINGS and they are the basic components of *Matchgates*.

A big surprise, against the putative form of a complexity classification for planar counting problems, is that we found out the complexity of Pl-Holant $(\neq_2 \mid =_k, \text{EXACTONE}_d)$ depends on the values of d and k, and the problem is tractable for all large k. These problems *cannot* be captured by a holographic reduction to Kasteleyn's algorithm, or any other known algorithm. Thus for planar problems the paradigm of holographic algorithms using matchgates (i.e., being $\mathcal{M}$-transformable) *is not universal*.

We show that if $k \geqslant 5$, then Pl-Holant $(\neq_2 \mid =_k, \text{EXACTONE}_d)$ is tractable. We first show this for $k \geqslant 6$, and then return to $=_5$. Let $\mathcal{EO} = \{\text{EXACTONE}_d \mid d \geqslant 3\}$.

To prove this, we first observe some possible degeneracy. Let G be the underlying graph of an instance Ω of Pl-Holant $(\neq_2 \mid =_k, \mathcal{EO})$. Any self loop on an EXACTONE_d by a $\neq_2$ makes it $[1, 0]^{\otimes(d-2)}$ with a factor of 2. Pinning signatures like $[1, 0]^{\otimes(d-2)}$ can be applied recursively. Any $[1, 0]$ is first transformed to $[0, 1]$ via $\neq_2$ on LHS and then applied either to $=_k$ producing $[0, 1]^{\otimes(k-1)}$, or to EXACTONE_d (for some d) producing $[1, 0]^{\otimes(d-1)}$. Similarly, any $[0, 1]$ is first transformed to $[1, 0]$ via $\neq_2$ on LHS and then applied either to $=_k$ producing $[1, 0]^{\otimes(k-1)}$, or to EXACTONE_d (for some d) producing EXACTONE_{d-1}. Note that if $d = 3$ then EXACTONE_{d-1} is just $\neq_2$ on RHS, which combined with its adjacent two copies of $\neq_2$ of LHS, is equivalent to a single $\neq_2$ of LHS. Moreover, whenever an EXACTONE_d and another EXACTONE_ℓ are connected by a $\neq_2$, we replace it by a single $\text{EXACTONE}_{d+\ell-2}$, contracting the edge between (and remove the connecting $\neq_2$). On the other hand, consider a connected component made of $=_k$ and $\neq_2$. We call such a component an E_k-block. Notice that each E_k-block has either exactly two or zero support vectors. This depends on whether or not there exists a contradiction, which is formed by an odd cycle of $=_k$'s connected by $\neq_2$'s. We say an E_k-block is trivial if it has no support. It is easy to check the triviality. The two support vectors of a nontrivial E_k-block are complements of each other. We mark dangling edges of a nontrivial E_k-block by "+" or "−" signs. Dangling edges marked by the same sign take the same value on both support vectors while dangling edges marked by different signs take opposite values on both support vectors. Let $n_\pm$ be the number of dangling edges marked $\pm$. Then it is easy to verify by induction that

$$n_+ \equiv n_- \mod k. \tag{6.1}$$

An example of E_6-block is illustrated in Figure 6.1, with 8 + signs and 2 − signs.

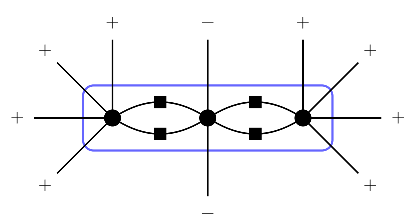


Figure 6.1: Example E_6-block. Circle vertices are assigned $=_6$ and square vertices are assigned $\neq_2$.

After contracting all edges between ExactOne$_d$'s and forming E_k-block's, we obtain a bipartite graph connected between ExactOne$_d$'s and E_k-block's by edges labeled $=_2$.

A key observation is that a planar (bipartite) graph cannot be simple, i.e., it must have parallel edges, if its degrees are large.

Lemma 6.2. *Let* $G = (L \cup R, E)$ *be a planar bipartite graph with parts* L *and* R. *If every vertex in* L *has degree at least* 6 *and every vertex in* R *has degree at least* 3*, then* G *is not simple.*

Proof. Suppose G is simple. Let v, e and f be the total number of vertices, edges, and faces, respectively. Let v_i be the number of vertices of degree i in L, where $i \geqslant 6$, and u_j be the number of vertices of degree j in R, where $j \geqslant 3$. Since G is simple and bipartite, each face has at least 4 edges. Thus,

$$2e \geqslant 4f. \tag{6.2}$$

Furthermore, it is easy to see that

$$v = \sum_{i \geqslant 6} v_i + \sum_{j \geqslant 3} u_j \qquad \text{and} \qquad e = \sum_{i \geqslant 6} i v_i = \sum_{j \geqslant 3} j u_j. \tag{6.3}$$

Then starting from Euler's characteristic equation for planar graphs, we have

$$\begin{aligned} 2 &= v - e + f \\ &\leqslant v - \frac{e}{2} && \text{(By (6.2))} \\ &= \sum_{i \geqslant 6} v_i + \sum_{j \geqslant 3} u_j - \frac{1}{6} \sum_{i \geqslant 6} i v_i - \frac{1}{3} \sum_{j \geqslant 3} j u_j && \text{(By (6.3))} \\ &= \sum_{i \geqslant 6} \frac{6 - i}{6} v_i + \sum_{j \geqslant 3} \frac{3 - j}{3} u_j \leqslant 0, \end{aligned}$$

a contradiction. □

Lemma 6.2 does not give us tractability for the case of $k \geqslant 6$ yet. The reason is that given an instance of Pl-Holant $(\neq_2 \mid =_k, \mathcal{EO})$, after the preprocessing and forming E_k-blocks to make the graph bipartite, it is possible to have E_k-blocks of arity less than 6, in which case Lemma 6.2 does not apply. However, for $k \geqslant 6$ and a nontrivial E_k-block of arity n where $n < 6$, by (6.1) and the fact that $0 \leqslant n_+, n_- \leqslant n < k$, we see that $n_+ = n_-$, and $n = n_+ + n_-$ must be even. Moreover, if $n = 2$, then this means that the E_k-block is just $\neq_2$, in which case we can replace it by a single $\neq_2$ connecting signatures from $\mathcal{EO}$ to produce a new ExactOne signature. The only problematic case is when $n = 4$. There are two possibilities of such E_k-blocks up to a rotation, shown in Figure 6.2a.

(a) Two different arity 4 E_k-blocks. (b) Replace them by parallel $\neq_2$'s.

Figure 6.2: Arity 4 E_k-blocks.

Formally we define a *contraction* process on the connected graph of E_k-block with dangling edges. Recursively, for any non-dangling non-loop edge e, we shrink it to a point, maintaining planarity. The local cyclic orders of incident edges of the two vertices of e are spliced along e to form the cyclic order of the new vertex. For any loop we simply remove it. This contraction process ends in a single point with a cyclic order of the dangling edges.

Figure 6.2a depicts the two possibilities of E_k-blocks of arity 4 up to a rotation. An E_k-block of arity 4 can be viewed as a pair of $\neq_2$ in parallel, but there is a correlation between them, namely their support vectors are paired up in a unique way. If we replace the contracted E_k-block of arity 4 by two parallel edges as indicated in Fig 6.2b, one can revert back to a planar realization in the E_k-block as it connects to the rest of the graph. This can be seen by reversing the contraction process step by step.

To prove the following lemma, we will show how to replace E_k-block of arity 4 by some other signatures while keeping track of the Holant value. We also observe that this tractable set is compatible with binary $\neq_2$ and unary $[1,0]$ or $[0,1]$ signatures.

Lemma 6.3. *For any integer* $k \geqslant 6$*,* Pl-Holant $(\neq_2 \mid =_k, \mathcal{EO}, \neq_2, [1,0], [0,1])$ *is tractable.*

Proof. Let Ω be an instance of Pl-Holant $(\neq_2 \mid =_k, \mathcal{EO}, \neq_2, [1,0], [0,1])$. Without loss of generality, we assume that Ω is connected. Any occurrence of $\neq_2$ of the right hand side can be removed as follows: It is connected to two adjacent copies of $\neq_2$ of the left hand side. We replace these 3 copies of $\neq_2$ by a single $\neq_2$ from the left hand side.

The given signatures have no weight, however the proof below can be adapted to the weighted case. For the unweighted case, we only need to count the number of satisfying assignments. We call an edge pinned if it has the same value in all satisfying assignments, *if there is any.* Clearly any edge incident to a vertex assigned $[1,0]$ or $[0,1]$ is pinned.

When an edge is pinned to a known value, we can get a smaller instance of the problem Pl-Holant$(\neq_2 \mid =_k, \mathcal{EO}, \neq_2, [1,0], [0,1])$ without changing the number of satisfying assignments. In our algorithm we may also find a contradiction and simply return 0. If e is a pinned edge, then it is adjacent to another edge e' via $\neq_2$ on the left hand side, and both e and e' are pinned. We remove e, e', and $\neq_2$, and perform the following on e (and on e' as well). If the other endpoint of e is $u = [1,0]$ or $[0,1]$ we either remove that u if the pinned value on e is consistent with u, or return 0 otherwise. If the other endpoint of e is $=_k$, then all edges of this $=_k$ are pinned to the same value which we can recursively apply. If the other endpoint of e is ExactOne$_d \in \mathcal{EO}$, then we replace this signature by ExactOne$_{d-1}$ when the pinned value is 0; or if the pinned value is 1 then the remaining $d-1$ edges of this ExactOne$_d$ are pinned to 0 which we recursively apply. Notice that we may create an ExactOne$_2$ (i.e. $\neq_2$) on the right hand side when we pin 0 on ExactOne$_3$. Such $\neq_2$'s are replaced as described at the beginning. It is easy to see that all these procedures do not change the number of satisfying assignments, and work in polynomial time.

We claim that there always exists an edge in Ω that is pinned, unless Ω does not contain $=_k$, or does not contain ExactOne$_d$ functions (for some $d \geqslant 3$), or there is a contradiction. Furthermore if there are $=_k$ or ExactOne$_d$ functions (for some $d \geqslant 3$), in polynomial time we can find a pinned edge with a known value, or return that there is a contradiction. (If there is a contradiction in Ω, we may still return a purported pinned edge with a known value, which we can apply and simplify Ω. The contradiction will eventually be found.) If Ω does not contain $=_k$, or does not contain ExactOne$_d$ functions (for some $d \geqslant 3$), then the problem is tractable, since Ω is an instance of $\mathcal{M}$, or an instance of $\mathcal{P}$. The lemma follows from the claim, for we either recurse on a smaller instance or have a tractable instance.

Suppose Ω is an instance where at least one $=_k$ and at least one ExactOne$_d \in \mathcal{EO}$ appear. We assume no $\neq_2$ appears on the right hand side. If any $[1, 0]$ or $[1, 0]$ appear, then we have found a pinned edge with a known value. Hence we may assume neither $[1, 0]$ nor $[1, 0]$ appears in Ω.

If a signature ExactOne$_d \in \mathcal{EO}$ is connected to itself by a self-loop through a $\neq_2$, then there are two choices for the assignment on this pair of edges through the $\neq_2$, but the remaining $d-2 \geqslant 1$ edges are pinned to 0. We can keep track of the factor 2 and have found a pinned edge with a known value. Thus we may assume there are no self-loops via $\neq_2$ on ExactOne signatures.

Next we consider the case that two separate signatures ExactOne$_d$ and ExactOne$_\ell$ from $\mathcal{EO}$ are connected by some number of $\neq_2$'s. Depending on the number of connecting edges, there are three cases:

1. The connection is by a single $\neq_2$. We contract the connecting edge, maintaining planarity, and replace these three signatures by an $\textsc{ExactOne}_{d+\ell-2}$ to get a new instance Ω'. If an edge is pinned in Ω' then it is also pinned in Ω to the same value. We continue with Ω'.

2. The connection is by two $\neq_2$'s. There are two choices for the assignment on these two pairs of edges through $\neq_2$, but the remaining $d+\ell-4 \geqslant 2$ edges are pinned to 0.

3. The connection is by at least three $\neq_2$'s. The three $\neq_2$'s cannot be all satisfied, so there is no satisfying assignment, a contradiction. We return the value 0.

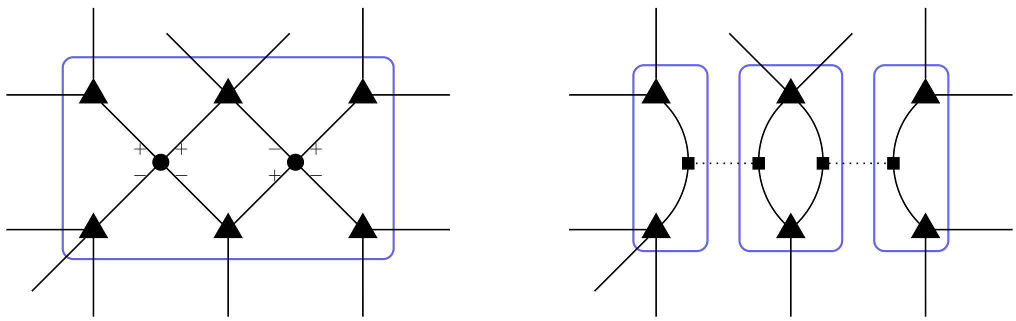
Hence, we may assume there is no connection via any number of $\neq_2$'s among ExactOne signatures.

Define an E_k-block as a connected component composed of $=_k$ and $\neq_2$. All external connecting edges of each E_k-block are marked with $+$ or $-$ and this can be found by testing bipartiteness of a E_k-block where we treat $\neq_2$'s as edges. If any E_k-block is not bipartite, we return 0. We contract all E_k-blocks and maintain planarity. For each E_k-block we contract two vertices that are connected by an edge, one edge at a time, and remove self loops in this contraction process. If a trivial E_k-block appears, then there is no satisfying assignment, and we return 0. Thus we may assume all E_k-blocks are nontrivial. If there is a nontrivial E_k-block of arity 2, as discussed earlier, its signature is $\neq_2$. We replace it with an edge labeled by $\neq_2$ to form an instance Ω', maintaining planarity, such that any pinned edge in Ω' corresponds to a pinned edge in Ω. This new edge is between ExactOne signatures and can be dealt with as described earlier. So we may assume the arity of any E_k-block is at least 4. Since $k \geqslant 6$, the only possible E_k-blocks of arity 4 are those in Figure 6.2a up to a rotation. Since there is at least one $\textsc{ExactOne}_d$ signature with $d \geqslant 3$, forming E_k-blocks does not consume all of Ω.

After these steps we may consider Ω a bipartite graph, with one side consisting of E_k-blocks and the other side ExactOne signatures. Edges are now labeled by $=_2$.

Suppose there are parallel edges between an E_k-block and an $\textsc{ExactOne}_d$ signature. We show that this always leads to some pinned edges. If two parallel edges are marked by the same sign in the E_k-block, then they must be pinned to 0. If they are marked by different signs, then the remaining $d-2 \geqslant 1$ edges of the $\textsc{ExactOne}_d$ signature must be pinned to 0.

Therefore, we may assume that there are no parallel edges between any E_k-block and any ExactOne signature.



(a) An EO-Eq-4-block. Triangles are assigned ExactOne signatures and circles are E_k-blocks of arity 4.

(b) Break the EO-Eq-4-block into three components. Squares are assigned $\neq_2$. The component in the middle contains a cycle, and hence is degenerate. The other two are equivalent to ExactOne signatures.

Figure 6.3: EO-Eq-4-blocks

The next thing we do is to consider E_k-blocks of arity 4 with ExactOne signatures together. Call a connected component consisting of E_k-blocks of arity 4 and ExactOne an EO-Eq-4-block. Figure 6.3a illustrates an example. Notice that the two possibilities of E_k-blocks of arity 4 can be viewed as two parallel $\neq_2$'s but with some correlation between them. This is illustrated in Figure 6.2b. Note that the two dotted lines in Figure 6.2b represent different correlations.

At this point we would like to replace every arity 4 E_k-block by two parallel $\neq_2$'s. However this replacement destroys the equivalence of the Holant values, before and after.

The surprising move of this proof is that we shall do so anyway!

Suppose we ignore the correlation for the time being and replace every arity 4 E_k-block by two parallel $\neq_2$'s as in Figure 6.2b. This replacement produces a *planar* signature grid Ω_1. Every edge in Ω_1 corresponds to a unique edge in Ω. The set of satisfying assignments of Ω_1 is a superset of that of Ω. Moreover, if there is an edge pinned in Ω_1 to a known value, the corresponding edge is also pinned in Ω to the same value. Once we find that in Ω_1 we revert back to work in Ω and apply the pinning to the pinned edge.

All that remains to be shown is that pinning always happens in Ω_1. Each EO-Eq-4-block splits into some number of connected components in Ω_1. If any component contains a cycle (which must alternate between $\neq_2$, which are the newly created ones from the E_k-blocks of arity

4, and $\textsc{ExactOne}_d$ signatures for $d \geqslant 3$), then any edges not in the cycle but incident to some vertex in the cycle is pinned to 0. Moreover such edges must exist, for $\textsc{ExactOne}_d$ signatures in the cycle are of arity at least 3. Note that the cycle has even length, and there are exactly two satisfying assignments, which assign exactly one 0 and one 1 to the two cycle edges incident to each $\textsc{ExactOne}_d$ signature. This produces pinned edges.

Hence we may assume there are no cycles in these components, and every such component forms a tree, whose vertices are $\textsc{ExactOne}$ functions and edges are $\neq_2$'s. Suppose there are $n \geqslant 2$ vertices in such a tree. As discussed in item 1 above, the whole tree is an $\textsc{ExactOne}_t$ function for some arity t. Since each vertex in the tree has degree at least 3, $t \geqslant 3n-2(n-1) = n+2 \geqslant 4$. We replace these components by $\textsc{ExactOne}_t$'s.

Thus, each connected component in the graph underlying Ω_1 is a planar bipartite graph with E_k-blocks of arity at least 6 on one side and $\textsc{ExactOne}_d$ signatures of arity at least 3 on the other. By Lemma 6.2, no component is simple, which means that there are parallel edges between some E_k-block and some $\textsc{ExactOne}_d$ signature. As discussed earlier, there must exist some pinned edge, and we can find a pinned edge with a known value in polynomial time. This finishes the proof. □

Unlike the situation in Lemma 6.2, a planar $(5, 3)$-regular bipartite graph *can* be simple. However, we show that such graphs must have a special induced subgraph. We call this structure a "wheel", which is pictured in Figure 6.4. There is a vertex v of degree 5 in the middle, and all faces adjacent to this vertex are 4-gons (i.e. quadrilaterals). Moreover, at least four neighbors of v have degree 3. Depending on the degree of the fifth neighbor (whether it is 3 or not), we have two types of wheel, pictured in Figure 6.4a and Figure 6.4b.

Lemma 6.4. *Let* $G = (L \cup R, E)$ *be a planar bipartite graph with parts* L *and* R*. Every vertex in* L *has degree at least* 5 *and every vertex in* R *has degree at least* 3*. If* G *is simple, then there exists one of the two wheel structures in Figure 6.4 in* G*.*

Proof. Let $V = L \cup R$ be the set of vertices and let F be the set of faces. We assign a *score* s_v to each vertex $v \in V$. We will define s_v so that $\sum_{v \in V} s_v = |V| - |E| + |F| = 2 > 0$. The base score is $+1$ for each vertex, which accounts for $|V|$. For each k-gon face, we assign $\frac{1}{k}$ to each of its

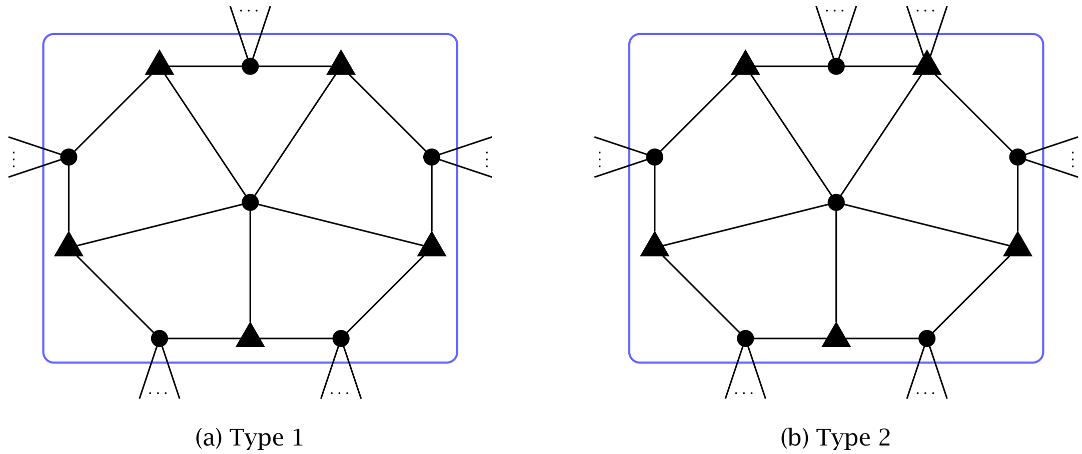


Figure 6.4: Two types of wheels. Each circle is an E_5-block and triangle an ExactOne signature.

vertex. This accounts for $|F|$. As G is a bipartite and a simple graph, $k \geqslant 4$ and a score from a face to a vertex is at most $\frac{1}{4}$.

For $-|E|$, we separate two cases. For any edge if one of the two endpoints has degree 3, we give the degree 3 vertex a score of $-\frac{7}{12}$, and the other one $-\frac{5}{12}$. This is well defined because all degree 3 vertices are in R. If the endpoints are not of degree 3, we give each endpoint $-\frac{1}{2}$. This accounts for $-|E|$.

Now we claim that $s_v \leqslant 0$ unless $v \in L$ and has degree 5. Suppose $v \in L$ and has degree $d \geqslant 6$, then

$$s_v \leqslant 1 + \frac{d}{4} - \frac{5}{12}d = 1 - \frac{d}{6} \leqslant 0.$$

Now suppose $v \in R$ and v has degree $d \geqslant 4$. Then every edge adjacent to v gives a score $-\frac{1}{2}$. Hence,

$$s_v \leqslant 1 + \frac{d}{4} - \frac{1}{2}d = 1 - \frac{d}{4} \leqslant 0.$$

The remaining case is that $v \in R$ and v has degree 3. Then,

$$s_v \leqslant 1 + \frac{d}{4} - \frac{7}{12}d = 1 - \frac{d}{3} \leqslant 0.$$

The claim is proved.

Since the total score is positive, there must exist $v \in L$, v has degree 5 and $s_v > 0$. We then claim that there must exist such a v so that all adjacent faces are 4-gons. Suppose otherwise. Then any such v is adjacent to at least one k-gon with $k \geqslant 6$. In this case,

$$s_v \leqslant 1 + \frac{1}{4} \cdot 4 + \frac{1}{6} - \frac{5}{12} \cdot 5 = \frac{1}{12}.$$

Moreover, if v is adjacent to more than one k-gon with $k \geqslant 6$, Then

$$s_v \leqslant 1 + \frac{1}{4} \cdot 3 + \frac{1}{6} \cdot 2 - \frac{5}{12} \cdot 5 = 0,$$

contrary to the assumption that $s_v > 0$. Hence v is adjacent to exactly one k-gon with $k \geqslant 6$. Call this face F_v.

In F_v, v has two neighbors in R. We match each vertex v that has a positive score to the vertex on F_v that is the next one in clockwise order from v. By the bipartiteness, every such v is matched to a vertex in R. We do this matching in all faces containing at least one positively scored vertex. It is possible that more than one such v are matched to the same $u \in R$. Suppose a vertex $u \in R$ is matched to from ℓ different such vertices of positive score. This means that u is adjacent to at least ℓ many k-gons with $k \geqslant 6$. Then, if u has degree 3 then u has score

$$s_u \leqslant 1 + \frac{1}{4} \cdot (3 - \ell) + \frac{1}{6} \cdot \ell - \frac{7}{12} \cdot 3 = -\frac{\ell}{12}.$$

If u has degree $d \geqslant 4$ then u has score

$$s_u \leqslant 1 + \frac{1}{4} \cdot (d - \ell) + \frac{1}{6} \cdot \ell - \frac{1}{2} \cdot d \leqslant -\frac{\ell}{12}.$$

Hence in any case, we have $s_u \leqslant -\frac{\ell}{12}$. It implies that the total score of u and all positively scored vertices matched to u is at most 0. However each positively scored vertex is matched to a vertex in R. Hence the total score cannot be positive. This is a contradiction.

Therefore there exists $v \in L$ such that $s_v > 0$, and has degree 5, and all adjacent faces are

4-gons. We further note that at most one neighbor of v can have degree $\geqslant 4$, for otherwise,

$$s_v \leqslant 1 + \frac{5}{4} - \frac{1}{2} \cdot 2 - \frac{5}{12} \cdot 3 = 0.$$

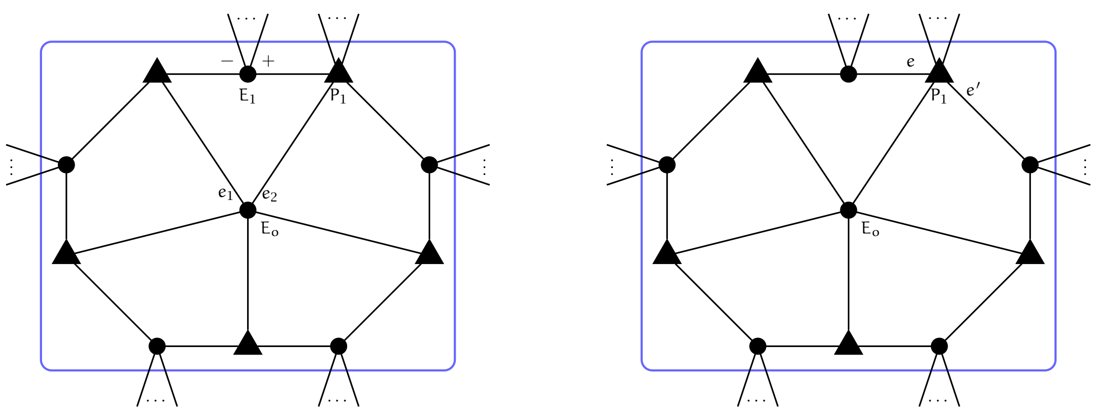
If all neighbors of v have degree 3, that is a wheel of type 1 as in Figure 6.4a. If one neighbor of v has degree $\geqslant 4$, that is a wheel of type 2 as in Figure 6.4b. □

As we shall see, either structure in Figure 6.4 leads to pinned edges.

Lemma 6.5. Pl-Holant $(\neq_2 \mid =_5, \mathcal{EO}, \neq_2, [1,0], [0,1])$ *is tractable.*

Proof. We proceed as in Lemma 6.3 up until the point of getting Ω_1. Note that due to (6.1) the only nontrivial E_5-blocks of arity $\leqslant 4$ are $\neq_2$ and those in Figure 6.2a. Moreover, each connected component of Ω_1 is planar and bipartite with vertices on one side having degree at least 5 and those on the other at least 3. We only need to show that there are edges pinned in Ω_1.

Unlike in Lemma 6.3, these components do not satisfy the condition of Lemma 6.2 but that of Lemma 6.4. If any such component is not simple, then there are pinned edges similar to Lemma 6.3. Otherwise by Lemma 6.4, the wheel structure in Figure 6.4 appears. All we need to show is that wheel structures of either type contain pinned edges.



(a) Different signs of an E_5-block along the cycle lead to pinning

(b) Edges e and e' are pinned in wheels of type 2

Figure 6.5: Degeneracies in the wheel structure.

First we claim that if a wheel of either type has a E_5-block, call it E_1, on the outer cycle which has different signs on the two edges incident to it along the cycle, then the middle $=_5$, denoted by E_o, is pinned. This is pictured in Figure 6.5a. It does not matter whether the wheel is type 1 or 2, or the position of E_1 relative to the special triangle P_1 in type 2. Because E_o is an equality, both e_1 and e_2, the two edges incident to E_o that are connected to the two ExactOne signatures flanking E_1, must take the same value. If both e_1 and e_2 are assigned 1, then the two incoming wires of E_1 along the cycle have to be both assigned 0, whereas they are marked by different signs. This is a contradiction. Hence both e_1 and e_2 are pinned to 0 as well as all edges of E_o.

We may therefore assume that each E_5-block has same signs along the outer cycle, either $++$ or $--$. If the wheel is of type 1, then there is no valid assignment such that E_o is assigned 0 because the cycle has odd length. In fact if E_o is assigned 0, then we can remove E_o and its incident edges, and effectively the five ExactOne signatures are now $\neq_2$'s forming a 5-cycle linked by binary equalities. Hence E_o and all its edges are pinned to 1.

Otherwise the wheel is of type 2, and each E_5-block has signs $++$ or $--$ along the outer cycle. We denote by P_1 the special ExactOne_d function that has arity $d > 3$. We claim that the two edges e and e' incident to P_1 along the cycle are both pinned to 0. This is illustrated in Figure 6.5b. As P_1 is ExactOne_d, at most one of e and e' is 1. If one of e and e' is 1, the other is 0, and as P_1 is an ExactOne_d function its edge to E_o is also 0, and thus all edges incident to E_o are 0. As all five neighbors of E_o are ExactOne functions, the four ExactOne_3 functions effectively become $(\neq_2)$ functions along the wheel, and we can remove E_o and its incident edges. This becomes the same situation as in the previous case of type 1, where effectively a cycle of five binary equalities are linked by five binary disequalities, which has no valid assignment. It implies that both e and e' are pinned to 0. This finishes the proof. □

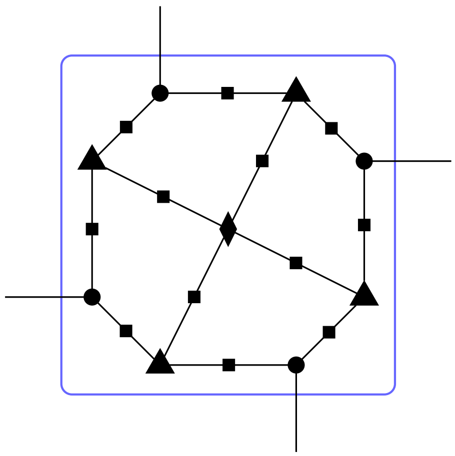
6.2 Complementing Hardness Results

On the other hand, we show that Pl-Holant $(\neq_2 \mid =_k, \text{ExactOne}_d)$ is #**P**-hard when $k = 3, 4$. Note that when $k = 2$ it is tractable as every signature is a matchgate.

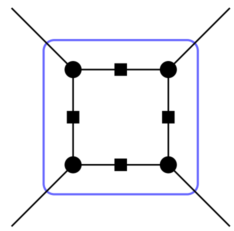
Lemma 6.6. Pl-Holant $(\neq_2 \mid =_3, [0, 1, 0, 0])$ *is #**P**-hard.*

Proof. By connecting two copies of $[0, 1, 0, 0]$ together via $\neq_2$, we have $[0, 1, 0, 0, 0]$ on the right.

Consider the gadget in Figure 6.6a. We assign $=_3$ to the triangle vertices, $[0, 1, 0, 0]$ to the circle vertices, $\neq_2$ to the square vertices, and $[0, 1, 0, 0, 0]$ on the diamond vertex in the middle. Let f be the signature of this gadget.



(a) Cycle-like gadget used twice



(b) Gadget to realize $\overline{g}$

Figure 6.6: Two gadgets used in the proof of Lemma 6.6.

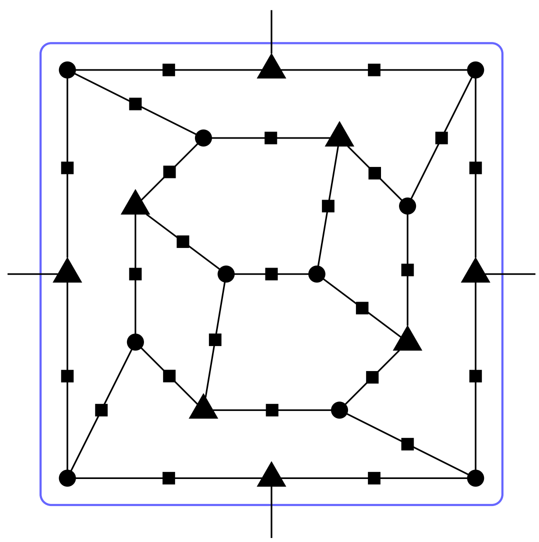


Figure 6.7: The whole gadget to realize $[0, 0, 0, 1, 0]$.

We claim that the support of f is $\{0011, 0110, 1100, 1001\}$. To see this, notice that $[0, 1, 0, 0, 0]$ in the middle must match exactly one of the half edges, which forces the corresponding equality signature to take the value 0 and all other equality signatures to take value 1. The two $[0, 1, 0, 0]$'s adjacent to the equality assigned 0 must have 0 going out, and the other two $[0, 1, 0, 0]$'s have 1 going out.

Now we consider the gadget in Figure 6.6a again. This time we place $[0, 1, 0, 0]$ on each triangle, $=_3$ on each circle, f on the middle diamond, and again $\neq_2$ on each square. Now notice that each support of f makes two $[0, 1, 0, 0]$'s that are cyclically adjacent on the outer cycle to become $[0, 1, 0]$ and the other two $[1, 0, 0]$. It is easy to see that the support of the resulting signature is $\{0111, 1011, 1101, 1110\}$. Therefore it is the reversed ExactOne$_4$ signature $[0, 0, 0, 1, 0]$ (namely AllButOne$_4$). The whole gadget is illustrated in Figure 6.7, where each circle is assigned $[0, 1, 0, 0]$, triangle $=_3$, and square $\neq_2$.

Finally, we build the gadget in Figure 6.6b. We place $=_3$ on each circle and $\neq_2$ on each square. It is easy to see that there are only two support vectors of the resulting signature, which are 0101 and 1010. Recall (5.4), the definition of the partial crossover $\overline{g}$. This gadget realizes exactly $\overline{g}$.

By Lemma 5.39, Pl-Holant $(\neq_2 \mid [0, 1, 0, 0, 0], [0, 0, 0, 1, 0], \overline{g})$ is #**P**-hard. We have constructed $[0, 1, 0, 0, 0]$, $[0, 0, 0, 1, 0]$, and $\overline{g}$ on the right side. Therefore Pl-Holant $(\neq_2 \mid =_3, [0, 1, 0, 0])$ is #**P**-hard. □

For $k = 4$, we need the following lemma.

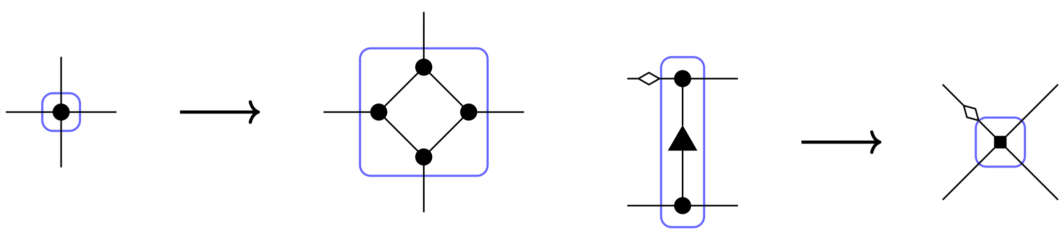
Lemma 6.7. *Let g be the arity 4 signature whose matrix is*

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

Then Pl-Holant(g) *is #**P**-hard.*

Proof. Let $h = [2, 1, 1]$. We show that Pl-#CSP$(h) \leqslant_T$ Pl-Holant(g) in two steps. In each step, we begin with a signature grid and end with a new signature grid such that the Holants of both signature grids are the same. Then we are done by Theorem 4.1, or more explicitly, since Pl-#CSP$(h) \equiv$ Pl-Holant $(\mathcal{EQ} \mid h)$ by (1.1), we are done by Theorem 4.2.

For step one, let $G = (U, V, E)$ be an instance of Pl-Holant $(\mathcal{EQ} \mid h)$. Fix an embedding of G in the plane. This defines a cyclic ordering of the edges incident to each vertex. Consider a vertex $u \in U$ of degree k. It is assigned the signature $=_k$. We decompose u into k vertices. Then we connect the k edges originally incident to u to these k new vertices so that each vertex is incident to exactly one edge. We also connect these k new vertices in a cycle according to the cyclic ordering induced on them by their incident edges. Each of these vertices has degree 3, and we assign them $=_3$. Clearly the Holant value is unchanged. This completes step one. An example of this step applied to a vertex of degree 4 is given in Figure 6.8a. The resulting graph has the following properties: (1) it is planar; (2) every vertex is either degree 2 (in V and assigned h) or degree 3 (newly created and assigned $=_3$); (3) each degree 2 vertex is connected to two degree 3 vertices; and (4) each degree 3 vertex is connected to one degree 2 vertex and two other degree 3 vertices.



(a) Step one: Degree 4 vertex example (b) Step two: Contract edges

Figure 6.8: A reduction from Pl-Holant $(\mathcal{EQ} \mid h)$ to Pl-Holant(g) for any binary signature h and a quaternary signature g that depends on h. The circle vertices are assigned $=_4$ or $=_3$ respectively, the triangle vertex is assigned h, and the square vertex is assigned the signature of the gadget to its left.

Now step two. For every $v \in V$, v has degree 2. We contract the two edges incident to v, or equivalently, we replace the two circle vertices and one triangle vertex boxed in Figure 6.8b with a single (square) vertex of degree 4. The resulting graph $G' = (V', E')$ is planar and 4-regular.

Next we determine what is the signature on $v' \in V'$ after this contraction. Clearly the two inputs to each original circle have to be the same. Therefore its support is 0000, 0110, 1001, 1111, listed starting from the diamond and going counterclockwise. Moreover, due to the triangle assigned h in the middle, the weight on 0000 is 2, and every other weight is 1. Hence it is exactly the signature g, with the diamond in Figure 6.8b marking the first input bit. This finishes the proof. $\square$

Remark From the planar embedding of the graph G, treating h vertices as edges, the resulting graph G' is known as the medial graph of G, which we have met before. See Figure 3.3 for an example. The (constructive) definition is usually phrased in the following way. The medial graph G_m of plane graph G has a vertex on each edge of G and two vertices in G_m are joined by an edge for each face of G in which their corresponding edges occur consecutively. However, our construction described in the proof above clearly extends to nonplanar graphs as well.

Lemma 6.8. Pl-Holant $(\neq_2 \mid =_4, [0,1,0,0])$ *is* #**P**-*hard.*

Proof. Consider the gadget in Figure 6.9. We assign binary disequality $\neq_2$ to the square vertices, $=_4$ to the circle vertices, and $[0,1,0,0]$ to the triangle vertices. We show that the support of the resulting signature is the set $\{00110011, 11001100, 11111111\}$, where each vector is the assignment ordered counterclockwise starting from the diamond point.

We call the equality signature $=_4$ in the middle the origin. There are two possible assignments at the origin. If it is assigned 0, then every adjacent perfect matching signature $[0,1,0,0]$ is matched to the half edge towards the origin, and every equality $=_4$ is forced to be 1. This gives the support vector 11111111.

The other possibility is that the origin is 1. In this case, we can remove the origin leaving the outer cycle, with every $[0,1,0,0]$ becoming $[0,1,0]$. This is effectively a cycle of four equalities connected by $\neq_2$. It is easy to see that there are only two support vectors, which are exactly 00110011 and 11001100.

Every pair of half edges at each corner always take the same value. We further connect each pair of these edges to different copy of $=_4$ via two copies of $\neq_2$. This results in a gadget with signature f whose support is the complement of the original support, that is, $\{11001100, 00110011, 00000000\}$.

Now consider the gadget in Figure 6.10a. We assign $\neq_2$ to the square vertices, $=_4$ to the circle vertices, $[0,1,0,0]$ to the triangle vertices, and f to the pentagon vertex. Notice that each pair of edges coming out of the pentagon vertex are from the same corner of the gadget in Figure 6.9 used to realize f. We now study the signature of this gadget.

Notice that if a $=_4$ on the outer cycle is assigned 0, then the two adjacent perfect matchings must match half edges toward that $=_4$, and their outgoing edges must be 0. Furthermore, the

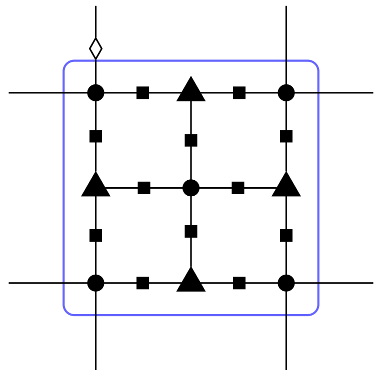
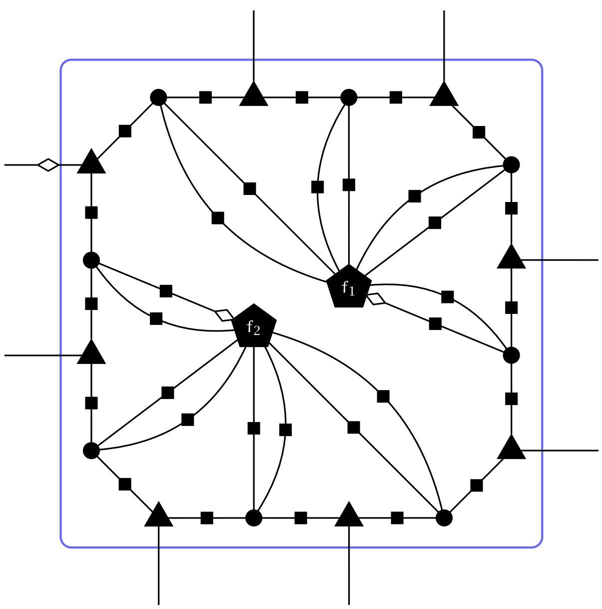


Figure 6.9: Grid-like gadget used in the proof of Lemma 6.8, whose support vectors are 00110011, 11001100, and 11111111. Each square is assigned a binary disequality $\neq_2$, circle $=_4$, and triangle $[0, 1, 0, 0]$.



(a) Gadget with signature g. Each square is assigned a binary disequality $\neq_2$, circle $=_4$, triangle $[0, 1, 0, 0]$, and pentagon f.

f_1	f_2	g
00000000	00000000	11111111
00110011	00000000	01111000
11001100	00000000	11110000
00000000	00110011	10000111
00110011	00110011	00000000
11001100	00110011	-
00000000	11001100	00001111
00110011	11001100	-
11001100	11001100	00000000

(b) Support of g. Each vector is an assignment ordered counterclockwise from the diamond.

Figure 6.10: Another gadget used in the proof of Lemma 6.8 and a Table listing the support of its signature.

two $=_4$ one more step away must be 1. A further observation is that any pair of consecutive $=_4$'s cannot be both 0, and if a pair of consecutive $=_4$'s are both 1, then the $[0, 1, 0, 0]$ in the middle must have a 1 going out. In Figure 6.10a, we call the pentagon connecting to four equalities $=_4$ on the upper right f_1 and the other one f_2. Let g be the signature of resulting gadget. We further order the external wires of f_1, f_2, and g counterclockwise, each starting from edge marked with a diamond. With this notation and these observations, we get Table 6.10b listing the support of g. The support of g is $\{11111111, 01111000, 11110000, 10000111, 00000000, 00001111, 00000000\}$, and 00000000 has multiplicity 2.

Next we use a *domain pairing* argument. First we move $=_4$ to the left hand side, by contracting four $\neq_2$ into it. We apply the domain pairing on the problem Pl-Holant $(=_4 | \; g)$. Specifically, we use $=_4$ as $=_2$, by pairing each pair of edges together. We also pair adjacent two outputs of g clockwise, starting from the diamond point. Each pair of output wires of g are connected to a pair of wires from $=_4$ on the left hand side. Note that $=_4$ enforces that each pair of edges always takes the same value. We re-interpret 00 or 11 as 0 or 1 in the Boolean domain. In this way, we can treat g as an arity 4 signature g' in the Boolean domain. So the reduction is

$$\text{Pl-Holant}\left(=_2 \mid g'\right) \leqslant_T \text{Pl-Holant}\left(=_4 \mid g\right).$$

We get the expression of g' next. The two support bit strings 01111000 and 10000111 of g are eliminated as they do not agree on adjacent paired outputs. So in the paired (Boolean) domain, the support of g' becomes $\{1111, 1100, 0011, 0000\}$ where 0000 has multiplicity 2. We further rotate g' as a Boolean domain signature such that the support is $\{1111, 0110, 1001, 0000\}$ (cf. Figure 3.2). Now it is easy to see that the matrix of g', an arity 4 signature in the Boolean domain, is

$$M_{g'} = \begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

By Lemma 6.7 Pl-Holant(g') is #**P**-hard. Hence Pl-Holant $(\neq_2 \mid =_4, [0, 1, 0, 0])$ is #**P**-hard. □

To extend Lemma 6.6 and Lemma 6.8 to general $\textsc{ExactOne}_d$ functions, we show that we can always realize constant functions $[1,0]$ and $[0,1]$ in this setting.

Lemma 6.9. *For any integer* $k \geqslant 3$ *and* $d \geqslant 3$ *and any signature set* $\mathcal{F}$*,*

$$\text{Pl-Holant}\,(\neq_2 \mid =_k, \textsc{ExactOne}_d, [0,1], [1,0], \mathcal{F}) \leqslant_T \text{Pl-Holant}\,(\neq_2 \mid =_k, \textsc{ExactOne}_d, \mathcal{F})\,.$$

Proof. Given an instance Ω of Pl-Holant $(\neq_2 \mid =_k, \textsc{ExactOne}_d, [0,1], [1,0], \mathcal{F})$ with an underlying planar graph G, if there is any $[1,0]$ on the right hand side, then it can be combined with $\neq_2$ as a $[0,1]$ on the left hand side, and then contracted into whatever function it is attached to. If it is connected to $[1,0]$ or $[0,1]$, we either know the Holant is 0 or remove the two vertices. If it is connected to $\textsc{ExactOne}_d$, then the contraction gives us $d-1$ many $[1,0]$ pinnings. Similarly, if it is connected to $=_k$, the whole function decomposes into $k-1$ many $[0,1]$'s. These additional pinnings by $[1,0]$'s or $[0,1]$'s can be recursively applied.

By a similar analysis, it is easy to show that the only nontrivial occurrences of $[0,1]$'s are those attached to $\textsc{ExactOne}_d$ via $\neq_2$. We may therefore assume there is no $[1,0]$ in Ω, and the only appearances of $[1,0]$'s are those applied to $\textsc{ExactOne}_d$ via $\neq_2$.

We can construct $=_{\ell k}$ for any integer $\ell \geqslant 1$, by $\neq_2$ on the left and $=_k$ on the right. In fact if we connect two copies of $=_k$ via $\neq_2$ we get a signature of arity $2k-2$ with $k-1$ consecutive external wires labeled $+$ and the others labeled $-$. As $k \geqslant 3$, we can take 2 wires of the $k-1$ wires labeled $-$ and attach to two copies of $=_k$ via two $\neq_2$. This creates a signature of arity $3(k-1)+(k-3)$ with $3(k-1)$ consecutive wires labeled $+$ and the other $k-3$ wires labeled $-$. Finally connect $k-3$ pairs of adjacent $+/-$ labeled wires by $\neq_2$ recursively. This creates a planar gadget with an equality signature of arity $3(k-1)-(k-3) = 2k$. This can be extended to any $=_{\ell k}$ by applying the same process on any consecutive k wires.

Next we construct $[0,1]^{\otimes r}$ for some integer $r \geqslant 1$. We get $[1,0]^{\otimes d-2}$ by a self-loop of $\textsc{ExactOne}_d$ via $\neq_2$, ignoring the factor 2. We pick an integer ℓ large enough so that $d-2 < \ell k$. Then we connect $[1,0]^{\otimes d-2}$ to $=_{\ell k}$ via $\neq_2$ to get $[0,1]^{\otimes(\ell k-d+2)}$. This is what we claim with $r = \ell k - d + 2$.

One more construction we will use is $\textsc{ExactOne}_{2+\ell(d-2)}$ for any integer $\ell \geqslant 1$. This is realizable by connecting ℓ many copies of $\textsc{ExactOne}_d$ sequentially via $\neq_2$.

Consider the dual graph G^* of G. Take a spanning tree T of G^*, with the external face as the root. In each face F, let c_F be the number of $[0,1]$'s in the face. We start from the leaves to recursively move all the pinnings of $[0,1]$ to the external face. Suppose we are working on the face F as a leaf of T. If $c_F = 0$ then we just remove the leaf from T and recurse on another leaf. Otherwise we remove all $[0,1]$'s in F. Let s be the smallest integer such that $sr \geqslant c_F$. We replace the $\neq_2$ edge bordering between F and its parent F' by a sequence of three signatures: $\neq_2$, $\textsc{ExactOne}_{2+\ell(d-2)}$ and $\neq_2$, where ℓ is a sufficiently large integer such that $\ell(d-2) \geqslant sr - c_F$. From $\textsc{ExactOne}_{2+\ell(d-2)}$ there are two edges connected to the two adjacent copies of $\neq_2$. Of the other $\ell(d-2)$ edges we will put $sr - c_F$ many dangling edges in F, and the remaining $\ell(d-2) - (sr - c_F)$ dangling edges in F'. Hence there are sr dangling edges in F, including those c_F many that were connected to $[0,1]$'s before we removed the $[0,1]$'s. We put s copies of $[0,1]^{\otimes r}$ inside the face F to pin all of them in a planar way. We add $\ell(d-2) - (sr - c_F)$ to $c_{F'}$. Remove the leaf F from T, and recurse.

After the process, all $[0,1]$'s are in the external face of G. Suppose the number is p. We put r disjoint copies of G together to form a planar signature grid. Apply a total of pr many $[0,1]$'s by p copies of $[0,1]^{\otimes r}$ in a planar way. This is now an instance of Pl-Holant $(\neq_2 \mid =_k, \textsc{ExactOne}_d, \mathcal{F})$ and the Holant value is the rth power of that of Ω. Since the Holant value of Ω is a nonnegative integer, we can take the rth root and finish the reduction. □

Remark Note that the spanning tree argument in the proof above is similar to the alternative algorithm to find a planar pairing mentioned after Lemma 3.37.

Once we have constant functions $[0,1]$ and $[1,0]$, it is easy to construct $\textsc{ExactOne}_3$ from $\textsc{ExactOne}_d$. Therefore combining Lemma 6.9 with Lemma 6.6 and Lemma 6.8 we get the following corollary.

Corollary 6.10. *If* $d \geqslant 3$ *and* $k \in \{3,4\}$, *then* Pl-Holant $(\neq_2 \mid =_k, \textsc{ExactOne}_d)$ *is #***P***-hard.*

6.3 Mixing $\mathcal{M}_4$ and $\mathcal{P}_2$ with Other Signatures

Now we prove some lemmas relating to $\mathcal{M}_4$ and $\mathcal{P}_2$ that are used in the proof of the full dichotomy.

Recall that $\textsc{AllButOne}_d$ is the signature $[0, \ldots, 0, 1, 0]$ of arity d, which is the reverse of $\textsc{ExactOne}_d$. After a Z transformation, $\mathcal{M}_4$ contains both $\textsc{AllButOne}_d$ and $\textsc{ExactOne}_d$. If both appear, then with any $=_k$ the problem is hard.

Lemma 6.11. *If integers* $d_1, d_2, k \geqslant 3$*, then* Pl-Holant$\left(\neq_2 \mid =_k, \textsc{ExactOne}_{d_1}, \textsc{AllButOne}_{d_2}\right)$ *is* #**P**-*hard.*

Proof. We apply Lemma 6.9 to create constant functions $[1, 0]$ and $[0, 1]$ first. Then we construct $\textsc{ExactOne}_4$ and $\textsc{AllButOne}_4$. With both $[1, 0]$ and $[0, 1]$ in hand, we may reduce d_1 or d_2 to 4 if $d_1 > 4$ or $d_2 > 4$. If either of the two arities is 3, then we connect two copies together via $\neq_2$ to realize an arity 4 copy.

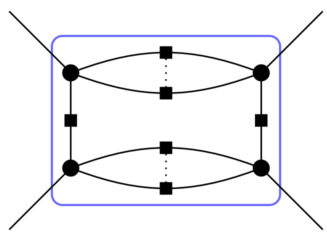


Figure 6.11: Gadget to realize $\overline{g}$ in Lemma 6.11. Circle vertices are assigned $=_k$ and square vertices are assigned $\neq_2$. The number of parallel edges is $k - 2$.

Moreover, we use the gadget illustrated in Figure 6.11 to create the function $\overline{g}$ in Lemma 5.39 as an E_k-block. Then by Lemma 5.39, Pl-Holant$\left(\neq_2 \mid =_k, \textsc{ExactOne}_{d_1}, \textsc{AllButOne}_{d_2}\right)$ is #**P**-hard. □

In general signatures in $\mathcal{P}_2$ are non-degenerate weighted equalities under the Z transformation. The next several lemmas show that the hardness criterion is the same regardless of the weight.

Lemma 6.12. *Let* $f \in \mathcal{P}_2$, $g_1 \in \mathcal{M}_4^+$, $g_2 \in \mathcal{M}_4^-$ *be non-degenerate signatures with arity* $\geqslant 3$*. Then* Pl-Holant(f, g_1, g_2) *is* #**P**-*hard.*

Proof. Suppose the arities of f, g_1, and g_2 are n, m_1, and m_2 respectively. Under a holographic transformation by Z, we have

$$\begin{aligned}\text{Pl-Holant}(f, g_1, g_2) &\equiv \text{Pl-Holant}\left(\neq_2 \mid \left(Z^{-1}\right)^{\otimes n} f, \left(Z^{-1}\right)^{\otimes m_1} g_1, \left(Z^{-1}\right)^{\otimes m_2} g_2\right) \\ &\equiv \text{Pl-Holant}\left(\neq_2 \mid \bar{f}, \textsc{ExactOne}_{m_1}, \textsc{AllButOne}_{m_2}\right),\end{aligned}$$

where $\bar{f} = (Z^{-1})^{\otimes n} f$ which has the form $[1, 0, \ldots, 0, c]$ up to a nonzero constant, with $c \neq 0$, as $f \in \mathcal{P}_2$. We do another diagonal transformation by $D = \left[\begin{smallmatrix} 1 & 0 \\ 0 & c^{1/n} \end{smallmatrix}\right]$. Then

$$\begin{aligned}&\text{Pl-Holant}(f, g_1, g_2) \\ &\quad\equiv \text{Pl-Holant}\left((\neq_2) D^{\otimes 2} \,\middle|\, (D^{-1})^{\otimes n} \bar{f}, (D^{-1})^{\otimes m_1} \textsc{ExactOne}_{m_1}, (D^{-1})^{\otimes m_2} \textsc{AllButOne}_{m_2}\right) \\ &\quad\equiv \text{Pl-Holant}\left(\neq_2 \mid =_n, \textsc{ExactOne}_{m_1}, \textsc{AllButOne}_{m_2}\right),\end{aligned}$$

where in the last line we ignored several nonzero factors. The lemma follows from Lemma 6.11.

□

We also need to consider the mixture of $\mathcal{P}_2$ and binary signatures.

Lemma 6.13. *Let $\mathcal{F}$ be a set of symmetric signatures. Suppose $\mathcal{F}$ contains a non-degenerate signature $f \in \mathcal{P}_2$ of arity $n \geqslant 3$ and a binary signature h. Then* Pl-Holant$(\mathcal{F})$ *is* #**P**-*hard unless* $h \in Z\mathcal{P}$, *or* Pl-#CSP$^2(DZ^{-1}\mathcal{F}) \leqslant_T$ Pl-Holant$(\mathcal{F})$ *for some diagonal transformation* D.

Proof. We do a Z transformation and get

$$\begin{aligned}\text{Pl-Holant}(\mathcal{F}) &\equiv \text{Pl-Holant}(\mathcal{F}, h, f) \\ &\equiv \text{Pl-Holant}\left(\neq_2 \mid Z^{-1}\mathcal{F}, \left(Z^{-1}\right)^{\otimes 2} h, \bar{f}\right),\end{aligned}$$

where $\bar{f} = (Z^{-1})^{\otimes n} f = [1, 0, \ldots, 0, t]$ up to a nonzero constant with $t \neq 0$. We further do another diagonal transformation of $D_1 = \begin{bmatrix} 1 & 0 \\ 0 & t^{1/n} \end{bmatrix}$. Then

$$\begin{aligned}
\text{Pl-Holant}(\mathcal{F}) &\equiv \text{Pl-Holant}\left((\neq_2) D_1^{\otimes 2} \mid (D_1^{-1})^{\otimes n} \bar{f}, (ZD_1)^{-1}\mathcal{F}, \left((ZD_1)^{-1}\right)^{\otimes 2} h \right) \\
&\equiv \text{Pl-Holant}\left(\neq_2 \mid =_n, (ZD_1)^{-1}\mathcal{F}, \left((ZD_1)^{-1}\right)^{\otimes 2} h \right) \\
&\geqslant_T \text{Pl-Holant}\left(=_n \mid (ZD_1)^{-1}\mathcal{F}, \left((ZD_1)^{-1}\right)^{\otimes 2} h \right),
\end{aligned}$$

where in the second line we ignore a nonzero factor on $\neq_2$. Hence by Theorem 1.15, Pl-Holant$(\mathcal{F})$ is #**P**-hard unless $\left((ZD_1)^{-1}\right)^{\otimes 2} h \in \mathcal{P}$ (cases 1, 2 or 3 in Theorem 1.15) or $\left((ZD_1)^{-1}\right)^{\otimes 2} h = [a, b, c]$ for some $a, b, c \in \mathbb{C}$ such that $ac \neq 0$ and $(a/c)^{2n} = 1$ (cases 4 or 5 in Theorem 1.15).

In the former case, $\left((ZD_1)^{-1}\right)^{\otimes 2} h \in \mathcal{P}$. Then $h \in ZD_1\mathcal{P} = Z\mathcal{P}$ as $D_1 \in \text{Stab}(\mathcal{P})$. In the latter case, we construct $=_{2n}$ on the right by connecting three copies of $=_n$ to one copy of $=_n$ via $\neq_2$. We do the same construction again to realize $=_{4n}$ using $=_{2n}$. We connect $n-1$ many $[a, b, c]$'s to $=_{2n}$ via $\neq_2$ to realize a binary weighted equality $[1, 0, r]$ with $r = (a/c)^{n-1} \neq 0$ ignoring a factor of c^{n-1}. Note that $r^{2n} = (a/c)^{2n(n-1)} = 1$. Then we do another diagonal transformation of $D_2 = \begin{bmatrix} 1 & 0 \\ 0 & r^{1/2} \end{bmatrix}$ to get Pl-Holant$\left(\neq_2 \mid (ZD_1D_2)^{-1}\mathcal{F}, =_2, \left(D_2^{-1}\right)^{\otimes 4n}(=_{4n})\right)$. Notice that

$$\left(D_2^{-1}\right)^{\otimes 4n}(=_{4n}) = [1, 0, \ldots, 0, r^{-2n}] = (=_{4n}),$$

as $r^{2n} = 1$.

Hence we have $=_2$ and $=_{4n}$ on the right. With $\neq_2$ on the left, we get $=_2$ on the left and therefore equalities of all even arities on the right. Let $D = (D_1D_2)^{-1}$. Then we have the reduction chain:

$$\begin{aligned}
\text{Pl-Holant}(\mathcal{F}) &\geqslant_T \text{Pl-Holant}\left(\neq_2 \mid DZ^{-1}\mathcal{F} \cup \{=_2, =_{4n}\} \right) \\
&\geqslant_T \text{Pl-Holant}\left(\neq_2 \mid DZ^{-1}\mathcal{F} \cup \mathcal{EQ}_2 \right) \\
&\geqslant_T \text{Pl-Holant}\left(\mathcal{EQ}_2 \mid DZ^{-1}\mathcal{F} \right).
\end{aligned}$$

The last problem is Pl-#CSP$^2(DZ^{-1}\mathcal{F})$. Thus Pl-#CSP$^2(DZ^{-1}\mathcal{F}) \leqslant_T$ Pl-Holant$(\mathcal{F})$. □

At last, we strengthen Corollary 6.10, Lemma 6.3, and Lemma 6.5 to weighted equalities.

We split the hardness and tractability cases. For a set $\mathcal{F}$ of signatures, denote by $\mathcal{F}_{nd}^{\geqslant 3}$ the set of non-degenerate signatures in $\mathcal{F}$ of arity at least 3. Moreover denote by $\mathcal{F}^*$ the signature set that is the same as $\mathcal{F}$ but with each degenerate signature $[a,b]^{\otimes m}$ in $\mathcal{F}$ replaced by the unary $[a,b]$.

Notice that $\mathcal{F} \cap \mathcal{P}_2$ and $\mathcal{F}^* \cap \mathcal{P}_2$ agree on signatures of arity at least 2, since signatures in $\mathcal{P}_2$ of arity at least 2 are non-degenerate. So $\mathcal{F} \cap \mathcal{P}_2 \subseteq \mathcal{F}^* \cap \mathcal{P}_2$, and the only possible extra elements are some unary $[x,y]$'s from $[x,y]^{\otimes m} \in \mathcal{F}$ for some integer $m \geqslant 2$ and $[x,y]$ is not a multiple of $[1, \pm i]$. Equivalently the only possible extra elements are unary signatures of the form $Z[a,b]$ for $ab \neq 0$, i.e., *not* of the form a multiple of $Z[1,0]$ or $Z[0,1]$, when $Z^{-1}\mathcal{F}$ contains some degenerate signatures of the form $[a,b]^{\otimes m}$ for some integer $m \geqslant 2$ and $ab \neq 0$.

Lemma 6.14. *Let $\mathcal{F}$ be a set of symmetric signatures. Let $\mathcal{F}_{nd}^{\geqslant 3}$ be the set of non-degenerate signatures in $\mathcal{F}$ of arity at least 3. Suppose $\mathcal{F}_{nd}^{\geqslant 3}$ contains $f \in \mathcal{M}_4$ of arity $d \geqslant 3$. Moreover, suppose $\mathcal{F}_{nd}^{\geqslant 3} \cap \mathcal{P}_2$ is nonempty, and let k be the greatest common divisor of the arities of signatures in $\mathcal{F}^* \cap \mathcal{P}_2$. If $k \leqslant 4$, then* Pl-Holant$(\mathcal{F})$ *is #**P**-hard.*

Proof. We may assume that $f \in \mathcal{M}_4^+$. Since $\mathcal{F}_{nd}^{\geqslant 3} \cap \mathcal{P}_2$ is nonempty, there exists $g \in \mathcal{F}_{nd}^{\geqslant 3} \cap \mathcal{P}_2$. By the definition of $\mathcal{F}_{nd}^{\geqslant 3}$, g has arity $n \geqslant 3$. We do a Z transformation,

$$\text{Pl-Holant}(\mathcal{F}) \equiv \text{Pl-Holant}\left(\neq_2 \mid \overline{g}, \textsc{ExactOne}_d, Z^{-1}\mathcal{F}\right),$$

where $\overline{g} = (Z^{-1})^{\otimes n} g$ has the form $[1,0,\ldots,0,c]$ of arity n for some $c \neq 0$ up to a nonzero factor. We further do a diagonal transformation $D = \left[\begin{smallmatrix} 1 & 0 \\ 0 & c^{1/n} \end{smallmatrix}\right]$ and get

$$\text{Pl-Holant}(\mathcal{F}) \equiv \text{Pl-Holant}\left(\neq_2 \mid =_n, \textsc{ExactOne}_d, (ZD)^{-1}\mathcal{F}\right),$$

where we ignore nonzero factors on $\neq_2$ and $\textsc{ExactOne}_d$. Then by Lemma 6.9,

$$\text{Pl-Holant}(\mathcal{F}) \geqslant_T \text{Pl-Holant}\left(\neq_2 \mid =_n, \textsc{ExactOne}_d, [0,1], [1,0], (ZD)^{-1}\mathcal{F}\right).$$

By a weighted equality we mean a signature of the form $[a,0,\ldots,0,b]$ of some arity $\geqslant 1$, where $ab \neq 0$. Recall that $\mathcal{P}_2$ consists of all weighted equalities in the Z basis. Let $\mathcal{G}$ be the set of weighted equalities in $(ZD)^{-1}\mathcal{F}$. In other words, $\mathcal{G} = (ZD)^{-1}(\mathcal{F} \cap \mathcal{P}_2)$ as $(ZD)^{-1}\mathcal{P}_2$ contains all weighted equalities. Moreover, up to a nonzero factor, $(=_n) \in \mathcal{G}$.

Let k' be the gcd of all arities of signatures in $\mathcal{G}$, or equivalently the gcd of all arities of signatures in $\mathcal{F} \cap \mathcal{P}_2$. If $k' \neq k$, then the only possibility is that $(ZD)^{-1}\mathcal{F}$ contains a degenerate signature $[a, b]^{\otimes m}$ for some $m \geqslant 2$ with $ab \neq 0$. In this case we use pinnings $[1, 0]$ or $[0, 1]$ to realize $[a, b]$ from $[a, b]^{\otimes m}$ and put $[a, b]$ in $\mathcal{G}$. Hence we may assume that $k' = k$.

Pick any $g_1, g_2 \in \mathcal{G}$ of arities ℓ_1 and ℓ_2. Let $r = \gcd(\ell_1, \ell_2)$. Let t_1, t_2 be two positive integers such that $t_1\ell_1 - t_2\ell_2 = r$. Then connecting t_1 copies of g_1 and t_2 copies of g_2 via $\neq_2$ in a bipartite and planar way, we get a weighted equality signature of arity r.

Apply the same argument repeatedly. Eventually we construct a weighted equality h of arity k. We further do a diagonal transformation D_1 to make it $=_k$, that is,

$$\begin{aligned}
\text{Pl-Holant}(\mathcal{F}) &\geqslant_T \text{Pl-Holant}\left(\neq_2 \mid \mathcal{G}, \textsc{ExactOne}_d\right) \\
&\geqslant_T \text{Pl-Holant}\left(\neq_2 \mid h, \textsc{ExactOne}_d, \mathcal{G}\right) \\
&\geqslant_T \text{Pl-Holant}\left((\neq_2)D_1^{\otimes 2} \mid =_k, \left(D_1^{-1}\right)^{\otimes d} \textsc{ExactOne}_d, D_1^{-1}\mathcal{G}\right) \\
&\geqslant_T \text{Pl-Holant}\left(\neq_2 \mid =_k, \textsc{ExactOne}_d, D_1^{-1}\mathcal{G}\right),
\end{aligned}$$

where in the last line we ignored nonzero factors of $\textsc{ExactOne}_d$ and $\neq_2$. If $k = 3$ or 4, then the hardness follows from Corollary 6.10.

If $k = 1$ or 2, then on the right hand side we have $=_k$, which is $=_1$ or $=_2$, and a weighted equality $\left(D_1^{-1}\right)^{\otimes n}(=_n) \in D_1^{-1}\mathcal{G}$. Call it $\overline{g}'$. We move the $=_k$ to the left hand side via $\neq_2$. Then we connect zero or more copies of this $=_k$, which is $=_1$ or $=_2$, to $\overline{g}'$ such that its arity is 3 or 4. It is possible that $n = 3$ or 4 to begin with, and if so we do nothing. We are done by yet another diagonal transformation and Corollary 6.10. □

Lemma 6.15. *Let $\mathcal{F}$ be a set of symmetric signatures. Suppose $\mathcal{F} \subseteq Z\mathcal{P} \cup \mathcal{M}_4^{\sigma}$ for some $\sigma \in \{+, -\}$ and the greatest common divisor of the arities of all signatures in $\mathcal{F}^* \cap \mathcal{P}_2$ is $k \geqslant 5$. Then* Pl-Holant$(\mathcal{F})$ *can be computed in polynomial time.*

Proof. We may assume that $\sigma = +$ and the case of $\sigma = -$ is similar. We do a Z transformation on Pl-Holant$(\mathcal{F})$, and get a problem of Pl-Holant$\left(\neq_2 \mid Z^{-1}\mathcal{F}\right)$.

In this bipartite setting, given $=_n$ on the right hand side, we can realize $=_{\ell n}$ for any integer $\ell \geqslant 1$ as an E_n-block on the right. The problem Pl-Holant $(\neq_2 \mid \mathcal{EQ}_n, \mathcal{EO}, \neq_2, [1,0], [0,1])$ is tractable for any $n \geqslant 5$ by Lemma 6.3 and Lemma 6.5, where $\mathcal{EQ}_n$ denotes the set of all equalities of arity ℓn for all integers $\ell \geqslant 1$.

The symmetric signatures in the set $Z\mathcal{P}$ consist of $\mathcal{P}_2$, $Z^{\otimes 2}(\neq_2)$, and degenerate signatures. If there is any degenerate signature of the form $(Z[a,b])^{\otimes m} \in \mathcal{F}$ with $ab \neq 0$, then $Z[a,b] \in \mathcal{F}^* \cap \mathcal{P}_2$. This contradicts $k \geqslant 5$. Hence all degenerate signatures in $\mathcal{F}$ are of the form $(Z[1,0])^{\otimes m}$ or $(Z[0,1])^{\otimes m}$, if any. Since $\mathcal{F} \subseteq Z\mathcal{P} \cup \mathcal{M}_4^+$, after a Z transformation, Pl-Holant$(\mathcal{F})$ is an instance of Pl-Holant $(\neq_2 \mid \mathcal{EQ}_k, \mathcal{EO}, \neq_2, [1,0], [0,1])$ except for the weights on the equalities. It can be checked that the tractability results of Lemma 6.3 and Lemma 6.5 also apply to weighted equalities. The lemma follows. □

6.4 #PM in Planar Hypergraphs

Let $\mathcal{G} = \{=_k \mid k \in S\}$ be a set of EQUALITY signatures, where S is a set of positive integers containing at least one $r \geqslant 3$. Moreover let $\mathcal{EO}^+ := \{\text{EXACTONE}_d \mid d \in \mathbb{Z}^+\} = \mathcal{EO} \cup \{\neq_2, [0,1]\}$. Then Pl-Holant $\left(\mathcal{G} \mid \mathcal{EO}^+\right)$ is the problem of counting perfect matchings over hypergraphs with planar incidence graphs, where the hyperedge sizes are prescribed by S. In the incidence graph, vertices assigned signatures in $\mathcal{G}$ on the left represent hyperedges, and vertices assigned signatures in $\mathcal{EO}^+$ on the right represent vertices of the hypergraph. Let $t = \gcd(S)$. By Lemma 6.3 and 6.5, this problem is tractable if $t \geqslant 5$ since we can reduce Pl-Holant $\left(\mathcal{G} \mid \mathcal{EO}^+\right)$ to Pl-Holant $(\neq_2 \mid =_t, \mathcal{EO}, \neq_2, [0,1])$. The reduction goes as follows. With $\neq_2$ on the left hand side and $=_t$ on the right hand side, we can construct all E_t-blocks and hence all of $\mathcal{EQ}_t$ on the right. Note that $\mathcal{G} \subseteq \mathcal{EQ}_t$. Then we move all signatures in $\mathcal{G}$ to the left via $\neq_2$.

If $t \leqslant 4$, then Pl-Holant $\left(\mathcal{G} \mid \mathcal{EO}^+\right)$ is #**P**-hard due to Corollary 6.10. The reason is as follows. We construct $\neq_2$ on the left using the gadget pictured in Figure 5.4b with $(=_r) \in \mathcal{G}$ on the left side assigned to circle vertices and $\neq_2$ on the right side assigned to square vertices. Then we move $\mathcal{G}$ to the right side via $\neq_2$ on the right side. We construct $=_t$ on the right side in the subtractive Euclidean process using $\mathcal{G}$ of the right side and $\neq_2$ of the left side. This gives us a reduction from Pl-Holant $(\neq_2 \mid =_t, \mathcal{EO})$, which is #**P**-hard by Corollary 6.10 if $t = 3, 4$. Otherwise

$t = 1, 2$. Recall that $(=_r) \in \mathcal{G}$ for some $r \geqslant 3$. We use $=_t$ to reduce the arity of $=_r$ to 3 or 4, if necessary. Again we are done by Corollary 6.10.

If we do not assume there is at least one hyperedge of size $\geqslant 3$ in Pl-Holant $(\mathcal{G} \mid \mathcal{EO}^+)$, and $t = \gcd(S) \leqslant 2$, then the problem is tractable if and only if $S \subseteq \{1, 2\}$. The tractability is due to Kasteleyn's algorithm, as there is no hyperedge. In summary, we have the following theorem.

Theorem 6.16. *The problem* Pl-Holant $(\mathcal{G} \mid \mathcal{EO}^+)$ *counts perfect matchings over hypergraphs with planar incidence graphs, where the hyperedge sizes are prescribed by a set* S *of positive integers. Let* $t = \gcd(S)$*. If* $t \geqslant 5$ *or* $S \subseteq \{1, 2\}$*, then the problem is computable in polynomial time. Otherwise* $t \leqslant 4$, $S \not\subseteq \{1, 2\}$*, and the problem is #***P***-hard.*

6.5 The Full Dichotomy

We are finally ready to prove our full dichotomy theorem. Recall that for a set $\mathcal{F}$ of signatures, $\mathcal{F}_{nd}^{\geqslant 3}$ denotes the set of non-degenerate signatures in $\mathcal{F}$ of arity at least 3, and $\mathcal{F}^*$ denotes $\mathcal{F}$ with all degenerate signatures $[a, b]^{\otimes m}$ replaced by unary $[a, b]$.

Theorem 6.17. *Let* $\mathcal{F}$ *be any set of symmetric, complex-valued signatures in Boolean variables. Then* Pl-Holant$(\mathcal{F})$ *is #***P***-hard unless* $\mathcal{F}$ *satisfies one of the following conditions:*

1. *All non-degenerate signatures in* $\mathcal{F}$ *are of arity at most 2;*
2. $\mathcal{F}$ *is* $\mathcal{A}$*-transformable;*
3. $\mathcal{F}$ *is* $\mathcal{P}$*-transformable;*
4. $\mathcal{F} \subseteq \mathcal{V}^\sigma \cup \{f \in \mathcal{R}_2^\sigma \mid \text{arity}(f) = 2\}$ *for some* $\sigma \in \{+, -\}$*;*
5. *All non-degenerate signatures in* $\mathcal{F}$ *are in* $\mathcal{R}_2^\sigma$ *for some* $\sigma \in \{+, -\}$*.*
6. $\mathcal{F}$ *is* $\mathcal{M}$*-transformable;*
7. $\mathcal{F} \subseteq Z\mathcal{P} \cup \mathcal{M}_4^\sigma$ *for some* $\sigma \in \{+, -\}$*, and the greatest common divisor of the arities of the signatures in* $\mathcal{F}^* \cap \mathcal{P}_2$ *is at least* 5.

In each exceptional case, Pl-Holant$(\mathcal{F})$ *is computable in polynomial time. If* $\mathcal{F}$ *satisfies conditions 1 to 5, then* Holant$(\mathcal{F})$ *is computable in polynomial time without planarity; otherwise* Holant$(\mathcal{F})$ *is #***P***-hard.*

Proof. Tractability for Cases 1 to 7 follows from Lemma 1.5, 1.7, 1.9, 3.15, 3.16, 1.10, and 6.15, respectively.

We may assume that $\mathcal{F}$ contains no identically 0 signatures. We note that removing any identically 0 signature from a set does not affect its complexity, being either tractable or #**P**-hard, and it does not affect the set $\mathcal{F}$ satisfying any of the exceptional conditions in Case 1 to 7.

Next we prove the claim for Pl-Holant$(\mathcal{F})$. Suppose Pl-Holant$(\mathcal{F})$ is not #**P**-hard. If all non-degenerate signatures in $\mathcal{F}$ are of arity at most 2, then the problem is tractable case 1. Otherwise, there is a non-degenerate signature $f \in \mathcal{F}$ of arity at least 3. By Theorem 5.41, Pl-Holant$(\mathcal{F})$ is #**P**-hard unless $f \in \mathcal{P}_1 \cup \mathcal{M}_2 \cup \mathcal{A}_3 \cup \mathcal{M}_3 \cup \mathcal{M}_4$ or f is vanishing. If $f \in \mathcal{P}_1$ or $f \in \mathcal{M}_2 \setminus \mathcal{P}_2$ or $f \in \mathcal{A}_3$ or $f \in \mathcal{M}_3$, then we are done by Corollary 5.25, or Corollary 5.31, or Corollary 5.28, or Lemma 5.33 respectively. Therefore, we assume that none of these is the case. This implies that $\mathcal{F}_{nd}^{\geqslant 3}$ is nonempty and that each of its signatures is in $\mathcal{P}_2$ or in $\mathcal{M}_4$ or vanishing. That is,

$$\emptyset \neq \mathcal{F}_{nd}^{\geqslant 3} \subseteq \mathcal{P}_2 \cup \mathcal{M}_4 \cup \mathcal{V}.$$

Suppose there exists some $f \in \mathcal{F}_{nd}^{\geqslant 3}$ which is in $\mathcal{V} \setminus \mathcal{M}_4$. We assume $f \in \mathcal{V}^+$ since the other case $\mathcal{V}^-$ is similar. In this case, we show that Pl-Holant$(\mathcal{F})$ is #**P**-hard, unless $\mathcal{F}$ is in Case 4 or Case 5. Assume that Pl-Holant$(\mathcal{F})$ is not #**P**-hard. We will discuss non-degenerate signatures of arity $\geqslant 3$, of arity 2, and degenerate signatures separately.

1. For any $g \in \mathcal{F}_{nd}^{\geqslant 3}$, we claim that $g \in \mathcal{V}^+$. Suppose otherwise, then $g \in \mathcal{P}_2$ or $g \in \mathcal{V}^-$. Notice that the latter covers the case where $g \in \mathcal{M}_4$ but $g \notin \mathcal{V}^+$ (namely $g \in \mathcal{M}_4^-$). If $g \in \mathcal{P}_2$, then Pl-Holant(f, g) is #**P**-hard by Lemma 3.48 and Lemma 5.8, with a possible diagonal transformation in the Z basis. Notice that a diagonal in the Z basis is equivalent to an orthogonal in the standard basis, which does not affect the complexity. If $g \in \mathcal{V}^-$, then Pl-Holant(f, g) is #**P**-hard by Lemma 3.45 as $f \notin \mathcal{M}_4$.

2. For any non-degenerate binary signature $h \in \mathcal{F}$, it must be that $h \in \mathcal{R}_2^+$ as otherwise Pl-Holant(f, h) is #**P**-hard by Lemma 3.43.

3. If $\mathrm{rd}^+(g) = 1$ for all $g \in \mathcal{F}_{nd}^{\geqslant 3}$, then $\mathcal{F}_{nd}^{\geqslant 3} \subseteq \mathcal{R}_2^+$ by Lemma 3.14. Together with the fact just proved that all non-degenerate binary in $\mathcal{F}$ are in $\mathcal{R}_2^+$, Case 5 is satisfied.

   Otherwise there exists $g \in \mathcal{F}_{nd}^{\geqslant 3}$ such that $\mathrm{rd}^+(g) \geqslant 2$. Then $g \in \mathcal{V}^+$ by the first item above. If $\mathcal{F}$ contains any degenerate signature $v = u^{\otimes m}$ for $m \geqslant 1$ and some unary u that is not a multiple of $[1, i]$, then by Lemma 3.41, Pl-Holant(g, v) is #**P**-hard. Hence all degenerate signatures are multiples of tensor powers of $[1, i]$, which are in $\mathcal{V}^+$. It implies that $\mathcal{F}$ is in Case 4.

Now we may assume that $\emptyset \neq \mathcal{F}_{nd}^{\geqslant 3} \subseteq \mathcal{P}_2 \cup \mathcal{M}_4$. We handle this in three cases.

1. Suppose $\mathcal{F}_{nd}^{\geqslant 3} \subseteq \mathcal{M}_4$. First suppose $\mathcal{F}_{nd}^{\geqslant 3} \subseteq \mathcal{M}_4^\sigma$ for some $\sigma \in \{+, -\}$. Assume $\sigma = +$ as $\sigma = -$ is similar. Then $\mathcal{F}_{nd}^{\geqslant 3} \subseteq \mathcal{R}_2^+$ by Lemma 5.17 and 3.14. If all non-degenerate binary signatures are in $\mathcal{R}_2^+$ as well, then this is Case 5 and tractable. Let h be a non-degenerate binary signature in $\mathcal{F}$ that is not in $\mathcal{R}_2^+$. We apply Lemma 3.44, and Pl-Holant$(\mathcal{F})$ is #**P**-hard unless $h = Z^{\otimes 2}[a, 0, 1]$ up to a nonzero factor, where $a \neq 0$. In this case we apply a Z transformation, and get Pl-Holant $\left(\neq_2 | \, [a, 0, 1], Z^{-1}\mathcal{F}\right)$. Then we do a diagonal transformation $D = \begin{bmatrix} a^{1/2} & 0 \\ 0 & 1 \end{bmatrix}$. Note that this only changes $\neq_2$ on the left hand side to a nonzero multiple of $\neq_2$. Hence we have the reduction chain:

$$\begin{aligned} \text{Pl-Holant}(\mathcal{F}) &\equiv \text{Pl-Holant}\left(\neq_2 \mid [a, 0, 1], Z^{-1}\mathcal{F}\right) \\ &\equiv \text{Pl-Holant}\left(\neq_2 \mid [1, 0, 1], D^{-1}Z^{-1}\mathcal{F}\right) \\ &\geqslant_T \text{Pl-Holant}(D^{-1}Z^{-1}\mathcal{F}) \end{aligned}$$

   Notice that $D^{-1}Z^{-1}\mathcal{F}$ contains $\textsc{ExactOne}_k$ with $k \geqslant 3$ that is in $\mathcal{M}_3$ with I_2. Then by Lemma 5.33, Pl-Holant$(\mathcal{F})$ is #**P**-hard unless $D^{-1}Z^{-1}\mathcal{F} \subseteq I_2\mathcal{M} = \mathcal{M}$, i.e., $\mathcal{F} \subseteq ZD\mathcal{M} = Z\mathcal{M}$. The exceptional case implies that $\mathcal{F}$ is $\mathcal{M}$-transformable via Z, and we are in the tractable Case 6.

Otherwise $\mathcal{F}_{nd}^{\geqslant 3}$ contains both $f \in \mathcal{M}_4^+$ and $g \in \mathcal{M}_4^-$. Similarly as above, by Lemma 3.44, any non-degenerate binary signature in $\mathcal{F}$ has to be in $\mathcal{R}_2^+ \cap \mathcal{R}_2^- = \{Z^{\otimes 2}(\neq_2)\}$ (cf. Lemma 3.14), or is a nonzero constant multiple of $Z^{\otimes 2}[a, 0, 1]$ where $a \neq 0$, as otherwise Pl-Holant($\mathcal{F}$) is #**P**-hard. Moreover, by Lemma 3.46, Pl-Holant($\mathcal{F}$) is #**P**-hard, unless all degenerate signatures in $\mathcal{F}$ are of the form $[1, \pm i]^{\otimes m}$. Note that $[1, i] = Z[1, 0]$ and $[1, -i] = Z[0, 1]$. When this is the case, $\mathcal{F}$ is $\mathcal{M}$-transformable via Z.

2. Suppose $\mathcal{F}_{nd}^{\geqslant 3} \subseteq \mathcal{P}_2$. If $\mathcal{F}$ contains a non-degenerate binary signature h, then we apply Lemma 6.13 and Pl-Holant($\mathcal{F}$) is #**P**-hard unless $h \in Z\mathcal{P}$, or Pl-#CSP$^2(DZ^{-1}\mathcal{F}) \leqslant_T$ Pl-Holant($\mathcal{F}$) for some diagonal transformation D. If it is the latter case, then by Theorem 5.22, either Pl-Holant($\mathcal{F}$) is #**P**-hard, or $DZ^{-1}\mathcal{F}$ is a subset of $T\mathcal{A}$, $\mathcal{P}$, or $T\left[\begin{smallmatrix}1 & 1\\ 1 & -1\end{smallmatrix}\right]\mathcal{M}$, for some diagonal matrix T. We claim that in any of these cases Pl-Holant($\mathcal{F}$) is tractable. In fact,

    a) if $DZ^{-1}\mathcal{F} \subseteq T\mathcal{A}$, then $\mathcal{F}$ is $\mathcal{A}$-transformable as $\mathcal{F} \subseteq ZD^{-1}T\mathcal{A}$ and $[1, 0, 1]$ (as a row vector) is transformed into $[1, 0, 1](ZD^{-1}T)^{\otimes 2}$, which is $[0, 1, 0] \in \mathcal{A}$ up to a nonzero constant;

    b) if $DZ^{-1}\mathcal{F} \subseteq \mathcal{P}$, then $\mathcal{F}$ is $\mathcal{P}$-transformable as $\mathcal{F} \subseteq ZD^{-1}\mathcal{P}$ and $[1, 0, 1](ZD^{-1})^{\otimes 2}$ is $[0, 1, 0] \in \mathcal{P}$ up to a nonzero constant;

    c) if $DZ^{-1}\mathcal{F} \subseteq T\left[\begin{smallmatrix}1 & 1\\ 1 & -1\end{smallmatrix}\right]\mathcal{M}$, then $\mathcal{F}$ is $\mathcal{M}$-transformable as $\mathcal{F} \subseteq ZD^{-1}T\left[\begin{smallmatrix}1 & 1\\ 1 & -1\end{smallmatrix}\right]\mathcal{M}$ and $[1, 0, 1]$ is transformed to $[1, 0, 1](ZD^{-1}T\left[\begin{smallmatrix}1 & 1\\ 1 & -1\end{smallmatrix}\right])^{\otimes 2}$, which is $[1, 0, -1] \in \mathcal{M}$ up to a nonzero constant.

    Hence we may assume that every non-degenerate binary in $\mathcal{F}$ is in $Z\mathcal{P}$. Notice that degenerate signatures are always in $\mathcal{P}$ under any transformation. Also $\mathcal{F}_{nd}^{\geqslant 3}$ is a subset of $Z\mathcal{P}$ because $\mathcal{F}_{nd}^{\geqslant 3} \subseteq \mathcal{P}_2$ and $\mathcal{P}_2$ is just weighted equalities under Z-transformation. It implies that $\mathcal{F}$ is $\mathcal{P}$-transformable under the Z transformation. Hence we are in Case 3.

3. Finally, suppose neither of the above is the case. Then there are $f, g \in \mathcal{F}_{nd}^{\geqslant 3}$ with $f \in \mathcal{M}_4$ and $g \in \mathcal{P}_2$. If $\mathcal{F}_{nd}^{\geqslant 3}$ contains both $f \in \mathcal{M}_4^+$ and $f' \in \mathcal{M}_4^-$, then Pl-Holant($\mathcal{F}$) is #**P**-hard by Lemma 6.12. Otherwise $\mathcal{F}_{nd}^{\geqslant 3} \cap \mathcal{M}_4 \subseteq \mathcal{M}_4^+$ or $\mathcal{M}_4^-$. Let $\mathcal{G} = \mathcal{F}^* \cap \mathcal{P}_2$, and let d be the gcd of the arities of the signatures in $\mathcal{G}$. Then $\mathcal{G}$ contains at least one non-degenerate

signature g of arity $\geqslant 3$. If $d \leqslant 4$, then Pl-Holant($\mathcal{F}$) is #**P**-hard by Lemma 6.14. Otherwise $d \geqslant 5$. If $\mathcal{F}$ contains a non-degenerate binary signature h, then we apply Lemma 6.13 and by a similar analysis as in the case of "$\mathcal{F}_{nd}^{\geqslant 3} \subseteq \mathcal{P}_2$" above, we are done unless every such h is in $Z\mathcal{P}$. Ignoring a nonzero factor, it implies that either $h = Z^{\otimes 2}[1, 0, a]$ where $a \neq 0$ or $h = Z^{\otimes 2}(\neq_2)$. If $h = Z^{\otimes 2}[1, 0, a]$, then $h \in \mathcal{F}^* \cap \mathcal{P}_2$, and it contradicts $d \geqslant 5$. Hence $h = Z^{\otimes 2}(\neq_2)$. If there is any degenerate $v = (Z[a, b])^{\otimes m}$ in $\mathcal{F}$ with $ab \neq 0$, then $Z[a, b] \in \mathcal{F}^* \cap \mathcal{P}_2$ and it also contradicts to $d \geqslant 5$.

In summary, Pl-Holant($\mathcal{F}$) is #**P**-hard unless $\mathcal{F}_{nd}^{\geqslant 3} \subseteq \mathcal{P}_2 \cup \mathcal{M}_4$, $\mathcal{F}_{nd}^{\geqslant 3} \cap \mathcal{M}_4 \subseteq \mathcal{M}_4^{\sigma}$ for some $\sigma \in \{+, -\}$, the greatest common divisor of the arities of the signatures in $\mathcal{F}^* \cap \mathcal{P}_2$ is at least 5. Every non-degenerate binary in $\mathcal{F}$ is of the form $Z^{\otimes 2}(\neq_2)$, and every degenerate in $\mathcal{F}$ is of the form $(Z[1, 0])^{\otimes m}$ or $(Z[0, 1])^{\otimes m}$. Notice that $\mathcal{P}_2$, $Z^{\otimes 2}(\neq_2)$, $(Z[1, 0])^{\otimes m}$, and $(Z[0, 1])^{\otimes m}$ are all in $Z\mathcal{P}$. Hence the exceptional case implies that $\mathcal{F} \subseteq Z\mathcal{P} \cup \mathcal{M}_4^{\sigma}$ for some $\sigma \in \{+, -\}$ and the greatest common divisor of the arities of the signatures in $\mathcal{F}^* \cap \mathcal{P}_2$ is at least 5. This is tractable Case 7.

Next we turn to Holant($\mathcal{F}$). Unless $\mathcal{F}$ is in Case 1, $\mathcal{F}_{nd}^{\geqslant 3} \neq \emptyset$. By Theorem 5.41, any $f \in \mathcal{F}_{nd}^{\geqslant 3}$ has to be in $\mathcal{P}_1 \cup \mathcal{P}_2 \cup \mathcal{A}_3 \cup \mathcal{V}$ and otherwise Holant($\mathcal{F}$) is #**P**-hard.

We deal with $\mathcal{P}_1$, $\mathcal{P}_2$, and $\mathcal{A}_3$ first. It is easy to check that reductions in Lemmas 5.24 and 5.27 do not require the planarity constraint. We will use them below.

- If there exists $f \in \mathcal{F}_{nd}^{\geqslant 3}$ such that $f \in \mathcal{P}_1$, then by Lemma 5.24, $\#\mathrm{CSP}^2(H\mathcal{F}) \leqslant_T \mathrm{Holant}(\mathcal{F})$ for some $H \in \mathbf{O}_2(\mathbb{C})$. Then by Theorem 6.1, $\#\mathrm{CSP}^2(H\mathcal{F}) \leqslant_T \mathrm{Holant}(\mathcal{F})$ is #**P**-hard, and so is Holant($\mathcal{F}$), unless $\mathcal{F}$ is $\mathcal{A}$- or $\mathcal{P}$-transformable.

- If there exists $f \in \mathcal{F}_{nd}^{\geqslant 3}$ such that $f \in \mathcal{A}_3$, then by Lemma 5.27, $\#\mathrm{CSP}^2(YH^{-1}\mathcal{F} \cup \{[1, -i, 1]\}) \leqslant_T \mathrm{Holant}(\mathcal{F})$, where $Y = \left[\begin{smallmatrix} \alpha & 1 \\ -\alpha & 1 \end{smallmatrix}\right]$ and $\alpha = e^{\pi i/4}$. Again by Theorem 6.1, it is easy to verify that $\#\mathrm{CSP}^2(YH^{-1}\mathcal{F} \cup \{[1, -i, 1]\})$ is #**P**-hard, and so is Holant($\mathcal{F}$), unless $\mathcal{F}$ is $\mathcal{A}$-transformable.

- Otherwise there exists $f \in \mathcal{F}_{nd}^{\geqslant 3}$ such that $f \in \mathcal{P}_2$. By Lemma 5.8, we may assume that $f = \left[\begin{smallmatrix} 1 \\ i \end{smallmatrix}\right]^{\otimes d} + \beta \left[\begin{smallmatrix} 1 \\ -i \end{smallmatrix}\right]^{\otimes d}$ where $\mathrm{arity}(f) = d$ and $\beta \neq 0$. We do a transformation by ZD, where $D = \begin{bmatrix} 1 & 0 \\ 0 & \beta^{1/d} \end{bmatrix}$,

$$\text{Holant}(\mathcal{F}) \equiv \text{Pl-Holant}\left(\neq_2 | =_d, (ZD)^{-1}\mathcal{F}\right).$$

With $\neq_2$ and $=_d$, we can realize any singature in $\mathcal{EQ}_d$ on the left as an E_d-block. (A detailed argument is provided at the beginning of this chapter.) Hence, we have that $\#\text{CSP}^d((ZD)^{-1}\mathcal{F}) \leqslant_T \text{Holant}(\mathcal{F})$. By Theorem 6.1, one can verify that $\#\text{CSP}^d((ZD)^{-1}\mathcal{F})$ is #**P**-hard, and so is $\text{Holant}(\mathcal{F})$, unless $\mathcal{F}$ is $\mathcal{A}$- or $\mathcal{P}$-transformable.

These are tractable Cases 2 and 3.

Now we may assume that $\emptyset \neq \mathcal{F}_{nd}^{\geqslant 3} \subseteq \mathcal{V}$ and $\text{Holant}(\mathcal{F})$ is not #**P**-hard. By Lemma 3.45 and Lemma 3.47, $\mathcal{F}_{nd}^{\geqslant 3}$ must be a subset of $\mathcal{V}^+$ or $\mathcal{V}^-$. Suppose $\mathcal{F}_{nd}^{\geqslant 3} \subseteq \mathcal{V}^+$ as the other case is similar. By Lemma 3.43, any non-degenerate binary signature in $\mathcal{F}$ has to be in $\mathcal{R}_2^+$. Now we have two cases.

- If there exists $f \in \mathcal{F} \subseteq \mathcal{V}^+$ such that $\text{rd}^+(f) \geqslant 2$, then by Lemma 3.41, the only unary signatures allowed in $\mathcal{F}$ are some multiples of $[1, i]$, and all degenerate signatures in $\mathcal{F}$ are some multiples of a tensor power of $[1, i]$. Thus, all non-degenerate signatures of arity at least 3 as well as all degenerate signatures belong to $\mathcal{V}^+$, and all non-degenerate binary signatures belong to $\mathcal{R}_2^+$. This is tractable Case 4.
- Otherwise $\mathcal{F}_{nd}^{\geqslant 3} \subseteq \mathcal{R}_2^+$. Since all non-degenerate binary signatures are also in $\mathcal{R}_2^+$, we have that all non-degenerate signatures in $\mathcal{F}$ are in $\mathcal{R}_2^+$. This is Case 5. □

Chapter 7

Anti-Ferromagnetic 2-Spin Systems

Starting from this chapter, we will show some results on approximate counting. Unlike sweeping dichotomy theorems for exact counting, such as Theorem 6.17, we have only delineated easy to approximate problems in very restricted settings. In the following several chapters, we will turn our attention to spin systems, which are well studied in the areas of Statistical Physics, Applied Probability and Computer Science as a general framework to model nearest-neighbour interactions in graphs. We will focus on 2-state spin systems, or 2-spin systems for short.

7.1 Definitions and Backgrounds

Let Σ be a finite alphabet. We want to approximate the value of a function $f : \Sigma^* \to \mathbb{R}$. A *randomized approximation scheme* is an algorithm that takes an instance $x \in \Sigma^*$ and a rational error tolerance $\varepsilon > 0$ as inputs, and outputs a rational number z such that, for every x and ε,

$$\Pr[e^{-\varepsilon} f(x) \leqslant z \leqslant e^{\varepsilon} f(x)] \geqslant \frac{3}{4}.$$

A *fully polynomial randomized approximation scheme* (FPRAS) is a randomized approximation scheme which runs in time bounded by a polynomial in $|x|$ and ε^{-1}. Note that the quantity $\frac{3}{4}$ can be changed to any value in the interval $\left(\frac{1}{2}, 1\right)$ or even $1 - 2^{-n^c}$ for a problem of size n without changing the set of problems that have fully polynomial randomized approximation schemes since the higher accuracy can be achieved with only polynomial delay by taking a majority vote of multiple samples.

Dyer *et al.* [DGGJ03] introduced the notion of approximation-preserving reductions. Suppose f and g are two functions from Σ^* to $\mathbb{R}$. An *approximation-preserving reduction* (AP-reduction) from f to g is a randomized algorithm $\mathcal{A}$ to approximate f using an oracle for g. The algorithm $\mathcal{A}$ takes an input $(x, \varepsilon) \in \Sigma^* \times (0, 1)$, and satisfies the following three conditions: (i) every oracle call made by $\mathcal{A}$ is of the form (y, δ), where $y \in \Sigma^*$ is an instance of g, and $0 < \delta < 1$ is an error bound satisfying $\delta^{-1} \leqslant \text{poly}(|x|, \varepsilon^{-1})$; (ii) the algorithm $\mathcal{A}$ meets the specification for being a randomized approximation scheme for f whenever the oracle meets the specification for being a randomized approximation scheme for g; (iii) the run-time of $\mathcal{A}$ is polynomial in $|x|$ and ε^{-1}.

If an AP-reduction from f to g exists, we write $f \leqslant_{\mathsf{AP}} g$, and say that f is *AP-reducible* to g. If $f \leqslant_{\mathsf{AP}} g$ and $g \leqslant_{\mathsf{AP}} f$, then we say that f and g are *AP-interreducible* or *AP-equivalent*, and write $f \equiv_{\mathsf{AP}} g$.

An instance of a 2-spin system is a graph $G = (V, E)$. A configuration σ assigns one of the two spins "0" and "1" to each vertex, that is, σ is one of the $2^{|V|}$ possible assignments $\sigma : V \to \{0, 1\}$. The local interaction along an edge is characterized by a matrix $\mathbf{A} = \left[\begin{smallmatrix} A_{0,0} & A_{0,1} \\ A_{1,0} & A_{1,1} \end{smallmatrix}\right]$, where $A_{i,j}$ is the local weight (or energy) when the two endpoints are assigned i and j respectively. Moreover, there is also an external field, specified by $\mathbf{b} = \left[\begin{smallmatrix} b_0 \\ b_1 \end{smallmatrix}\right]$ on each vertex, where b_i is the local weight when the vertex is assigned spin i. All parameters are non-negative. The total weight $w(\sigma)$ of a configuration σ is given by the following product

$$w(\sigma) = \prod_{(u,v) \in E} A_{\sigma(u), \sigma(v)} \prod_{v \in V} b_{\sigma(v)}.$$

We study symmetric edge interactions, that is, $A_{0,1} = A_{1,0}$. We normalize $\mathbf{A}$ and $\mathbf{b}$ so that $\mathbf{A} = \left[\begin{smallmatrix} \beta & 1 \\ 1 & \gamma \end{smallmatrix}\right]$ and $\mathbf{b} = \left[\begin{smallmatrix} \lambda \\ 1 \end{smallmatrix}\right]$. In this case, $w(\sigma)$ simplified into

$$w(\sigma) = \beta^{m_0(\sigma)} \gamma^{m_1(\sigma)} \lambda^{n_0(\sigma)} \tag{7.1}$$

where $m_0(\sigma)$ is the number of $(0, 0)$ edges given by the configuration σ, $m_1(\sigma)$ is the number of $(1, 1)$ edges, and $n_0(\sigma)$ is the number of vertices assigned 0. A 2-spin system is specified by these parameters $\beta, \gamma \geqslant 0$ and $\lambda > 0$. Two important special cases are the Ising model, where $\beta = \gamma$, and the hardcore gas model, where $\beta = 0$ and $\gamma = 1$.

We say a real number $z \neq 0$ is *efficiently approximable* if there is an FPRAS for the problem of computing z. We will always assume parameters β, γ, λ are efficiently approximable. Notice that in statistic physics literature, parameters are usually chosen to be the logarithms of our parameters above. Parameterizations do not affect the complexity of the system.

The Gibbs measure is a natural distribution in which each configuration σ is drawn with probability proportional to its weight, that is, $\Pr_{G;\beta,\gamma,\lambda}(\sigma) \sim w(\sigma)$. The normalizing factor of the Gibbs measure is called the partition function, defined by

$$Z_{\beta,\gamma,\lambda}(G) = \sum_{\sigma:V\to\{0,1\}} w(\sigma). \tag{7.2}$$

The partition function encodes rich information regarding the macroscopic behavior of the spin system. We consider the following computation problem.

Name #2SPIN(β, γ, λ)

Instance A graph $G = (V, E)$.

Output $Z_{\beta,\gamma,\lambda}(G)$.

#2SPIN$(\beta, \gamma, 1)$ is exactly the problem Holant $(\mathcal{EQ} \mid f)$ where f is a binary signature such that $M_f = \left[\begin{smallmatrix} \beta & 1 \\ 1 & \gamma \end{smallmatrix}\right]$. Thus by Theorem 4.2, #2SPIN$(\beta, \gamma, 1)$ is #**P**-hard unless $\beta = \gamma = 0$ or $\beta\gamma = 1$. Moreover, on Δ-regular graphs where $\Delta \geqslant 3$, #2SPIN(β, γ, λ) becomes Holant $([\lambda, 0, \cdots, 0, 1] \mid f)$. We do a holographic transformation $\left[\begin{smallmatrix} 1 & 0 \\ 0 & \lambda^{1/\Delta} \end{smallmatrix}\right]$ to normalize the left to $=_\Delta$. Again by Theorem 1.15, #2SPIN(β, γ, λ) is #**P**-hard to compute unless $\beta = \gamma = 0$ or $\beta\gamma = 1$.

Due to the apparent intractability of exact evaluation, much effort has focused on approximating #2SPIN(β, γ, λ). The system is called *ferromagnetic* if the edge interaction is attractive ($\beta\gamma > 1$), and *anti-ferromagnetic* if repulsive ($\beta\gamma < 1$). As it turns out, the behavior of the system depends heavily on this property.

For ferromagnetic systems, a seminal result by Jerrum and Sinclair [JS93] gave the first *fully polynomial-time randomized approximation scheme* (FPRAS) for the Ising model, that is, $\beta = \gamma > 1$. Their algorithm is based on Markov Chain Monte Carlo (MCMC) techniques. It was later generalized to other ferromagnetic cases by Goldberg et al. [GJP03], which was in turn improved by Liu et al. [LLZ14a]. For $\beta \leqslant \gamma$, the current best approximable bound is that $\lambda \leqslant \frac{\gamma}{\beta}$ [LLZ14a]. In Chapter 9, we will improve it to $\lambda \leqslant \left(\frac{\gamma}{\beta}\right)^{\Delta_0}$ when $\beta \leqslant 1$, where $\Delta_0 = \frac{\sqrt{\beta\gamma}}{\sqrt{\beta\gamma}-1}$. The other case of $\beta \geqslant \gamma$ is completely symmetric.

The very first step of aforementioned MCMC algorithms is to transform the spin system to the so-called "subgraph world", which is essentially a holographic transformation by $H_2 = \left[\begin{smallmatrix} 1 & 1 \\ 1 & -1 \end{smallmatrix}\right]$. For anti-ferromagnetic systems ($\beta\gamma < 1$), a transformation by H_2 will inevitably make signature entries negative, which prevents us to define a Markov chain properly. It is not clear how to generalize the "subgraph world" to the anti-ferromagnetic case.

A breakthrough result by Weitz [Wei06] gave the first *fully polynomial-time approximation scheme* (FPTAS) of the hardcore model up to the uniqueness threshold on graphs with degree bound Δ. This threshold characterizes the uniqueness of the Gibbs measure on infinite $(\Delta-1)$-regular trees. Weitz's algorithm introduces a completely different idea, called correlation decay. In fact, our aforementioned results for ferromagnetic systems are also based on correlation decay.

Formally, we define the problem #2SPIN(β, γ, λ) on bounded degree graphs.

Name #Δ-2SPIN(β, γ, λ)

Instance A graph $G = (V, E)$ with maximum degree Δ.

Output $Z_{\beta,\gamma,\lambda}(G)$.

Since Weitz's result [Wei06], it is widely believed that, for anti-ferromagnetic 2-spin systems, the phase transition of Gibbs measures on infinite regular trees coincides with the computational complexity transition of approximating partition functions. In this chapter, we will give positive answer of this conjecture on the algorithmic side.

On the hardness side, the inapproximability of partition functions has in fact been extensively studied. In [JS93], Jerrum and Sinclair showed that there is no FPRAS for #2SPIN(β, β, λ) where $0 < \beta < 1$ and $\lambda > 0$. This hardness result is later generalized to all $0 < \beta, \gamma \leqslant 1$ and slightly beyond [GJP03]. Using gadgets based on random regular bipartite graphs, the hardcore gas model, that is #Δ-2SPIN$(0, 1, \lambda)$, is showed to be inapproximable for a small interval of λ beyond the uniqueness threshold via a series of research [DFJ02, MWW09, Sly10]. Sly and

Sun [SS14] finally showed that beyond the uniqueness threshold, unless **NP** = **RP**, there is no FPRAS for the partition function of #Δ-2Spin(β, β, λ) where $0 < \beta < 1$ and $\lambda > 0$, as well as #Δ-2Spin$(0, 1, \lambda)$ where $\lambda > 0$. Note that the result of Sly and Sun [SS14] is proved independently by Galanis, Štefankovič and Vigoda [GŠV12] for #Δ-2Spin$(\beta, \beta, 1)$ where $0 < \beta < 1$ and $\lambda > 0$, as well as #Δ-2Spin$(0, 1, \lambda)$ where $\lambda > 0$, under the same assumption of beyond the uniqueness threshold.

On Δ-regular graphs, the system with parameters (β, γ, λ) satisfies the uniqueness condition if and only if the system $(\sqrt{\beta\gamma}, \sqrt{\beta\gamma}, \lambda(\beta/\gamma)^{\Delta/2})$ does. It is easy to see this through a diagonal transformation by $\begin{bmatrix} 1 & 0 \\ 0 & (\beta/\gamma)^{1/2} \end{bmatrix}$. Therefore hardness results by Sly and Sun [SS14] can be translated to all anti-ferromagnetic 2-spin systems. We should note that we are not aware of transformations of the same property when the graph is not regular. The aforementioned conjecture is thus almost confirmed. The only problem only is at the critical threshold.

Unlike the clear picture of anti-ferromagnetic 2-spin models, approximating #2Spin(β, γ, λ) where $\beta\gamma > 1$ is not completely understood. The complexity is at most as hard as approximating independent sets in bipartite graphs (#BIS) [GJ07], which is conjectured to have no approximation algorithms [DGGJ03]. As mentioned earlier, the celebrated Jerrum-Sinclair [JS93] chain for ferromagnetic Ising models is generalized to a broader range of parameters [GJP03, LLZ14a]. It is not likely that the rest region is all as hard as #BIS and some surprising hardness results have been shown recently [LLZ14a]. Their results are achieved via reduction from anti-ferromagnetic 2-spin systems in bipartite graphs, which were studied in [CGG+14]. In bipartite graphs, the uniqueness threshold turns out to be the correct approximability boundary again except for the special case of #2Spin$(\beta, \beta, 1)$. We will return to hardness results in Chapter 8.

The understanding of multi-spin systems, either ferromagnetic or anti-ferromagnetic is also much less complete. Hardness results are obtained regarding the special case of ferromagnetic Potts models [GJ12a] and below the first order phase transition threshold [GŠV14, GŠVY14]. Correlation decay and FPTAS are also studied [LY13], but the whole picture is still far from clear.

The idea of approximate counting via correlation decay is introduced independently by Weitz [Wei06] and Bandyopadhyay and Gamarnik [BG08]. Aside from results presented in this dissertation, other important examples include [GK07, BGK+07, LWZ14, LLL14, LLZ14b, LL15b, LL15a].

In this chapter, we will give algorithmic results regarding anti-ferromagnetic 2-spin systems. Throughout this chapter, we will assume that $\beta \leqslant \gamma$ as the two parameters are symmetric. With a slight abuse of notation, we call (β, γ, λ) *anti-ferromagnetic* if $0 \leqslant \beta \leqslant \gamma$, $\gamma > 0$, $\beta\gamma < 1$, and $\lambda > 0$. Due to the symmetry of β and γ, and the triviality of $\beta = \gamma = 0$, it in fact captures all nontrivial anti-ferromagnetic two-state spin systems.

7.2 The Self-Avoiding Walk Tree

We briefly describe Weitz's algorithm [Wei06] in bounded degree graphs. Our algorithms presented later will follow roughly the same paradigm.

The Gibbs measure defines a marginal distribution of spins for each vertex. Let p_v denote the probability of a vertex v colored blue. Since the system is self-reducible, #2SPIN(β, γ, λ) is equivalent to computing p_v [JVV86] (for details, see for example Lemma 7.15).

Let $\sigma_\Lambda \in \{0,1\}^\Lambda$ be a configuration of $\Lambda \subset V$. We call vertices in Λ *fixed* and other vertices *free*. We use $p_v^{\sigma_\Lambda}$ to denote the marginal probability of v being assigned "0" conditional on the configuration σ_Λ of Λ.

Suppose the instance is a tree T with root v. Let $R_T^{\sigma_\Lambda} := p_v^{\sigma_\Lambda}/(1 - p_v^{\sigma_\Lambda})$ be the ratio between the two probabilities that the root v is 0 and 1, while imposing some condition σ_Λ (with the convention that $R_T^{\sigma_\Lambda} = \infty$ when $p_v^{\sigma_\Lambda} = 1$). Suppose that v has d children $v_i, \ldots v_d$. Let T_i be the subtree with root v_i. Due to the independence of subtrees, it is straightforward to get the following recursion for calculating $R_T^{\sigma_\Lambda}$:

$$R_T^{\sigma_\Lambda} = F_d\left(R_{T_1}^{\sigma_\Lambda}, \ldots, R_{T_d}^{\sigma_\Lambda}\right), \tag{7.3}$$

where the function $F_d(x_1, \ldots, x_d)$ is defined as

$$F_d(x_1, \ldots, x_d) := \lambda \prod_{i=1}^{d} \frac{\beta x_i + 1}{x_i + \gamma}.$$

We allow x_i's to take the value ∞ as in that case the function F_d is clearly well defined. In general we use capital letters like $F, G, A, \ldots$ to denote multivariate functions, and small letters

$f, g, \alpha, \ldots$ to denote their symmetric versions. Here we define $f_d(x) := \lambda\left(\frac{\beta x+1}{x+\gamma}\right)^d$ to be the symmetric version of $F_d(\mathbf{x})$.

Let $G(V, E)$ be a graph. Similarly define that $R_{G,v}^{\sigma_\Lambda} := p_v^{\sigma_\Lambda}/(1 - p_v^{\sigma_\Lambda})$. In contrast to the case of trees, there is no easy recursion to calculate $R_{G,v}^{\sigma_\Lambda}$ for a general graph G. The reason is dependencies caused by cycles. Weitz [Wei06] reduced computing the marginal distribution of v in a general graph G to that in a tree, called the self-avoiding walk (SAW) tree, denoted by $T_{SAW}(G, v)$. To be specific, given a graph $G = (V, E)$ and a vertex $v \in V$, $T_{SAW}(G, v)$ is a tree with root v that enumerates all self-avoiding walks originating from v in G, with additional vertices closing cycles as leaves of the tree. Each vertex in the new vertex set V_{SAW} of $T_{SAW}(G, v)$ corresponds to a vertex in G, but a vertex in G may be mapped to more than one vertices in V_{SAW}. A boundary condition is imposed on leaves in V_{SAW} that close cycles. The imposed color of such leaves depends on whether the cycle is formed from a small vertex to a large vertex or conversely, where the ordering is arbitrarily chosen in G. Vertex sets $S \subset \Lambda \subset V$ are mapped to respectively $S_{SAW} \subset \Lambda_{SAW} \subset V_{SAW}$, and any configuration $\sigma_\Lambda \in \{0, 1\}^\Lambda$ is mapped to $\sigma_{\Lambda_{SAW}} \in \{0, 1\}^{\Lambda_{SAW}}$. With abuse of notations we may write $S = S_{SAW}$ and $\sigma_\Lambda = \sigma_{\Lambda_{SAW}}$ when no ambiguity is caused.

Theorem 7.1 (Theorem 3.1 of Weitz [Wei06])**.** *Let $G = (V, E)$ be a graph, $v \in V$, $\sigma_\Lambda \in \{0, 1\}^\Lambda$ be a configuration on $\Lambda \subset V$, and $S \subset V$. Let $T = T_{SAW}(G, v)$ be constructed as above. It holds that*

$$R_{G,v}^{\sigma_\Lambda} = R_T^{\sigma_\Lambda}.$$

Moreover, the maximum degree of T is at most the maximum degree of G, $\mathrm{dist}_G(v, S) = \mathrm{dist}_T(v, S_{SAW})$, and any neighborhood of v in T can be constructed in time proportional to the size of the neighborhood.

The SAW tree construction does not solve a #**P**-hard problem, since $T_{SAW}(G, v)$ is potentially exponential in size of G. For a polynomial time approximation algorithm, we may run the tree recursion within some polynomial size, or equivalently a logarithmic depth. At the boundary where we stop, we may plug in some randomly guessed values. The question is then how large is the error of our random guess. To guarantee the performance of the algorithm, we need the following notion of *strong spatial mixing*.

For a subset of vertices $\Lambda \subseteq V$, a configuration σ_Λ is *feasible* if there exists a $\sigma \in \{0,1\}^E$ with Gibbs measure $\rho(\sigma) > 0$ such that σ is consistent with σ_Λ on Λ. Notice that if $\beta, \gamma > 0$ then all configurations are feasible.

Definition 7.2. *A spin system on a family $\mathcal{G}$ of graphs is said to exhibit* strong spatial mixing *(SSM) if for any graph $G = (V, E) \in \mathcal{G}$, any $v \in V, \Lambda \subset V$ and any feasible $\sigma_\Lambda, \tau_\Lambda \in \{0,1\}^\Lambda$,*

$$|p_v^{\sigma_\Lambda} - p_v^{\tau_\Lambda}| \leqslant \exp(-\Omega(\mathrm{dist}(v, S))),$$

where $S \subset \Lambda$ is the subset on which σ_Λ and τ_Λ differ, and $\mathrm{dist}(v, S)$ is the shortest distance from v to any vertex in S.

The *weak spatial mixing* can be defined by measuring the decay with respect to $\mathrm{dist}(v, \Lambda)$ instead of $\mathrm{dist}(v, S)$. The spatial mixing property is also called correlation decay in Statistical Physics.

If SSM holds, then the error caused by early termination in $T_{SAW}(G, v)$ and random boundary values is only exponentially small in the depth. Hence our algorithm is an FPTAS. In a lot of cases, the existence of FPTASes then boils down to showing SSM holds.

The Uniqueness Condition

Recall that the Gibbs distribution is the distribution in which a configuration σ is drawn with probability

$$\Pr_{G;\beta,\gamma,\lambda}(\sigma) = \frac{w(\sigma)}{Z_{\beta,\gamma,\lambda}(G)}. \tag{7.4}$$

Let $\mathbb{T}_\Delta$ denote the infinite Δ-regular tree, also known as the *Bethe lattice* or the *Cayley tree*. A *Gibbs measure* on $\mathbb{T}_\Delta$ is a measure such that for any finite subtree $T \subset \mathbb{T}_\Delta$, the induced distribution on T conditioned on the outer boundary is the same as that given by (7.4). There may be one or more Gibbs measures (see, e.g., [Geo11] for more details). A Gibbs measure is called *translation-invariant* if it is invariant under all automorphisms of $\mathbb{T}_\Delta$, and is *semi-translation-invariant* if it is invariant under all parity-preserving automorphisms of $\mathbb{T}_\Delta$. In our context, the Gibbs measures that will be of interest are the two extremal semi-translation-invariant Gibbs measures corresponding to the all 1's and all 0's boundary conditions. These two measures are different in the non-uniqueness region of $\mathbb{T}_\Delta$.

If we pick an arbitrary vertex as the root of $\mathbb{T}_d$, then the root has d children and every other vertex has $d-1$ children. Notice that the difference between $\mathbb{T}_d$ and an infinite $(d-1)$-ary tree, denoted by $\widehat{\mathbb{T}}_{d-1}$, is only the degree of the root. We consider the uniqueness of Gibbs measures on $\mathbb{T}_d$. Due to the symmetric structure of $\mathbb{T}_d$, the standard recursion (7.3) thus becomes $R_v = f_{d-1}(R_{v_i})$ for any child v_i of v, where $f_d(x) = \lambda\left(\frac{\beta x+1}{x+\gamma}\right)^d$ is the symmetrized version of $F_d(\mathbf{x})$.

Notice that $f_d'(x) < 0$ for any $\beta\gamma < 1$ and $x > 0$. It implies that there exists a unique positive fixed point $\widehat{x}_d$ such that $\widehat{x}_d = f(\widehat{x}_d)$. Denote $\mathrm{Ctr}(\beta,\gamma,\lambda,d) := |f_d'(\widehat{x}_d)|$. It is straightforward to calculate that

$$\mathrm{Ctr}(\beta,\gamma,\lambda,d) = \frac{\lambda d(1-\beta\gamma)(\beta\widehat{x}_d+1)^{d-1}}{(\widehat{x}_d+\gamma)^{d+1}} = \frac{d(1-\beta\gamma)\widehat{x}_d}{(\beta\widehat{x}_d+1)(\widehat{x}_d+\gamma)}.$$

It is known [Kel85, Geo11, MSW07] that the Gibbs measure of anti-ferromagnetic two-state spin systems in $\mathbb{T}_d$ is unique if and only if

$$\mathrm{Ctr}(\beta,\gamma,\lambda,d-1) = |f_{d-1}'(\widehat{x}_{d-1})| \leqslant 1. \tag{7.5}$$

Roughly speaking this is because $|f_{d-1}'(\widehat{x}_{d-1})| \leqslant 1$ if and only if the dynamical system defined by $f_{d-1}(x)$ converges to its unique fixed point $\widehat{x}_{d-1}$. If $|f_{d-1}'(\widehat{x}_{d-1})| > 1$ then the dynamical system will eventually be oscillating between $x_1 \neq x_2$ such that $x_1 = f_{d-1}(x_2)$ and $x_2 = f_{d-1}(x_1)$. This motivates the following definition.

Definition 7.3. *Let* (β,γ,λ) *be anti-ferromagnetic and* $d \geqslant 1$ *be an integer. Then* (β,γ,λ,d) *satisfies the strict uniqueness condition, or* StrUnique(β,γ,λ,d) *holds, if and only if*

$$\mathrm{Ctr}(\beta,\gamma,\lambda,d) < 1$$

In particular, (β,γ,λ) *is called* universally strictly unique *if* StrUnique(β,γ,λ,d) *holds for any* *integer* $d \geqslant 1$*.*

See Figure 7.1 for the universally strictly uniqueness region of the $\lambda = 1$ plane.

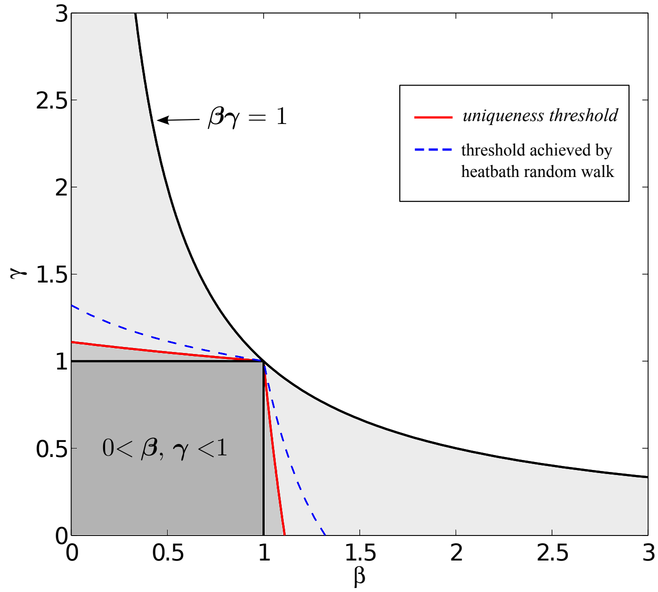


Figure 7.1: The universally strictly uniqueness region of the $\lambda = 1$ plane. In general graphs, our FPTAS works for the region between the red critical curve of the uniqueness threshold and the black curve $\beta\gamma = 1$. The heat-bath random walk in [GJP03] works for the region between the blue dashed line and $\beta\gamma = 1$.

Proposition 7.4. *For any integer* $d \geqslant 2$*, if* $|f'_{d-1}(\widehat{x}_{d-1})| > 1$*, then weak spatial mixing fails in* $\mathbb{T}_d$*, and therefore so does strong spatial mixing.*

We will discuss various thresholds of the uniqueness condition in Section 7.7.

At the end of this section we observe two properties of the universal strict uniqueness.

Lemma 7.5. *Let* (β, γ, λ) *be anti-ferromagnetic. If* (β, γ, λ) *is universally strictly unique then* $\gamma > 1$*.*

Proof. Suppose $\gamma \leqslant 1$. Suppose for the sake of contradiction that $\widehat{x}_d$ goes to 0 as d goes to infinity. It is easy to see that this violates $\widehat{x}_d = \lambda \left(\frac{\beta \widehat{x}_d + 1}{\widehat{x}_d + \gamma} \right)^d$. Hence there exists a increasing subsequence d_i such that $\widehat{x}_{d_i}$ is bounded away from 0 as i goes to ∞. Then $\left| f'_{d_i}(\widehat{x}_{d_i}) \right| > 1$ for sufficiently large i, which violates universally strictly uniqueness. $\square$

The uniqueness condition is defined by $|f'_d(\hat{x}_d)| < 1$. The following lemma shows that if (β, γ, λ) is universally strictly unique $|f'_d(\hat{x}_d)|$ is bounded away from 1. The same statement is obviously true if $\mathrm{StrUnique}(\beta, \gamma, \lambda, d)$ holds for finitely many d.

Lemma 7.6. *Let (β, γ, λ) be anti-ferromagnetic. If (β, γ, λ) is universally strictly unique, then $\alpha := \max_{d \geqslant 1}\{\mathrm{Ctr}(\beta, \gamma, \lambda, d)\} < 1$.*

Proof. If (β, γ, λ) is universally strictly unique, by Lemma 7.5, $\gamma > 1$. For anti-ferromagnetic (β, γ, λ), $\beta \leqslant \frac{1}{\gamma}$, thus the fixed point $\hat{x}_d = \lambda \left(\frac{\beta \hat{x}_d + 1}{\hat{x}_d + \gamma}\right)^d \leqslant \frac{\lambda}{\gamma^d}$. Therefore $\mathrm{Ctr}(\beta, \gamma, \lambda, d) = |f'_d(\hat{x}_d)| = \frac{d(1-\beta\gamma)\hat{x}_d}{(\beta \hat{x}_d + 1)(\hat{x}_d + \gamma)} \leqslant d\lambda\gamma^{-d}$.

Clearly there exists d_0 such that for any $d \geqslant d_0$, $d\lambda\gamma^{-d}$ is decreasing. Let $d_1 \geqslant d_0$ be the first such d so that $d_1\lambda\gamma^{-d_1} < 1$. For any $d \geqslant d_1$, we have that $\mathrm{Ctr}(\beta, \gamma, \lambda, d) \leqslant d\lambda\gamma^{-d} \leqslant d_1\lambda\gamma^{-d_1} < 1$. Let c be the larger of $\max_{1 \leqslant d \leqslant d_1}\{\mathrm{Ctr}(\beta, \gamma, \lambda, d)\}$ and $d_1\lambda\gamma^{-d_1}$. Clearly $\alpha \leqslant c$ and $c < 1$ due to universally strict uniqueness. □

7.3 The Potential Method

We would like prove the strong spatial mixing in arbitrary trees, sometimes with bounded degree Δ, under certain conditions. This is sufficient for approximation algorithms due to the self-avoiding walk tree construction. Our main technique in the analysis is the potential method.

To study correlation decay on trees, we use the standard recursion given in (7.3). Recall that T is a tree with root v. Vertices $v_1, \dots, v_d$ are d children of v, and T_i is the subtree rooted by v_i. A configuration σ_Λ is on a subset Λ of vertices, and R_T^σ denote the ratio of marginal probabilities at v given a partial configuration σ on T.

We want to study the influence of another set of vertices, say S, upon v. In particular, we want to study the range of ratios at v over all possible configurations on S. To this end, we define the lower and upper bounds as follows. Notice that as S will be fixed, we may assume that it is a subset of Λ.

Definition 7.7. *Let $T, v, \Lambda, \sigma_\Lambda, S, R_T^\sigma$ be as above. Define $R_v := \min_{\tau_\Lambda} R_T^{\tau_\Lambda}$ and $R^v := \max_{\tau_\Lambda} R_T^{\tau_\Lambda}$, where τ_Λ can only differ from σ_Λ on S. Define $\delta_v := R^v - R_v$.*

Our goal is thus to prove that $\delta_v \leqslant \exp(-\Omega(\mathrm{dist}(v,S)))$. We can recursively calculate R_v and R^v as follows. The base cases are:

1. $v \in S$, in which case $R_v = 0$ and $R^v = \infty$ and $\delta_v = \infty$;

2. $v \in \Lambda \setminus S$, i.e. v is fixed to be the same value in all τ_Λ, in which case $R_v = R^v = 0$ (or ∞) if v is fixed to be blue (or green), and $\delta_v = 0$;

3. $v \notin \Lambda$ and v is the only node of T, in which case $R_v = R^v = \lambda$ and $\delta_v = 0$.

For $v \notin \Lambda$, since F_d is monotonically decreasing with respect to any x_i for any anti-ferromagnetic (β, γ, λ),

$$R_v = F_d(R^{v_1}, ..., R^{v_d}) \text{ and } R^v = F_d(R_{v_1}, ..., R_{v_d}),$$

where R_{v_i} and R^{v_i} are recursively defined lower and upper bounds of $R_{T_i}^{\tau_\Lambda}$ for $1 \leqslant i \leqslant d$.

Our goal is to show that δ_v decays exponentially in the depth of the recursion when the uniqueness holds. A straightforward approach would be to prove that δ_v contracts by a constant ratio at each recursion step. This is a sufficient, but not necessary condition for the exponential decay. Indeed there are circumstances that δ_v does not necessarily decay in every step but does decay in the long run. To amortize this behaviour, we use a *potential* function $\Phi(x)$ and show that the correlation of a new recursion decays by a constant ratio.

To be more precise, the potential function $\Phi : \mathbb{R}^+ \to \mathbb{R}^+$ is a differentiable and monotonically increasing function. It maps the domain of the original recursion to a new one. Let $y_i = \Phi(x_i)$. We want to consider the recursion for y_i's. The new recursion function, which is the pullback of F_d, is defined as

$$G_d(y_1, \ldots, y_d) := \Phi(F_d(\Phi^{-1}(x_1), \ldots, \Phi^{-1}(x_d))).$$

The relationship between $F_d(\mathbf{x})$ and $G_d(\mathbf{y})$ is illustrated in Figure 7.2.

Morally we can choose whatever function as the potential function. However, we would like

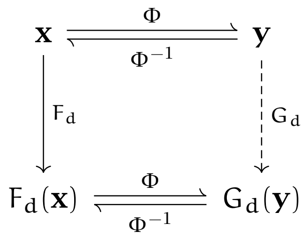


Figure 7.2: Commutative diagram between F_d and G_d.

to pick "good" ones so as to help the analysis of the contraction Define $\varphi(x) := \Phi'(x) > 0$ and

$$C_{\varphi,d}(\mathbf{x}) := \varphi(F_d(\mathbf{x})) \cdot \sum_{i=1}^{d} \left| \frac{\partial F_d}{\partial x_i} \right| \frac{1}{\varphi(x_i)}.$$

Definition 7.8. *Let $\Phi : \mathbb{R}^+ \to \mathbb{R}^+$ be a differentiable and monotonically increasing function. Let $\varphi(x)$ and $C_{\varphi,d}(\mathbf{x})$ be defined as above. Then $\Phi(x)$ is a* good potential function for degree d *if it satisfies the following conditions:*

1. *there exists a constant $C_1, C_2 > 0$ such that $C_1 \leqslant \varphi(x) \leqslant C_2$ for any $x \in [\lambda\beta^d, \lambda\gamma^{-d}]$ if $\beta\gamma < 1$ or $x \in [\lambda\gamma^{-d}, \lambda\beta^d]$ if $\beta\gamma > 1$;*

2. *there exists a constant $\alpha < 1$ such that $C_{\varphi,d}(\mathbf{x}) \leqslant \alpha$ for all $x_i \in [\lambda\beta^d, \lambda\gamma^{-d}]$ if $\beta\gamma < 1$ or $x_i \in [\lambda\gamma^{-d}, \lambda\beta^d]$ if $\beta\gamma > 1$.*

In Definition 7.8, Condition 1 is rather easy to satisfy. The crux is in fact Condition 2. We call α in Condition 2 the contraction ratio of $\Phi(x)$.

Lemma 7.9. *Let $\Phi(x)$ be a good potential function with contraction ratio α. Then $\alpha \geqslant \mathrm{Ctr}(\beta, \gamma, \lambda, d)$.*

Proof. Recall that $\mathrm{Ctr}(\beta, \gamma, \lambda, d) = \left| f'_d(\widehat{x}_d) \right|$, where $\widehat{x}_d$ is unique fixed point of the symmetrized recursion such that $f_d(\widehat{x}_d) = \widehat{x}_d$. By Definition 7.8, $\alpha \geqslant C_{\varphi,d}(\mathbf{x})$ for all $x_i \in [\lambda\beta^d, \lambda\gamma^{-d}]$.

Plugging in $x_i = \widehat{x}_d$ for all $1 \leqslant i \leqslant d$, we have that

$$\begin{aligned}
\alpha &\geqslant C_{\varphi,d}(\widehat{x}_d, \cdots, \widehat{x}_d) \\
&= \varphi(F_d(\widehat{x}_d, \ldots, \widehat{x}_d)) \cdot \sum_{i=1}^{d} \left| \frac{\partial F_d}{\partial x_i}(\widehat{x}_d, \ldots, \widehat{x}_d) \right| \frac{1}{\varphi(\widehat{x}_d)} \\
&= \varphi(f_d(\widehat{x}_d)) \cdot \sum_{i=1}^{d} \left| \frac{\partial F_d}{\partial x_i}(\widehat{x}_d, \ldots, \widehat{x}_d) \right| \frac{1}{\varphi(\widehat{x}_d)} \\
&= \varphi(\widehat{x}_d) \cdot \sum_{i=1}^{d} \left| \frac{\partial F_d}{\partial x_i}(\widehat{x}_d, \ldots, \widehat{x}_d) \right| \frac{1}{\varphi(\widehat{x}_d)} \\
&= \sum_{i=1}^{d} \left| \frac{\partial F_d}{\partial x_i}(\widehat{x}_d, \ldots, \widehat{x}_d) \right| \\
&= d \cdot \frac{(1-\beta\gamma) f_d(\widehat{x}_d)}{(\beta\widehat{x}_d + 1)(\widehat{x} + \gamma)} = \left| f_d'(\widehat{x}_d) \right| = \mathrm{Ctr}(\beta, \gamma, \lambda, d).
\end{aligned}$$

□

Recall Definition 7.3. We then have the following corollary.

Corollary 7.10. *A good potential function for degree* d *exists only if* $\mathrm{StrUnique}(\beta, \gamma, \lambda, d)$ *holds.*

We also define the upper and lower bounds of y. Define $y_v = \Phi(R_v)$ and accordingly $y_{v_i} = \Phi(R_{v_i})$, for $1 \leqslant i \leqslant d$, as well as $y^v = \Phi(R^v)$ and $y^{v_i} = \Phi(R^{v_i})$, for $1 \leqslant i \leqslant d$. We have that

$$y_v = G_d(y^{v_1}, \ldots, y^{v_d}) \text{ and } y^v = G_d(y_{v_1}, \ldots, y_{v_d}). \tag{7.6}$$

Let $\varepsilon_v = y^v - y_v$. For a good potential function, exponential decay of ε_v is sufficient to imply that of δ_v.

Lemma 7.11. *Let* $\Phi(x)$ *be a good potential function. Then there exists a constant* C *such that* $\delta_v \leqslant C\varepsilon_v$ *for any* $\mathrm{dist}(v, S) \geqslant 2$.

Proof. By (7.6) and the Mean Value Theorem, there exists an $\widetilde{R} \in [R_v, R^v]$ such that

$$\varepsilon_v = \Phi(R^v) - \Phi(R_v) = \Phi'(\widetilde{R}) \cdot \delta_v = \varphi(\widetilde{R}) \cdot \delta_v. \tag{7.7}$$

Since $\mathrm{dist}(v, S) \geqslant 2$, we have that $R_v \geqslant \lambda\beta^d$ and $R^v \leqslant \lambda\gamma^{-d}$. Hence $\widetilde{R} \in [\lambda\beta^d, \lambda\gamma^{-d}]$, and by Condition 1 of Definition 7.8, there exists a constant C_1 such that $\varphi(\widetilde{R}) \geqslant C_1$. Therefore

$\delta_\nu \leqslant 1/C_1 \varepsilon_\nu$. □

The next lemma explains Condition 2 of Definition 7.8.

Lemma 7.12. *Let* $\Phi(x)$ *be a good potential function with contraction ratio* α. *Then,*

$$\varepsilon_\nu \leqslant \alpha \max_{1 \leqslant i \leqslant d} \{\varepsilon_{\nu_i}\}.$$

Proof. First we use (7.6):

$$\varepsilon_\nu = y^\nu - y_\nu = G_d(y_{\nu_1}, \dots, y_{\nu_d}) - G_d(y^{\nu_1}, \dots, y^{\nu_d}).$$

Let $\mathbf{y}_0 = (y^{\nu_1}, \dots, y^{\nu_d})$ and $\mathbf{y}_1 = (y_{\nu_1}, \dots, y_{\nu_d})$. Let $\mathbf{z}(t) = t\mathbf{y}_1 + (1-t)\mathbf{y}_0$ be a linear combination of $\mathbf{y}_0$ and $\mathbf{y}_1$ where $t \in [0,1]$. Then we have that

$$\varepsilon_\nu = G_d(\mathbf{z}(1)) - G_d(\mathbf{z}(0)).$$

By the Mean Value Theorem, there exist $\tilde{t}$ such that $\varepsilon_\nu = \frac{d\, G_d(\mathbf{z}(t))}{d\, t}\Big|_{t=\tilde{t}}$. Let $\widetilde{y_i} = \tilde{t} y_{\nu_i} + (1-\tilde{t}) y^{\nu_i}$ for all $1 \leqslant i \leqslant d$. Then we have that

$$\varepsilon_\nu = \left| \nabla G_d(\widetilde{y_1}, \dots, \widetilde{y_d}) \cdot (\varepsilon_{\nu_1}, \dots, \varepsilon_{\nu_d}) \right|. \tag{7.8}$$

It is straightforward to calculate that

$$\frac{\partial G_d(\mathbf{y})}{\partial y_i} = \frac{\varphi(F_d(\mathbf{R}))}{\varphi(R_i)} \cdot \frac{\partial F_d(\mathbf{R})}{\partial R_i}, \tag{7.9}$$

where $R_i = \Phi^{-1}(y_i)$ and $\mathbf{y}$ and $\mathbf{R}$ are vectors composed by y_i's and R_i's. Plugging (7.9) into (7.8) we get that

$$\begin{aligned} \varepsilon_\nu &= \varphi(F_d(\widetilde{\mathbf{R}})) \cdot \sum_{i=1}^{d} \left| \frac{\partial F_d}{\partial R_i} \right| \frac{1}{\varphi(\widetilde{R_i})} \cdot \varepsilon_{\nu_i} \\ &\leqslant C_{\varphi,d}(\widetilde{R}_1, \dots, \widetilde{R}_d) \cdot \max_{1 \leqslant i \leqslant d} \{\varepsilon_{\nu_i}\} \leqslant \alpha \max_{1 \leqslant i \leqslant d} \{\varepsilon_{\nu_i}\}, \end{aligned}$$

where $\widetilde{R_i} = \Phi^{-1}(\widetilde{y_i})$, $\widetilde{\mathbf{R}}$ is the vector composed by $\widetilde{R_i}$'s, and in the last line we use Condition 2 of Definition 7.8. $\square$

Note that the two conditions of a good potential function does not necessarily deal with all cases in the tree recursion. At the root we have one more child than other vertices in a SAW tree. Also, if v has a child $u \in S$, then $\varepsilon_u = \infty$ and the range in both conditions of Definition 7.8 does not apply. To bound the recursion at the root, we have the following straightforward bound of the original recursion.

Lemma 7.13. *Let v be a vertex and v_i be its children for $1 \leqslant i \leqslant d$. Suppose $\delta_{v_i} \leqslant C_0$ for some $C_0 > 0$ and all $1 \leqslant i \leqslant d$. Then,*

$$\delta_v \leqslant d\lambda\gamma^{-d+1}C_0.$$

Proof. By the same argument as in Lemma 7.12 and (7.3), there exists x_i's such that

$$\begin{aligned}\delta_v &= \left|\nabla F_d(x_1,\dots,x_d)\cdot(\delta_{v_1},\dots,\delta_{v_d})\right| \\ &\leqslant C_0\sum_{i=1}^{d}\left|\frac{\partial F_d(\mathbf{x})}{\partial x_i}\right|,\end{aligned}$$

where $\mathbf{x}$ is the vector composed by x_i's. Then, we have that

$$\left|\frac{\partial F_d(\mathbf{x})}{\partial x_i}\right| = \frac{d(1-\beta\gamma)F_d(\mathbf{x})}{(x_i+\gamma)(\beta x_i+1)} \leqslant d\lambda\gamma^{-d+1},$$

where we use the fact that $F_d(\mathbf{x}) \leqslant \lambda\gamma^{-d}$ for any $x_i \in [0,\infty)$. $\square$

Lemma 7.14. *For a set of anti-ferromagnetic parameters (β,γ,λ) , if a good potential function $\Phi(x)$ exists for all integers $d \in [1,\Delta-1]$, then strong special mixing holds for any graphs with degree bound Δ.*

Proof. Let G be the graph with degree bound Δ and v be a vertex. We construct the SAW tree $T = T_{SAW}(G,v)$. Due to Theorem 7.1, we only need to show strong special mixing in T with respect to v and an arbitrary vertex set S. Let σ_Λ be a configuration on Λ where $S \subseteq \Lambda$. Let

δ_ν be defined as in Definition 7.7 with respect to T, ν, Λ, σ_Λ, and S. We want to show that $\delta_\nu = \exp(-\Omega(\mathrm{dist}(\nu, S)))$.

The maximum degree of T is at most Δ. Thus the root ν has at most Δ children in T, and any other vertex in T has at most $\Delta - 1$ children. Assume ν has $k \geqslant 1$ children as otherwise we are done. We may also assume that $\nu \notin S$ and let $t = \mathrm{dist}(\nu, S) - 1 \geqslant 1$. We recursively construct a path $u_0 = \nu, u_1, \ldots, u_l$ of length $l \leqslant t$ as follows. Given u_i, if there is no child of u_i, then we stop and let $l = i$. Otherwise u_i has at least one child. If $i = t$ then we stop and let $l = t$. Otherwise $l < t$ and let u_{i+1} be the child of u_i such that $\varepsilon_{u_{i+1}}$ takes the maximum ε among all children of u_i. In other words, by Lemma 7.12, we have that

$$\varepsilon_{u_i} \leqslant \alpha \varepsilon_{u_{i+1}}, \tag{7.10}$$

for all $1 \leqslant i \leqslant l - 1$. Notice that (7.10) may not hold for $i = 0$ since $\nu = u_0$ has possibly Δ children.

First we note that for all $1 \leqslant i \leqslant l$, $\mathrm{dist}(\nu, u_i) = i \leqslant l \leqslant t$, and therefore $u_i \notin S$. If we met any vertex u_l with no child, then we claim that $\varepsilon_{u_l} = 0$. This is because u_l is either a free vertex with no child or $u_l \in \Lambda$ but $u_l \notin S$. However since ε_{u_l} takes the maximum ε among all children of u_{l-1}, we have that for all children of u_{i-1}, $\varepsilon = 0$, which implies that $\varepsilon_{u_{i-1}} = 0$. Recursively we get that $\varepsilon_\nu = \varepsilon_{u_0} = 0$ and clearly the theorem holds by (7.7).

Hence we may assume that $l = t$. Since $u_l \notin S$, we have that $\delta_{u_l} \leqslant \lambda\gamma^{-(\Delta-1)}$ if $\gamma \leqslant 1$, or $\delta_{u_l} \leqslant \lambda$ if $\gamma > 1$. Hence by (7.7), we have that $\varepsilon_{u_l} \leqslant C_0$ for some constant C_0. Moreover applying (7.10) inductively we have that

$$\varepsilon_{u_1} \leqslant \alpha^l \varepsilon_{u_l} \leqslant \alpha^t C_0.$$

Hence by Lemma 7.11, we there exists another constant C_1 such that $\delta_{u_1} \leqslant \alpha^t C_1$. To get a bound on δ_{u_0}, we use Lemma 7.13, which states that

$$\delta_{u_0} \leqslant d_0 \lambda \gamma^{-d_0-1} \delta_{u_1} \leqslant d_0 \lambda \gamma^{-d_0-1} \alpha^t C_1,$$

where $d_0 \leqslant \Delta$ is the degree of $\nu = u_0$. Hence we have that $\delta_\nu = \exp(-\Omega(t))$ and the lemma holds. □

It has the following algorithmic implication.

Lemma 7.15. *Let* (β, γ, λ) *be a set of anti-ferromagnetic parameters. If there exists a good potential function for all* $d \in [1, \Delta - 1]$ *with contraction ratio* $\alpha < 1$*, then* #Δ-2SPIN(β, γ, λ) *can be approximated within* ε *in deterministic time* $O\left(n \left(\frac{n}{\varepsilon}\right)^{\frac{\log(\Delta-1)}{-\log \alpha}}\right)$*, where* n *is the number of vertices of the instance.*

Proof. Let G be a graph with degree bound Δ and v be a vertex in G. A self-avoiding walk tree $T = T_{SAW}(G, v)$ can be constructed so that $R_{G,v}^{\sigma_\Lambda} = R_T^{\sigma_\Lambda}$ by Theorem 7.1. We use the recursive procedure described above to compute upper and lower bounds of $R_T^{\sigma_\Lambda}$, with the base case that for any vertex u at level t that is not fixed, trivial bounds $R_u = 0$ and $R^u = \infty$ are used. In other words, we let S be the set of vertices whose distance to v is larger than t. Since a good potential function exists for all $d \in [1, \Delta - 1]$, by Lemma 7.14 the recursive procedure returns R_v and R^v such that $R_v \leqslant R_T^{\sigma_\Lambda} \leqslant R^v$, and $R^v - R_v = O(\alpha^t)$ where $\alpha < 1$ is the contraction ratio. Note that $R_T^{\sigma_\Lambda} = R_{G,v}^{\sigma_\Lambda} = \frac{p_v^{\sigma_\Lambda}}{1 - p_v^{\sigma_\Lambda}}$. Let $p_0 = \frac{R_v}{R_v + 1}$ and $p_1 = \frac{R^v}{R^v + 1}$. Then $p_0 \leqslant p_v^{\sigma_\Lambda} \leqslant p_1$ and

$$p_1 - p_0 = \frac{R^v}{R^v + 1} - \frac{R_v}{R_v + 1} \leqslant R^v - R_v = O(\alpha^t). \tag{7.11}$$

The recursive procedure runs in time $O(\Delta^t)$ since it only needs to construct the first t levels of the self-avoiding walk tree. For any $\varepsilon > 0$, let $t = O(\log_\alpha \varepsilon)$. As Δ is bounded, this gives an algorithm which approximates $p_v^{\sigma_\Lambda}$ within an additive error ε in time $O\left(\varepsilon^{\frac{\log(\Delta-1)}{\log \alpha}}\right)$.

Then we use self-reducibility to reduce computing $Z_{\beta,\gamma,\lambda}(G)$ to computing marginal probabilities with certain boudary conditions. To be specific, let σ be a configuration on a subset of V and τ be sampled according to the Gibbs measure. Let $p_v^\sigma := \Pr(\tau(v) = 1 \mid \sigma)$ be the conditional marginal probability. We can compute $Z_{\beta,\gamma,\lambda}(G)$ from p_v^σ by the following standard procedure. Let $v_1, \ldots, v_n$ enumerate vertices in G. For $0 \leqslant i \leqslant n$, let σ_i be the configuration fixing the first i vertices $v_1, \ldots, v_i$ as follows: $\sigma_i(v_j) = \sigma_{i-1}(v_j)$ for $1 \leqslant j \leqslant i - 1$ and $\sigma_i(v_i)$ is fixed to the spin s so that $p_i := \Pr(\tau(v_i) = s \mid \sigma_{i-1}) \geqslant 1/3$. This is always possible because clearly

$$\Pr(\tau(v_i) = 0 \mid \sigma_{i-1}) + \Pr(\tau(v_i) = 1 \mid \sigma_{i-1}) = 1.$$

In particular, $\sigma_n \in \{0,1\}^V$ is a configuration of V. The Gibbs measure of σ_n is $\rho(\sigma_n) = \frac{w(\sigma_n)}{Z_{\beta,\gamma,\lambda}(G)}$. On the other hand, we can rewrite $\rho(\sigma_n) = p_1 p_2 \cdots p_n$ by conditional probabilities. Thus $Z_{\beta,\gamma,\lambda}(G) = \frac{w(\sigma_n)}{p_1 p_2 \cdots p_n}$. The weight $w(\sigma_n) = \prod_{(u,v)\in E} A_{\sigma_n(u),\sigma_n(v)} \prod_{v \in V} b_{\sigma_n(v)}$ can be computed exactly in time polynomial in n. Note that p_i equals to either $p_{v_i}^{\sigma_{i-1}}$ or $1 - p_{v_i}^{\sigma_{i-1}}$. Since we can approximate $p_v^{\sigma_\Lambda}$ within an additive error ε in time $O\left(\varepsilon^{\frac{\log(\Delta-1)}{\log \alpha}}\right)$, the configurations σ_i can be efficiently constructed, which guarantees that all p_i's are bounded away from 0. Thus the product $p_1 p_2 \cdots p_n$ can be approximated within a factor of $(1 \pm n\varepsilon')$ in time $O\left(n\varepsilon'^{\frac{\log(\Delta-1)}{\log \alpha}}\right)$. Now let $\varepsilon' = \frac{\varepsilon}{n}$. We get the claimed FPTAS for $Z_{\beta,\gamma,\lambda}(G)$. □

When the degree is unbounded, there is a slight problem since the SAW tree may grow super polynomially even if the depth is of order $\log n$. We use a refined metric replacing the naive graph distance used in Definition 7.2. Strong spatial mixing under this metric is also called *computationally efficient correlation decay.*

Definition 7.16. *Let* T *be a rooted tree and* $M > 1$ *be a constant. For any vertex* v *in* T*, define the* M-based depth *of* v*, denoted* $\ell_M(v)$*, such that* $\ell_M(v) = 0$ *if* v *is the root, and* $\ell_M(v) = \ell_M(u) + \lceil \log_M(d+1) \rceil$ *if* v *is a child of* u *and* u *has degree* d*.*

We then define a slightly stronger notion of potential functions.

Definition 7.17. *Let* $\Phi : \mathbb{R}^+ \to \mathbb{R}^+$ *be a differentiable and monotonically increasing function. Let* $\varphi(x)$ *and* $C_{\varphi,d}(\mathbf{x})$ *defined in the same way as in Definition 7.8. Let* $\beta, \gamma > 0$ *be two parameters such that* $\beta\gamma < 1$ *and* $\beta > 1$ *or* $\gamma > 1$*. Then* $\Phi(x)$ *is a* universal potential function *for* (β, γ, λ) *if it satisfies the following conditions:*

1. *there exists a constant* $C_1 > 0$ *such that* $\varphi(x) > C_1$ *for any* $x \in (0, \lambda)$*;*
2. $\varphi(x)$ *is decreasing and there exists a constant* $C_2 > 0$ *such that* $\varphi(x) < C_2 x^{-1}$ *for any* $x \in (0, \lambda)$*;*
3. *there exists a constant* $\alpha < 1$ *and* $M > 1$ *such that for all* d*,* $C_{\varphi,d}(\mathbf{x}) \leqslant \alpha^{\lceil \log_M(d+1) \rceil}$ *for all* $x_i \in (0, \infty)$*;*

We also call α the contraction ratio and M the base.

Then we have the following analogue of Lemma 7.14. Note that the condition $\gamma > 1$ is necessary due to Lemma 7.5.

Lemma 7.18. *For a set of anti-ferromagnetic parameters* (β, γ, λ) *where* $\gamma > 1$*, if a good potential function* $\Phi(x)$ *exists for all integers* $d \in [1, \Delta - 1]$*, then strong spatial mixing holds for any graphs with degree bound* Δ*.*

Proof. The proof goes almost the same as in Lemma 7.14. Let G be a graph and v be a vertex. We construct the SAW tree $T = T_{SAW}(G, v)$. Let $t > 0$ be an integer which denotes the boundary distance. Let S be the set of vertices whose distance to v is larger than t. Since $\gamma > 1$, for any vertex v that is not in S, $R^v < \lambda$ and $R_v > 0$. Construct the path $v = u_0, u_1, \cdots, u_t$ as in Lemma 7.14.

For the base case of u_t, we use Condition 2 of Definition 7.17. Since $u_t \notin S$, then $\delta_{u_t} \leqslant \lambda\gamma^{-d_t}$, where d_t is the degree of u_t. Moreover, $R_{u_t} \geqslant \lambda\beta^{d_t}$. By (7.7) and Condition 2, we have that

$$\varepsilon_{u_t} = \delta_{u_t}\varphi(\widetilde{R}) \leqslant \lambda\gamma^{-d_t}\varphi(\lambda\beta^{d_t}) \leqslant C_2(\beta\gamma)^{-d_t} \leqslant C_2,$$

where $\widetilde{R} \in [R_{u_t}, R^{u_t}]$.

By Condition 3 and Lemma 7.12, we get stepwise decay until the last level, such that $\varepsilon_{u_1} \leqslant \alpha^t C_2$. By Condition 1 of Definition 7.17 and (7.7), we have that $\delta_{u_1} = \varepsilon_{u_1}/\varphi(\widetilde{R}') < \alpha^t C_2/C_1$, for some $\widetilde{R}' \in [R_{u_1}, R^{u_1}] \subset (0, \lambda)$. Then by Lemma 7.13, we have that

$$\delta_{u_0} \leqslant d_0\lambda\gamma^{-d_0-1}\alpha^t C_2/C_1,$$

where d_0 is the degree of $v = u_0$. Clearly there is a constant bound on $d_0\lambda\gamma^{-d_0-1}$ when $\gamma > 1$, and hence $\delta_{u_0} = \exp(\Omega(-t))$ holds. □

Lemma 7.19. *Let* (β, γ, λ) *be a set of anti-ferromagnetic parameters where* $\gamma > 1$*. If there exists a universal potential function* $\Phi(x)$ *with contraction ratio* α *and base* M*, then* $\#2\textsc{Spin}(\beta, \gamma, \lambda)$ *can be approximated within* ε *in deterministic time* $O\left(n^3\left(\frac{n}{\varepsilon}\right)^{\frac{\log M}{-\log\alpha}}\right)$*, where* n *is the number of vertices of the instance.*

Proof. By the same proof of Lemma 7.15, we only need to approximate the marginal probability at the root v of a tree T. By Condition 3 of Definition 7.17, there exists a constant $M > 1$ such that for all $d \geqslant M$, $C_{\varphi,d}(x_1, \cdots, x_d) < \alpha^{\lceil \log_M(d+1) \rceil}$. Denote by $B(\ell)$ the set of all vertices whose M-based depths is at most ℓ. It can be verified by induction that $|B(\ell)| \leqslant M^{\ell}$. Let $S = \{u \mid \mathrm{dist}(u, B(\ell)) > 1\}$, which is essentially the same S as in Lemma 7.14 and Lemma 7.18. We can recursively compute upper and lower bounds R^{v} and R_{v} of $R_T^{\sigma\Lambda}$ such that $R_v \leqslant R_T^{\sigma\Lambda} \leqslant R^{v}$, with the base case that for any vertex $u \in S$ trivial bounds $R_u = 0$ and $R^u = \infty$ are used.

We proceed as in the proof of Lemma 7.18. Without loss of generality, we construct a path $u_0 u_1 \cdots u_k$ in T from the root $u_0 = v$ to a u_k with $\ell_M(u_{k-1}) \leqslant \ell$ and $\ell_M(u_k) > \ell$. As in the proof of Lemma 7.12, $\varepsilon_{u_j} \leqslant C^{\varphi}_{d_j}(x_{j,1}, \ldots, x_{j,d_j}) \cdot \varepsilon_{u_{j+1}}$ for all $0 \leqslant j \leqslant k-1$, where d_j is the number of children of u_j and $x_{j,i} \in [0, \infty)$, $1 \leqslant i \leqslant d_j$. Hence we have that

$$\begin{aligned}\varepsilon_v &\leqslant \varepsilon_{u_k} \cdot \prod_{j=0}^{k-1} \alpha^{\lceil \log_M(d_j+1) \rceil} \leqslant \varepsilon_{u_k} \cdot \alpha^{\sum_{j=0}^{k-1} \lceil \log_M(d_j+1) \rceil} \\ &= \varepsilon_{u_k} \cdot \alpha^{\ell_M(u_k)} \leqslant \varepsilon_{u_k} \cdot \alpha^{\ell}.\end{aligned}$$

Note that $\mathrm{dist}(u_k, B(\ell)) = 1$ and hence $u_k \notin S$. So $\varepsilon_{u_k} \leqslant C_2$ as in Lemma 7.18. The rest of the proof is the same as that of Lemma 7.15. The running time has an extra n^2 factor since we need to go down two more levels (in the worst case) outside of $B(\ell)$. □

7.4 A Degree Dependent Potential Function

In this section, we show that there exists a good potential function for any degree d with contraction ratio $\mathrm{Ctr}(\beta, \gamma, \lambda, d)$. However, the potential function does depend on d, and hence it does not always satisfy the condition of Lemma 7.15, which requires a good potential function for all $d \leqslant \Delta - 1$. As a result, this potential function implies FPTASes only when $\gamma < 1$.

To find a good potential function, it all boils down to satisfying Condition 2 of Definition 7.8. Essentially, we want the function $C_{\varphi,d}(\mathbf{x})$ to take its maximum at $\mathbf{x} = \widehat{\mathbf{x}}_d := (\widehat{x}_d, \cdots, \widehat{x}_d)$, since $C_{\varphi,d}(\widehat{\mathbf{x}}_d) = \mathrm{Ctr}(\beta, \gamma, \lambda, d)$. We break it down into two promises. First, $C_{\varphi,d}(\mathbf{x})$ should be a concave function so that its maximum is achieved when all x_i's are equal. Second, $c_{\varphi,d}(x)$ should take its maximum when $x = \widehat{x}_d$, where $c_{\varphi,d}(x) = C_{\varphi,d}(x, x, \cdots, x)$ is the symmetrized version.

To satisfy the second condition, we would like $\widehat{x}_d$ to be the unique root to $c'_{\varphi,d}(x) = 0$. We will derive a potential function using this condition, and then go back to verify the rest. We will pick $\Phi(x)$ to be monotonically increasing, It is straightforward to calculate that

$$c_{\varphi,d}(x) = \frac{\varphi(f_d(x))(-f'_d(x))}{\varphi(x)}$$

and

$$\frac{c'_{\varphi,d}(x)}{c_{\varphi,d}(x)} = \frac{\varphi'(f_d(x))f'_d(x)}{\varphi(f_d(x))} + \frac{f''_d(x)}{f'_d(x)} - \frac{\varphi'(x)}{\varphi(x)}. \tag{7.12}$$

Moreover, by direct calculation,

$$\frac{f'_d(x)}{f_d(x)} = -\frac{d(1-\beta\gamma)}{(\beta x+1)(x+\gamma)} \tag{7.13}$$

and

$$\frac{f''_d(x)}{f'_d(x)} = \frac{f'_d(x)}{f_d(x)} - \frac{\beta}{\beta x+1} - \frac{1}{x+\gamma}. \tag{7.14}$$

Plugging (7.13) and (7.14) into (7.12), we get

$$\begin{aligned}(\beta x+1)(x+\gamma)\frac{c'_{\varphi,d}(x)}{c_{\varphi,d}(x)} =& - d(1-\beta\gamma)\left(\frac{\varphi'(f_d(x))f_d(x)}{\varphi(f_d(x))}+1\right)\\ &- \left(2\beta x+1+\gamma\beta+\frac{\varphi'(x)(\beta x+1)(x+\gamma)}{\varphi(x)}\right).\end{aligned} \tag{7.15}$$

Note that in the right hand side of (7.15), we have separated terms involving $f_d(x)$ from those involving x. Now let $p(y) := -d(1-\beta\gamma)\left(\frac{\varphi'(y)y}{\varphi(y)}+1\right)$ and $q(x) := 2\beta x+1+\gamma\beta+\frac{\varphi'(x)(\beta x+1)(x+\gamma)}{\varphi(x)}$. Hence we want $p(f_d(x)) = q(x)$ if and only if $x = \widehat{x}_d$, or in other words, $x = f_d(x)$. We rewrite $p(y)$ and $q(x)$ as

$$p(y) = -d(1-\beta\gamma)\frac{\varphi(y)+\varphi'(y)y}{\varphi(y)} = \frac{(-d(1-\beta\gamma)\varphi(y)y)'}{\varphi(y)}$$

and

$$q(x) = \frac{(2\beta x + 1 + \gamma\beta)\,\varphi(x) + \varphi'(x)(\beta x + 1)(x + \gamma)}{\varphi(x)}$$
$$= \frac{(\varphi(x)(\beta x + 1)(x + \gamma))'}{\varphi(x)}.$$

An obvious pick is to let $p(y)$ and $q(x)$ to be the same function in each own variables, so that $p(y) = q(x)$ when $x = y$. To ensure that, we would like to let $d(1-\beta\gamma)\varphi(x)x + \varphi(x)(\beta x+1)(x+\gamma) = C$ for some constant $C > 0$. The value of C does not matter here, and we pick $C = 1$. This leads to our choice of potential functions:

$$\varphi_d(x) := \frac{1}{d(1-\beta\gamma)x + (\beta x + 1)(x + \gamma)}. \tag{7.16}$$

Let $\Phi_d(x) := \int_0^x \varphi'_d(t)dt$. In fact, $\Phi_d(x)$ takes the form

$$\Phi_d(x) = \frac{\beta}{K_2 - K_1} \log \frac{x + K_1}{x + K_2}, \tag{7.17}$$

where $K_2 > K_1 > 0$ are the two roots of

$$K_1 + K_2 = \frac{d(1-\beta\gamma) + 1 + \beta\gamma}{\beta},$$
$$K_1 K_2 = \frac{\gamma}{\beta}.$$

It is easy to verify that $K_1, K_2 \in \mathbb{R}$ and $\Phi_d(x)$ is well defined. Note that $\varphi_d(x)$ is a rational function. Therefore we can factor out $f_d(x) - x$ from $p(f_d(x)) - q(x)$. We still need to verify the factor leftover is positive.

Lemma 7.20. *Let* (β, γ, λ) *be anti-ferromagnetic and* d *be an integer such that* $\mathrm{StrUnique}(\beta, \gamma, \lambda, d)$ *holds. Then* $C_{\varphi_d, d}(\mathbf{x}) \leqslant \mathrm{Ctr}(\beta, \gamma, \lambda, d)$ *for any* $x_i \geqslant 0$*, where* $\Phi_d(x)$ *is defined by* (7.17)*.*

Proof. We first claim that $C_{\varphi_d, d}(\mathbf{x}) \leqslant c_{\varphi_d, d}(\widehat{x})$, where $\mathbf{x} = (x_1, \cdots, x_d)$ and $\widehat{x}$ is the unique solution such that $f_d(\widehat{x}) = F_d(\mathbf{x})$. To show this, we do a change of variables. Let $h(x) := \log \frac{\beta x_i + 1}{x_i + \gamma}$, $z_i = h(x_i)$ for all $1 \leqslant i \leqslant d$, and $\widehat{z} = h(\widehat{x})$. Moreover, let $D(\mathbf{z}) := C_{\varphi_d, d}(h^{-1}(z_1), \cdots, h^{-1}(z_d))$ where $\mathbf{z} = (z_1, \cdots, z_d)$. Since $\sum_{i=1}^d z_i = d\widehat{z}$, we only need to show $D(\mathbf{z})$ is a concave function to prove the claim. Indeed, this can be directly verified since the Hessian matrix of $D(\mathbf{z})$ is negative semidefinite.

Next we show that $c_{\varphi_d,d}(x)$ takes its maximum at $\widehat{x}_d$ where $\widehat{x}_d = f_d(\widehat{x}_d)$. Let $y = f_d(x)$. Recall (7.12) and (7.15). The derivative is

$$\begin{aligned}
\frac{c'_{\varphi_d,d}(x)}{c_{\varphi_d,d}(x)} =& \frac{\varphi'_d(f_d(x))f'_d(x)}{\varphi_d(f_d(x))} + \frac{f''_d(x)}{f'_d(x)} - \frac{\varphi'_d(x)}{\varphi_d(x)} \\
=& \frac{1}{(\beta x+1)(x+\gamma)}\Bigg(- d(1-\beta\gamma)\left(\frac{\varphi'_d(y)y}{\varphi_d(y)} + 1\right) \\
& - \left(2\beta x + 1 + \gamma\beta + \frac{\varphi'_d(x)(\beta x+1)(x+\gamma)}{\varphi_d(x)}\right)\Bigg) \\
=& \frac{1}{(\beta x+1)(x+\gamma)}\left(\frac{d(1-\beta\gamma)(\beta y^2-\gamma)}{d(1-\beta\gamma)y + (\beta y+1)(y+\gamma)} - \frac{d(1-\beta\gamma)(\beta x^2-\gamma)}{d(1-\beta\gamma)x + (\beta x+1)(x+\gamma)}\right) \\
=& \frac{d(1-\beta\gamma)(y-x)}{(\beta x+1)(x+\gamma)} \cdot \frac{2\beta\gamma(x+y) + (\beta xy+\gamma)(d+1-\beta\gamma(d-1))}{(d(1-\beta\gamma)y + (\beta y+1)(y+\gamma))(d(1-\beta\gamma)x + (\beta x+1)(x+\gamma))}.
\end{aligned}$$

Note that $\beta\gamma < 1$ and hence $d+1 > \beta\gamma(d-1)$. Moreover, $f_d(x)$ is monotonically decreasing in x. If $x < \widehat{x}_d$, we have that $f_d(x) > x$ and $c'_{\varphi_d,d}(x) > 0$. Similarly, if $x > \widehat{x}_d$, $f_d(x) < x$ and $c'_{\varphi_d,d}(x) < 0$. Hence $c'_{\varphi_d,d}(x) > 0$ takes its maximum at $x = \widehat{x}_d$. It is easy to verify that $c_{\varphi_d,d}(\widehat{x}_d) = \mathrm{Ctr}(\beta,\gamma,\lambda,d)$. This finishes the proof. □

The potential function Φ_d achieves the optimal contraction ratio. However, to apply Lemma 7.15, we need a good potential function for all degrees $d \leqslant \Delta - 1$, where Δ is the degree bound. We may pick $\Phi_{\Delta-1}$, but it only works when $\gamma \leqslant 1$. This is because $\mathrm{Ctr}(\beta,\gamma,\lambda,d)$ is monotone if and only if $\gamma \leqslant 1$. We will see this in Section 7.6.

Lemma 7.21. *Let* (β,γ,λ) *be anti-ferromagnetic and* Δ *an integer such that* $\mathrm{StrUnique}(\beta,\gamma,\lambda,\Delta)$ *holds and* $\gamma \leqslant 1$*. Then* $\Phi_\Delta(x)$ *defined by* (7.17) *is a good potential function for all degrees* $d \in [1,\Delta]$ *with contraction ratio* $\mathrm{Ctr}(\beta,\gamma,\lambda,\Delta)$*.*

Proof. We verify the two conditions in Definition 7.8. Condition 1 for any $d \leqslant \Delta$ is straightforward. Lemma 7.20 implies Condition 2 for Δ. We are left to show that $C_{\varphi_\Delta,d}(\mathbf{x}) \leqslant \mathrm{Ctr}(\beta,\gamma,\lambda,\Delta)$ for all integers $1 \leqslant d < \Delta$.

Fix such a d. We will show that $C_{\varphi_\Delta,d}(\mathbf{x}) \leqslant \mathrm{Ctr}(\beta,\gamma,\lambda,\Delta)$ where $\mathbf{x} = \{x_1,\cdots,x_d\}$ and $x_i \geqslant 0$. Let $\rho = \frac{1-\gamma}{1-\beta} \geqslant 0$. Then $\frac{\beta\rho+1}{\rho+1} = 1$. By Lemma 7.20, we have that $C_{\varphi_\Delta,\Delta}(\mathbf{x},\rho,\cdots,\rho) \leqslant$

$\mathrm{Ctr}(\beta,\gamma,\lambda,\Delta)$, where we append $\mathbf{x}$ by $\Delta-d$ many ρ's. On the other hand, we see that $F_\Delta(\mathbf{x},\rho,\cdots,\rho) = F_d(\mathbf{x})$ and hence

$$\begin{aligned}
C_{\varphi_\Delta,\Delta}(\mathbf{x},\rho,\cdots,\rho) =& \varphi(F_\Delta(\mathbf{x},\rho,\cdots,\rho)) \cdot \sum_{i=1}^{\Delta} \left|\frac{\partial F_\Delta}{\partial x_i}(\mathbf{x},\rho,\cdots,\rho)\right| \frac{1}{\varphi(x_i)} \\
=& \varphi(F_d(\mathbf{x})) \cdot \sum_{i=1}^{d} \left|\frac{\partial F_\Delta}{\partial x_i}(\mathbf{x},\rho,\cdots,\rho)\right| \frac{1}{\varphi(x_i)} \\
&+ \varphi(F_d(\mathbf{x})) \cdot \sum_{i=d+1}^{\Delta} \left|\frac{\partial F_d}{\partial x_i}(\mathbf{x},\rho,\cdots,\rho)\right| \frac{1}{\varphi(x_i)} \\
\geqslant& \varphi(F_d(\mathbf{x})) \cdot \sum_{i=1}^{d} \left|\frac{\partial F_\Delta}{\partial x_i}(\mathbf{x},\rho,\cdots,\rho)\right| \frac{1}{\varphi(x_i)} \\
=& \varphi(F_d(\mathbf{x})) \cdot \sum_{i=1}^{d} \left|\frac{\partial F_d}{\partial x_i}(\mathbf{x})\right| \frac{1}{\varphi(x_i)},
\end{aligned}$$

where we use the fact that for $1 \leqslant i \leqslant d$,

$$\begin{aligned}
\frac{\partial F_\Delta}{\partial x_i}(\mathbf{x},\rho,\cdots,\rho) &= \frac{\beta\gamma-1}{(\beta x_i+1)(x_i+\gamma)} F_\Delta(\mathbf{x},\rho,\cdots,\rho) \\
&= \frac{\beta\gamma-1}{(\beta x_i+1)(x_i+\gamma)} F_d(\mathbf{x}) = \frac{\partial F_d}{\partial x_i}(\mathbf{x}).
\end{aligned}$$

□

Combining Lemma 7.21 with Lemma 7.15, we have the following Theorem.

Theorem 7.22. *Let* (β,γ,λ) *be anti-ferromagnetic and* Δ *an integer such that* $\mathrm{StrUnique}(\beta,\gamma,\lambda,\Delta-1)$ *holds and* $\gamma \leqslant 1$*. Then* $\#\Delta$-2SPIN(β,γ,λ) *can be approximated within additive error* ε *in deterministic time* $O\left(\left(\frac{n}{\varepsilon}\right)^{\frac{\log(\Delta-1)}{-\log\alpha}}\right)$*, where* n *is the number of vertices of the instance and* $\alpha = \mathrm{Ctr}(\beta,\gamma,\lambda,\Delta-1)$*.*

7.5 An Improved Potential Function

We will derive another potential function using a different approach. This potential function has the advantage to work for all $\beta\gamma < 1$, not restricted to $\gamma \leqslant 1$. In particular, it gives the approximation algorithm when (β,γ,λ) is universally strictly unique. On the other hand, the contraction ratio achieved is not optimal. If $\gamma \leqslant 1$, the contraction ratio is $\sqrt{\mathrm{Ctr}(\beta,\gamma,\lambda,\Delta-1)}$, worse than $\mathrm{Ctr}(\beta,\gamma,\lambda,\Delta-1)$ achieved by Φ_d introduced in Section 7.4, (7.17).

Choosing the Potential Function

We assume that the system is at the critical threshold of the uniqueness condition for a certain degree d, that is $\mathrm{Ctr}(\beta, \gamma, \lambda, d) = 1$, or equivalently $f'_d(\widehat{x}_d) = -1$, where $\widehat{x}_d = f_d(\widehat{x}_d)$ is the unique positive fixed point of the recursion $f_d(x)$. We then have the following two equations:

$$\widehat{x}_d = \lambda \left(\frac{\beta \widehat{x}_d + 1}{\widehat{x}_d + \gamma} \right)^d \quad \text{and} \quad \frac{d(1-\beta\gamma)\widehat{x}_d}{(\beta\widehat{x}_d + 1)(\widehat{x}_d + \gamma)} = 1. \tag{7.18}$$

We want to find a potential function to satisfy Condition 2 of Definition 7.8, which states that $C_{\varphi,d}(\mathbf{x}) \leqslant \alpha$ for some $\alpha < 1$. We might as well consider the symmetrized version, which is easier to analyze. Hence we want $c_{\varphi,d}(x) := \frac{\varphi(f_d(x))|f'_d(x)|}{\varphi(x)} \leqslant \alpha$. Again, we have that $c_{\varphi,d}(\widehat{x}_d) = 1$. Similar to the analysis in Section 7.4, $c_{\varphi,d}(x)$ should achieve its maximum at $x = \widehat{x}_d$. We therefore want that $c'_{\varphi,d}(\widehat{x}_d) = 0$, that is

$$\left(\frac{f'_d(x)\varphi(f_d(x))}{\varphi(x)} \right)' \Bigg|_{x=\widehat{x}_d} = 0.$$

We can rewrite the above equation as:

$$\begin{aligned}
&\left(\frac{f'_d(x)\varphi(f_d(x))}{\varphi(x)} \right)' \Bigg|_{x=\widehat{x}_d} = 0 \\
\Leftrightarrow\ & \left[f'_d(x)\varphi(f_d(x)) \right]' \varphi(x) \Big|_{x=\widehat{x}_d} = f'_d(x)\varphi(f(x))\varphi'(x)\big|_{x=\widehat{x}_d} \\
\Leftrightarrow\ & \left[f''_d(\widehat{x}_d)\varphi(f_d(\widehat{x}_d)) + f'_d(\widehat{x}_d)\varphi'(f_d(\widehat{x}_d))f'_d(\widehat{x}_d) \right] \varphi(\widehat{x}_d) = f'_d(\widehat{x}_d)\varphi(f_d(\widehat{x}_d))\varphi'(\widehat{x}_d) \\
\Leftrightarrow\ & f''_d(\widehat{x}_d)\varphi(\widehat{x}_d) + \varphi'(\widehat{x}_d) = -\varphi'(\widehat{x}_d) \\
\Leftrightarrow\ & -\frac{f''_d(\widehat{x}_d)}{2} = \frac{\varphi'(\widehat{x}_d)}{\varphi(\widehat{x}_d)} = (\log(\varphi(\widehat{x}_d)))',
\end{aligned}$$

where in the fourth line we use facts that $\widehat{x}_d = f_d(\widehat{x}_d)$ and $f'_d(\widehat{x}_d) = -1$. It all amounts to solve an equation

$$(\log(\varphi(\widehat{x}_d)))' = -\frac{f''_d(\widehat{x}_d)}{2}. \tag{7.19}$$

To solve (7.19), we need to calculate the second derivative of $f_d(x)$, which is,

$$f''_d(x) = \lambda d(\beta\gamma - 1)\frac{(\beta x + 1)^{d-2}}{(x+\gamma)^{d+2}} \cdot ((d-1)\beta(x+\gamma) - (d+1)(\beta x + 1)).$$

Using this expression, we are already able to solve the equation (7.19). However the solution is too complicated, and more importantly, it depends on the degree d. We want the potential function to work for various degrees, and it is better to be independent from d. Keeping this in mind, we use (7.18) to simplify the expression of $f''_d(x)$ at $x = \widehat{x}_d$:

$$\begin{aligned} f''_d(\widehat{x}_d) =& \lambda d(\beta\gamma - 1)\frac{(\beta\widehat{x}_d + 1)^{d-2}}{(\widehat{x}_d+\gamma)^{d+2}} \cdot ((d-1)\beta(\widehat{x}_d+\gamma) - (d+1)(\beta\widehat{x}_d + 1)) \\ =& \frac{(d+1)(\beta\widehat{x}_d+1) - (d-1)\beta(\widehat{x}_d+\gamma)}{(\beta\widehat{x}_d+1)(\widehat{x}_d+\gamma)} \\ =& \frac{d+1}{\widehat{x}_d+\gamma} - \frac{(d-1)\beta}{\beta\widehat{x}_d+1} = \frac{d(1-\beta\gamma)}{(\beta\widehat{x}_d+1)(\widehat{x}_d+\gamma)} + \frac{1}{\widehat{x}_d+\gamma} + \frac{\beta}{\beta\widehat{x}_d+1} \\ =& \frac{1}{\widehat{x}_d} + \frac{1}{\widehat{x}_d+\gamma} + \frac{\beta}{\beta\widehat{x}_d+1}. \end{aligned} \tag{7.20}$$

Plugging (7.20) into (7.19), we have that

$$(\log(\varphi(\widehat{x})))' = -\frac{1}{2}\left(\frac{1}{\widehat{x}_d} + \frac{1}{\widehat{x}_d+\gamma} + \frac{\beta}{\beta\widehat{x}_d+1}\right). \tag{7.21}$$

Now we make our guess and impose that (7.21) holds for all x. This gives us a differential equation, to which the solution is

$$\log(\varphi(x)) = -\frac{1}{2}\log(x(x+\gamma)(\beta x+1)) + C_1,$$

where C_1 is some arbitrary constant. Hence we get

$$\varphi(x) = \frac{C_2}{(x(\beta x+1)(x+\gamma))^{\frac{1}{2}}},$$

where $C_2 \neq 0$ is some arbitrary constants. We set $C_2 = 1$ and define

$$\varphi_*(x) := \frac{1}{(x(\beta x+1)(x+\gamma))^{\frac{1}{2}}}. \tag{7.22}$$

Let $\Phi_*(x) := \int_s^x \varphi_*(t)dt$ where $s > 0$ such that $\Phi_*'(x) = \varphi_*(x)$. Clearly $\Phi_*(x)$ is well defined for any $x \in [s, \infty)$. Since after at least one step of recursion, there is a lower bound on the range of R_ν, we will set s to be the lower bound. There is no elementary expression for $\Phi_*(x)$ in general. However, in Definition 7.8 and Definition 7.17, all it matters is $\varphi_*(x)$. This will be the potential function we choose in this section.

Verifying the Potential Function

We will verify that Φ_* given by (7.22) is a good potential function for all degrees $d \leqslant \Delta$ if for all $d \leqslant \Delta$, StrUnique$(\beta, \gamma, \lambda, d)$ holds. We will also verify that Φ_* is a universal potential function if (β, γ, λ) is universally strictly unique.

We first do some calculation:

$$
\begin{aligned}
C_{\varphi_*,d}(\mathbf{x}) &= -\varphi_*(F_d(\mathbf{x})) \cdot \sum_{i=1}^{d} \frac{\partial F_d}{\partial x_i} \frac{1}{\varphi_*(x_i)} \\
&= \frac{(1-\beta\gamma)\,(F_d(\mathbf{x}))^{\frac{1}{2}}}{(\beta F_d(\mathbf{x})+1)^{\frac{1}{2}}\,(F_d(\mathbf{x})+\gamma)^{\frac{1}{2}}} \cdot \sum_{i=1}^{d} \frac{x_i^{\frac{1}{2}}}{(\beta x_i+1)^{\frac{1}{2}}(x_i+\gamma)^{\frac{1}{2}}},
\end{aligned}
$$

where we used that $\frac{\partial F_d}{\partial x_i} = -\frac{F_d(\mathbf{x})(1-\beta\gamma)}{(\beta x_i+1)(x_i+\gamma)}$. It is easy to see that $C_{\varphi_*,d}(\mathbf{x}) > 0$ unless $x_i = 0$ for all $1 \leqslant i \leqslant d$.

Similar to the proof of Lemma 7.20, we will bound $C_{\varphi_*,d}(\mathbf{x})$ in two steps. The first step is to show that the symmetrized version dominates $C_{\varphi_*,d}(\mathbf{x})$, which is

$$
c_{\varphi_*,d}(x) = d(1-\beta\gamma) \cdot \left(\frac{x}{(\beta x+1)(x+\gamma)}\right)^{\frac{1}{2}} \cdot \left(\frac{f_d(x)}{(\beta f_d(x)+1)\,(f_d(x)+\gamma)}\right)^{\frac{1}{2}}.
$$

We will then show that $c_{\varphi_*,d}(x)$ takes its maximum at $\hat{x}_d$.

Lemma 7.23. *Let* (β, γ, λ) *be anti-ferromagnetic. For any integer* d *and any* $x_i \in [0, +\infty)$, $1 \leqslant i \leqslant d$, *there exists* $\widehat{x} \in [0, +\infty)$ *such that* $C_{\varphi_*,d}(\mathbf{x}) \leqslant c_{\varphi_*,d}(\widehat{x})$.

Proof. Let $z_i = \frac{\beta x_i+1}{x_i+\gamma}$. Then $z_i \in (\beta, \frac{1}{\gamma}]$ as $x_i \geqslant 0$, and $x_i = \frac{1-\gamma z_i}{z_i-\beta}$. Express $C_{\varphi_*,d}(\mathbf{x})$ in terms of

z_i's:

$$C_{\varphi_*,d}(\mathbf{x}) = \frac{\left(\lambda \prod_{i=1}^{d} z_i\right)^{\frac{1}{2}}}{\left(\beta\lambda \prod_{i=1}^{d} z_i + 1\right)^{\frac{1}{2}} \left(\lambda \prod_{i=1}^{d} z_i + \gamma\right)^{\frac{1}{2}}} \cdot \sum_{i=1}^{d} (z_i^{-1} - \gamma)^{\frac{1}{2}} (z_i - \beta)^{\frac{1}{2}}. \tag{7.23}$$

By the Cauchy-Schwarz inequality, we have

$$\begin{aligned} \sum_{i=1}^{d} (z_i^{-1} - \gamma)^{\frac{1}{2}} (z_i - \beta)^{\frac{1}{2}} &\leqslant d \left(\frac{1}{d} \sum_{i=1}^{d} (z_i^{-1} - \gamma)(z_i - \beta) \right)^{\frac{1}{2}} \\ &= d \left(1 + \beta\gamma - \frac{1}{d} \sum_{i=1}^{d} (z_i \gamma + \beta z_i^{-1}) \right)^{\frac{1}{2}}. \end{aligned} \tag{7.24}$$

The inequality of arithmetic and geometric means implies that

$$\begin{aligned} d \left(1 + \beta\gamma - \frac{1}{d} \sum_{i=1}^{d} (z_i \gamma + \beta z_i^{-1}) \right)^{\frac{1}{2}} &\leqslant d \left(1 + \beta\gamma - \gamma \left(\prod_{i=1}^{d} z_i \right)^{\frac{1}{d}} - \beta \left(\prod_{i=1}^{d} z_i \right)^{-\frac{1}{d}} \right)^{\frac{1}{2}} \\ &= d \left(1 + \beta\gamma - \gamma\widehat{z} - \beta\widehat{z}^{-1} \right)^{\frac{1}{2}}, \end{aligned} \tag{7.25}$$

where $\widehat{z} = \left(\prod_{i=1}^{d} z_i\right)^{\frac{1}{d}}$. Plugging (7.24) and (7.25) into (7.23) we have

$$\begin{aligned} C_{\varphi_*,d}(\mathbf{x}) &\leqslant \frac{(\lambda\widehat{z}^d)^{\frac{1}{2}} \cdot d(1 + \beta\gamma - \gamma\widehat{z} - \beta\widehat{z}^{-1})^{\frac{1}{2}}}{(\beta\lambda\widehat{z}^d + 1)^{\frac{1}{2}} (\lambda\widehat{z}^d + \gamma)^{\frac{1}{2}}} \\ &= d \cdot \left(\frac{\lambda\widehat{z}^d (\widehat{z}^{-1} - \gamma)(\widehat{z} - \beta)}{(\beta\lambda\widehat{z}^d + 1)(\lambda\widehat{z}^d + \gamma)} \right)^{\frac{1}{2}}. \end{aligned}$$

Let $\widehat{x} = \frac{1-\gamma\widehat{z}}{\widehat{z}-\beta}$, and therefore $\frac{\beta\widehat{x}+1}{\widehat{x}+\gamma} = \widehat{z}$. It is easy to see that $\widehat{z} \in (\beta, \frac{1}{\gamma}]$ as $z_i \in (\beta, \frac{1}{\gamma}]$ for any $1 \leqslant i \leqslant d$. Then $\widehat{x} \in [0, +\infty)$. By substituting $\frac{\beta\widehat{x}+1}{\widehat{x}+\gamma}$ with $\widehat{z}$, we have

$$\begin{aligned} C_{\varphi_*,d}(\mathbf{x}) &\leqslant d(1 - \beta\gamma) \cdot \left(\frac{\widehat{x}}{(\beta\widehat{x} + 1)(\widehat{x} + \gamma)} \right)^{\frac{1}{2}} \cdot \left(\frac{f_d(\widehat{x})}{(\beta f_d(\widehat{x}) + 1)\,(f_d(\widehat{x}) + \gamma)} \right)^{\frac{1}{2}} \\ &= c_{\varphi_*,d}(\widehat{x}). \end{aligned}$$

This finishes the proof. □

Lemma 7.24. *Let* (β,γ,λ) *be anti-ferromagnetic and* $\Delta \geqslant 2$ *be an integer or* $\Delta = \infty$*. If* StrUnique (β,γ,λ,d) *holds for any integer* $1 \leqslant d < \Delta$*, then* $\alpha := \max_{1\leqslant d<\Delta}\{\mathrm{Ctr}(\beta,\gamma,\lambda,d)\} < 1$ *and* $c_{\varphi_*,d}(x) \leqslant \alpha^{\frac{1}{2}}$ *for all* $x \geqslant 0$*.*

Proof. We first characterize the maximum of $c_{\varphi_*,d}(x)$. Let $c_{\varphi_*,d}(x) = d(1-\beta\gamma)h_d(x)^{\frac{1}{2}}$, where

$$h_d(x) := \frac{x}{(\beta x+1)(x+\gamma)} \cdot \frac{f_d(x)}{(\beta f_d(x)+1)(f_d(x)+\gamma)}.$$

Recall that $f_d(x) = \lambda\left(\frac{\beta x+1}{x+\gamma}\right)^d$. Then take the derivative of $c_{\varphi_*,d}(x)$ with respect to x,

$$c'_{\varphi_*,d}(x) = d(1-\beta\gamma)\cdot\frac{1}{2}h'_d(x)h_d(x)^{-\frac{1}{2}}.$$

The derivative of $h_d(x)$ is

$$h'_d(x) = \frac{f_d(x)\cdot d(1-\beta\gamma)x}{(\beta f_d(x)+1)(f_d(x)+\gamma)(\beta x+1)^2(x+\gamma)^2}\cdot\left(\frac{\gamma-\beta x^2}{d(1-\beta\gamma)x} - \frac{\gamma-\beta f_d(x)^2}{(\beta f_d(x)+1)(f_d(x)+\gamma)}\right).$$

As x ranges over $[0,\infty)$, the function $\frac{\gamma-\beta x^2}{d(1-\beta\gamma)x}$ is strictly decreasing in x and ranges from $+\infty$ to $-\infty$. On the other hand, the function $\frac{\gamma-\beta f_d(x)^2}{(\beta f_d(x)+1)(f_d(x)+\gamma)}$ is strictly increasing in x as $f_d(x)$ is decreasing in x, and it has a bounded range since $f_d(x)$ is bounded. Thus, the equation

$$\frac{\gamma-\beta x^2}{d(1-\beta\gamma)x} = \frac{\gamma-\beta f_d(x)^2}{(\beta f_d(x)+1)\left(f_d(x)+\gamma\right)}. \tag{7.26}$$

has a unique solution in $(0,\infty)$, denoted by x_d. In addition, it holds that

$$h'_d(x)\begin{cases} >0 & \text{if } 0 \leqslant x < x_d, \\ =0 & \text{if } x = x_d, \\ <0 & \text{if } x > x_d. \end{cases} \tag{7.27}$$

Since $h_d(x) > 0$ and $1-\beta\gamma > 0$, the sign of $c'_{\varphi_*,d}(x)$ is the same as that of $h'_d(x)$. Thus, for any integer d, $c_{\varphi_*,d}(x)$ achieves its maximum when $x = x_d$.

We define a new function $\alpha_d(x)$ that

$$\alpha_d(x) := \left(d(1-\beta\gamma) \cdot \frac{\gamma - \beta x^2}{(\beta x + 1)(x+\gamma)} \cdot \frac{f_d(x)}{\gamma - \beta f_d(x)^2} \right)^{\frac{1}{2}}.$$

Therefore, for all $x \geqslant 0$,

$$\begin{aligned} c_{\varphi_*,d}(x) &\leqslant c_{\varphi_*,d}(x_d) \\ &= d(1-\beta\gamma) \left(\frac{x}{(\beta x_d + 1)(x_d+\gamma)} \cdot \frac{f_d(x_d)}{(\beta f_d(x_d)+1)\,(f_d(x_d)+\gamma)} \right)^{\frac{1}{2}} \\ &= \alpha_d(x_d), \end{aligned} \tag{7.28}$$

where in the last equation we substituted $(\beta f_d(x_d)+1)\,(f_d(x_d)+\gamma)$ with $\frac{d(1-\beta\gamma)x_d(\gamma-\beta f_d(x_d)^2)}{\gamma-\beta x_d^2}$ by (7.26).

We claim that for any integer $1 \leqslant d < \Delta$,

$$\alpha_d(x_d) \leqslant \alpha_d(\widehat{x}_d), \tag{7.29}$$

where $\widehat{x}_d$ is the unique positive fixed point of $f_d(x)$, that is, $\widehat{x}_d = f_d(\widehat{x}_d)$. We will use the assumption that StrUnique(β,γ,λ,d) holds for any integer $1 \leqslant d < \Delta$.

To prove the claim, there are two cases depending on the ordering of $\widehat{x}_d$ and x_d. Observe that both $\frac{\gamma-\beta x^2}{(\beta x+1)(x+\gamma)}$ and $\frac{f_d(x)}{\gamma-\beta f_d(x)^2}$ are decreasing for any $x \geqslant 0$ as $f_d(x)$ is decreasing in x.

- Case 1: $\widehat{x}_d \leqslant x_d$. We would like to show that $\alpha_d(x)$ is decreasing in the range $x \in [\widehat{x}_d, x_d]$. Due to the observation above, it suffices to show that both $\frac{\gamma-\beta x^2}{(\beta x+1)(x+\gamma)}$ and $\frac{f_d(x)}{\gamma-\beta f_d(x)^2}$ are positive for $x \in [\widehat{x}_d, x_d]$.

  By (7.27), we have $h_d'(\widehat{x}_d) \geqslant 0$. Note that

$$h_d'(\widehat{x}_d) = \frac{d(1-\beta\gamma)(\gamma-\beta\widehat{x}_d^2)\widehat{x}_d^2}{(\beta\widehat{x}_d+1)^3(\widehat{x}_d+\gamma)^3} \cdot \left(\frac{1}{d(1-\beta\gamma)\widehat{x}_d} - \frac{1}{(\beta\widehat{x}_d+1)(\widehat{x}_d+\gamma)} \right).$$

  Because StrUnique(β,γ,λ,d) holds, we have that $|f_d'(\widehat{x}_d)| = \frac{d(1-\beta\gamma)\widehat{x}_d}{(\beta\widehat{x}_d+1)(\widehat{x}_d+\gamma)} < 1$, which implies that $\frac{1}{d(1-\beta\gamma)\widehat{x}_d} - \frac{1}{(\beta\widehat{x}_d+1)(\widehat{x}_d+\gamma)} > 0$. Therefore $h_d'(\widehat{x}_d) \geqslant 0$ implies that $\gamma - \beta\widehat{x}_d^2 \geqslant$

0. Since $f_d(x)$ is monotonically decreasing in x and $\widehat{x}_d$ is its fixed point, we have that

$$\gamma - \beta f_d(x_d)^2 \geqslant \gamma - \beta f_d(\widehat{x}_d)^2$$
$$= \gamma - \beta \widehat{x}_d^2 \geqslant 0.$$

Since x_d satisfies (7.26), $\gamma - \beta x_d^2$ and $\gamma - \beta f_d(x_d)^2$ must be simultaneously positive or negative. Thus it also holds that $\gamma - \beta x_d^2 \geqslant 0$. Then both $\frac{\gamma-\beta x^2}{(\beta x+1)(x+\gamma)}$ and $\frac{f_d(x)}{\gamma-\beta f_d(x)^2}$ are positive and monotonically decreasing in the range $x \in [\widehat{x}_d, x_d]$, and so is $\alpha_d(x)$. We conclude that $\alpha_d(x_d) \leqslant \alpha_d(\widehat{x}_d)$ as $\widehat{x}_d \leqslant x_d$ and (7.29) holds.

- Case 2: $\widehat{x}_d > x_d$. By a similar argument to the one above, it holds that

$$\gamma - \beta f_d(\widehat{x}_d)^2 = \gamma - \beta \widehat{x}_d^2 < 0, \gamma - \beta f_d(x_d)^2 < 0, \text{ and } \gamma - \beta x_d^2 < 0.$$

Thus both $\frac{\gamma-\beta x^2}{(\beta x+1)(x+\gamma)}$ and $\frac{f_d(x)}{\gamma-\beta f_d(x)^2}$ are negative and monotonically decreasing in $x \in [x_d, \widehat{x}_d]$. It implies that their product is positive and increasing in $x \in [x_d, \widehat{x}_d]$, and so is $\alpha_d(x)$. We conclude that $\alpha_d(x_d) \leqslant \alpha_d(\widehat{x}_d)$ as $\widehat{x}_d > x_d$ and (7.29) holds.

Combining (7.28) and (7.29), we have that for any $x \geqslant 0$,

$$c_{\varphi_*,d}(x) \leqslant c_{\varphi_*,d}(x_d) = \alpha_d(x_d) \leqslant \alpha_d(\widehat{x}_d)$$
$$= \sqrt{\frac{d(1-\beta\gamma)\widehat{x}_d}{(\beta\widehat{x}_d + 1)(\widehat{x}_d + \gamma)}} = (\mathrm{Ctr}(\beta, \gamma, \lambda, d))^{\frac{1}{2}}.$$

If Δ is finite, then $\alpha = \max_{1 \leqslant d < \Delta}\{\mathrm{Ctr}(\beta, \gamma, \lambda, d)\} < 1$ since $\mathrm{StrUnique}(\beta, \gamma, \lambda, d)$ holds for any integer $d \in [1, \Delta)$, and $c_{\varphi_*,d}(x) \leqslant \alpha^{\frac{1}{2}}$ for any integer $d \in [1, \Delta)$ and all $x \geqslant 0$.

Otherwise $\Delta = \infty$, and (β, γ, λ) is universally strict unique. By Lemma 7.6, we have that $\alpha = \max_{d \geqslant 1}\{\mathrm{Ctr}(\beta, \gamma, \lambda, d)\} < 1$. Also, $c_{\varphi_*,d}(x) \leqslant \alpha^{\frac{1}{2}}$ for any integer $d \in [1, \infty)$ and all $x \geqslant 0$. □

Lemma 7.23 and Lemma 7.24 together imply that Φ_* is a good potential function for all $d \in [1, \Delta)$ if $\mathrm{StrUnique}(\beta, \gamma, \lambda, d)$ holds for all $d \in [1, \Delta)$. Hence, we have the following theorem, which applies to a wider range of parameters than Theorem 7.22. Note that the running time is quadratic in that of Theorem 7.22.

Theorem 7.25. *Let* (β,γ,λ) *be anti-ferromagnetic and* Δ *an integer such that* $\mathrm{StrUnique}(\beta,\gamma,\lambda,d)$ *holds for all integers* $d \in [1,\Delta)$*. Then* $\#\Delta\text{-}2\textsc{Spin}(\beta,\gamma,\lambda)$ *can be approximated within additive error* ε *in deterministic time* $O\left(\left(\frac{n}{\varepsilon}\right)^{\frac{2\log(\Delta-1)}{-\log\alpha}}\right)$*, where* n *is the number of vertices of the instance and* $\alpha = \max_{1\leqslant d\leqslant \Delta-1} \mathrm{Ctr}(\beta,\gamma,\lambda,d)$.

In addition, the strength of Φ_* is that it also applies to cases without degree bound.

Theorem 7.26. *Let* (β,γ,λ) *be anti-ferromagnetic and universally strictly unique. Then* #2Spin (β,γ,λ) *can be approximated within additive error* ε *in deterministic time* $O\left(\left(\frac{n}{\varepsilon}\right)^{\frac{2\log M}{-\log\alpha}}\right)$*, where* n *is the number of vertices of the instance,* $\alpha = \max_{d\geqslant 1} \mathrm{Ctr}(\beta,\gamma,\lambda,d)$*, and* M *is a constant.*

Proof. We only need to verify that Φ_* is a universal potential function with contraction ratio α. Condition 1 of Definition 7.17 is straghtforward. Condition 2 clearly holds with $C_2 = 1$.

For Condition 3, by Lemma 7.23 and Lemma 7.24, $C_{\varphi_*,d}(\mathbf{x}) \leqslant \alpha$. Moreover, note that $\gamma > 1$ by Lemma 7.5 since (β,γ,λ) is universally strictly unique. Hence $F_d(\mathbf{x}) \leqslant \lambda\gamma^{-d}$ for any $x_i \in [0,\infty)$ and $\left(\frac{z}{(\beta z+1)(z+\gamma)}\right)^{\frac{1}{2}} \leqslant \frac{1}{1+\sqrt{\beta\gamma}} \leqslant 1$ for any $z \in [0,\infty)$. It implies that,

$$\begin{aligned} C_{\varphi_*,d}(\mathbf{x}) &= \frac{(1-\beta\gamma)\,(F_d(\mathbf{x}))^{\frac{1}{2}}}{(\beta F_d(\mathbf{x})+1)^{\frac{1}{2}}\,(F_d(\mathbf{x})+\gamma)^{\frac{1}{2}}} \cdot \sum_{i=1}^{d} \frac{x_i^{\frac{1}{2}}}{(\beta x_i+1)^{\frac{1}{2}}(x_i+\gamma)^{\frac{1}{2}}} \\ &\leqslant \frac{\lambda\gamma^{-\frac{d}{2}}}{\gamma^{\frac{1}{2}}} \cdot d = d\lambda\gamma^{-\frac{d+1}{2}}. \end{aligned}$$

Hence there exists an integer $M > 1$ such that for any integer $d < M$, $C_{\varphi_*,d}(\mathbf{x}) \leqslant \alpha \leqslant \alpha^{\lceil \log_M(d+1)\rceil}$, and for any $d \geqslant M$, $C_{\varphi_*,d}(\mathbf{x}) \leqslant d\lambda\gamma^{-\frac{d+1}{2}} \leqslant \alpha^{\lceil \log_M(d+1)\rceil}$. Condition 3 holds as well. □

7.6 Monotonicity of the Uniqueness

In this section, we study the monotonicity of the uniqueness with respect to the degree d. There are two cases, summarized as follows. Proofs are given in the two subsections following.

Theorem 7.27. *For an anti-ferromagnetic* 2*-spin system with parameters* (β,γ,λ) *, if* $\beta \leqslant 1$ *and* $\gamma \leqslant 1$*, then there exists a unique integer* $\Delta_c = \Delta_c(\beta,\gamma,\lambda)$ *such that* $\mathrm{StrUnique}(\beta,\gamma,\lambda,d)$ *holds for all integers* $1 \leqslant d < \Delta_c$*, and* $\mathrm{StrUnique}(\beta,\gamma,\lambda,d)$ *fails for all integers* $d \geqslant \Delta_c$.

Theorem 7.28. *For an anti-ferromagnetic* 2*-spin system with parameters* (β, γ, λ) *, if* $\beta > 1$ *or* $\gamma > 1$*, then either* (β, γ, λ) *is universally strictly unique, or there exists two critical integers* $\Delta_c \leqslant \overline{\Delta_c}$ *such that* $\text{StrUnique}(\beta, \gamma, \lambda, d)$ *holds for any integer* $d \in [1, \Delta_c)$ *or* $d \in (\overline{\Delta_c}, \infty)$*, and* $\text{StrUnique}(\beta, \gamma, \lambda, d)$ *fails for any integer* $d \in [\Delta_c, \overline{\Delta_c}]$*.*

Due to Theorem 7.28, if $\gamma > 1$, the uniqueness condition for some large Δ does not necessarily imply uniqueness for smaller $d < \Delta$. In fact, for a fixed set of anti-ferromagnetic parameters (β, γ, λ) with $\gamma > 1$, there always exists a large enough integer Δ_0, such that $\text{StrUnique}(\beta, \gamma, \lambda, \Delta)$ holds for any integer $\Delta \geqslant \Delta_0$, but apparently not all such parameters are universally strictly unique.

Recall Definition 7.3. Our proofs of Theorem 7.27 and Theorem 7.28 relies heavily on the analysis of $\text{Ctr}(\beta, \gamma, \lambda, d)$ as d varies. To simplify the notation, let

$$c(d) := \text{Ctr}(\beta, \gamma, \lambda, d) = |f_d'(\widehat{x}_d)| = \frac{d(1-\beta\gamma)\widehat{x}_d}{p(\widehat{x}_d)},$$

where $p(x) = (\beta x + 1)(x + \gamma)$ and $\widehat{x}_d$ is the unique fixed point of $f_d(x)$. Notice that $c(1) < 1$ as $(\beta x + 1)(x + \gamma) - (1 - \beta\gamma)x > 0$ for any $x > 0$. Hence $\text{StrUnique}(\beta, \gamma, \lambda, 1)$ always holds. In most analysis of the following two sections we will treat d as a real positive parameter. Also notice that $c(d)$ actually depends on (β, γ, λ) as well but we will focus on its dependence on d in this section.

We always assume that $0 \leqslant \beta \leqslant \gamma$, $\gamma > 0$, $\beta\gamma < 1$, and $\lambda > 0$. It turns out that if $0 \leqslant \beta \leqslant \gamma \leqslant 1$, then $c(d)$ is monotone increasing in d; otherwise $c(d)$ is a single-peaked function in d, and there exists a unique maximum point. Moreover, $c(d)$ is increasing before its maxima and decreasing afterwards. Figure 7.3 illustrates two examples.

Take the derivative of $c(d)$ against d, we get

$$c'(d) = (1 - \beta\gamma)\left(\frac{\widehat{x}_d}{p(\widehat{x}_d)} + d \cdot \frac{\partial \widehat{x}_d}{\partial d} \frac{p(\widehat{x}_d) - \widehat{x}_d p'(\widehat{x}_d)}{p^2(\widehat{x}_d)}\right) \tag{7.30}$$

However, $\widehat{x}_d$ satisfies that $\widehat{x}_d = f(\widehat{x}_d)$, that is:

$$\widehat{x}_d = \lambda\left(\frac{\beta\widehat{x}_d + 1}{\widehat{x}_d + \gamma}\right)^d.$$

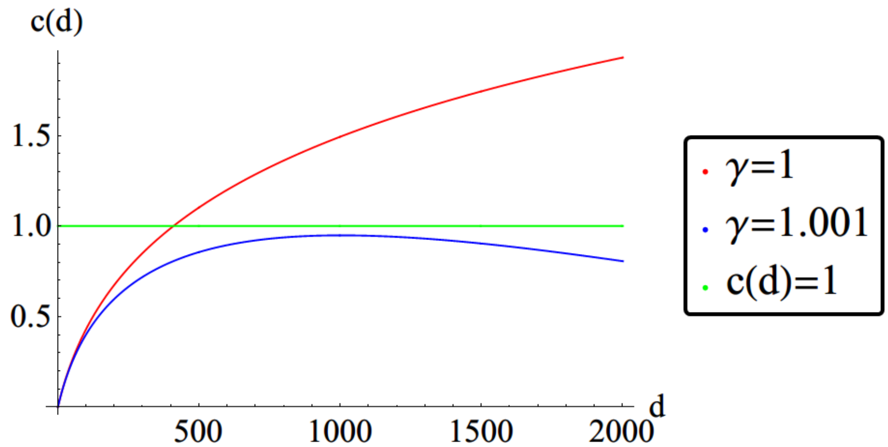


Figure 7.3: The function $c(d) = |f'_d(\hat{x}_d)|$ with argument d. The green line is the threshold $c(d) = 1$. For the other two curves, we fix $\beta = 0$ and $\lambda = 150$. The red curve above is for $\gamma = 1$ and the blue one below is for $\gamma = 1.001$. The red curve is monotone whereas the blue one has a unique maximum. Moreover, $(0, 1.001, 150)$ is universally strictly unique, while $\mathrm{StrUnique}(0, 1, 150, d)$ holds only for integer $d \in [1, 409]$.

Let $q(x) := \frac{\beta x+1}{x+\gamma}$. As $\hat{x}_d > 0$, we take logarithm on both sides,

$$\log \hat{x}_d = \log \lambda + d \log q(\hat{x}_d),$$

and then take the partial derivative with respect to d,

$$\frac{1}{\hat{x}_d} \cdot \frac{\partial \hat{x}_d}{\partial d} = \log q(\hat{x}_d) + d \frac{\partial \hat{x}_d}{\partial d} \cdot \frac{\beta\gamma - 1}{p(\hat{x}_d)}.$$

Hence, we have:

$$\frac{\partial \hat{x}_d}{\partial d} = \log q(\hat{x}_d) \cdot \frac{p(\hat{x}_d)\hat{x}_d}{p(\hat{x}_d) + d(1-\beta\gamma)\hat{x}_d} \tag{7.31}$$

Using (7.31) we substitute $\frac{\partial \hat{x}_d}{\partial d}$ in (7.30) and get

$$\begin{aligned} c'(d) &= (1-\beta\gamma)\left(\frac{\hat{x}_d}{p(\hat{x}_d)} + d \log q(\hat{x}_d) \cdot \frac{p(\hat{x}_d)\hat{x}_d}{p(\hat{x}_d) + d(1-\beta\gamma)\hat{x}_d} \cdot \frac{p(\hat{x}_d) - \hat{x}_d p'(\hat{x}_d)}{p^2(\hat{x}_d)}\right) \\ &= \frac{(1-\beta\gamma)\hat{x}_d}{p(\hat{x}_d)}\left(1 - d \log q(\hat{x}_d) \cdot \frac{\beta\hat{x}_d^2 - \gamma}{p(\hat{x}_d) + d(1-\beta\gamma)\hat{x}_d}\right). \end{aligned}$$

Since $\frac{\widehat{x}_d}{p(\widehat{x}_d)} > 0$ for any $\widehat{x}_d > 0$, we will focus on the sign inside the parentheses. We define the following function:

$$s(d) := 1 - d \log q(\widehat{x}_d) \frac{\beta \widehat{x}_d^2 - \gamma}{p(\widehat{x}_d) + d(1 - \beta\gamma)\widehat{x}_d}.$$

Hence the sign of $c'(d)$ will be the same as that of $s(d)$. We will show that if $\gamma \leqslant 1$, then $s(d) > 0$ for all $d > 0$; otherwise $\gamma > 1$, then there exists a unique $d_c = d_c(\beta, \gamma, \lambda)$ such that the sign of $s(d)$ is the same as that of $(d_c - d)$.

The proof of Theorem 7.27

In this section we will establish the monotonicity of the case $0 \leqslant \beta \leqslant \gamma \leqslant 1$. We have the following lemma.

Lemma 7.29. *Let* (β, γ, λ) *be anti-ferromagnetic. If* $\gamma \leqslant 1$*, then* $c(d)$ *is strictly increasing in* d *for all* $d > 0$*.*

Proof. As discussed above it is enough to show that if $\gamma \leqslant 1$, then $s(d) > 0$ and therefore $c'(d) > 0$ for all $d > 0$. We discuss cases based on the sign of $\beta \widehat{x}_d^2 - \gamma$.

1. $\beta \widehat{x}_d^2 - \gamma = 0$, then $s(d) = 1$ and the lemma holds.

2. $\beta \widehat{x}_d^2 - \gamma > 0$. This implies that $\widehat{x}_d > \frac{1-\gamma}{1-\beta}$, because the function $\beta x^2 - \gamma$ is increasing in x and $\beta \left(\frac{1-\gamma}{1-\beta}\right)^2 - \gamma = \frac{(\beta-\gamma)(1-\gamma\beta)}{(1-\beta)^2} \leqslant 0$. Also notice that $q(x)$ is decreasing in x, and $q\left(\frac{1-\gamma}{1-\beta}\right) = 1$. It implies that in this case $q(\widehat{x}_d) < 1$. So we have

$$d \log q(\widehat{x}_d) \frac{\beta \widehat{x}_d^2 - \gamma}{p(\widehat{x}_d) + d(1 - \beta\gamma)\widehat{x}_d} < 0$$

   and hence $s(d) > 1 > 0$.

3. $\beta \widehat{x}_d^2 - \gamma < 0$. To show $s(d) > 0$ we only need to show that

$$\frac{p(\widehat{x}_d) + d(1 - \gamma\beta)\widehat{x}_d}{\beta \widehat{x}_d^2 - \gamma} > d \log q(\widehat{x}_d).$$

Note that $\beta \leqslant \gamma \leqslant 1$, and thus $d \log q(\widehat{x}_d) \leqslant d \log \frac{q(\widehat{x}_d)}{\beta}$. We apply the inequality $\log x \leqslant x - 1$ on $\log \frac{q(\widehat{x}_d)}{\beta}$ and get

$$\begin{aligned} d \log q(\widehat{x}_d) &\leqslant d \log \frac{q(\widehat{x}_d)}{\beta} \\ &\leqslant d\left(\frac{1}{\beta} \cdot \frac{\beta \widehat{x}_d + 1}{\widehat{x}_d + \gamma} - 1\right) = \frac{d(1-\beta\gamma)}{\beta(\widehat{x}_d + \gamma)} = \frac{d(1-\beta\gamma)\widehat{x}_d}{\beta \widehat{x}_d^2 + \gamma \widehat{x}_d} \\ &\leqslant \frac{d(1-\beta\gamma)\widehat{x}_d}{\beta \widehat{x}_d^2 - \gamma} < \frac{p(\widehat{x}_d) + d(1-\gamma\beta)\widehat{x}_d}{\beta \widehat{x}_d^2 - \gamma}. \end{aligned}$$

It implies that $s(d) > 0$.

To sum up, we always have that $s(d) > 0$ and hence $c'(d) > 0$. □

Moreover, we show that $c(d)$ has no upper bound if $\gamma \leqslant 1$ as d goes to infinity.

Lemma 7.30. *Let (β, γ, λ) be anti-ferromagnetic. If $\gamma \leqslant 1$, then $c(d)$ goes to ∞ as d goes to ∞.*

Proof. First we claim that as d goes to infinity, $\widehat{x}_d$ is bounded away from ∞ and 0. Note that $\widehat{x}_d = \lambda \left(\frac{\beta \widehat{x}_d + 1}{\widehat{x}_d + \gamma}\right)^d$. Assume otherwise there is a subsequence of $\widehat{x}_d$ that goes to infinity, then the left hand side goes to infinity, while the right goes to $\lambda\beta^d \to 0$ or λ. Contradiction. Similarly if there is a subsequence of $\widehat{x}_d$ goes to 0, then the left hand side goes to 0, while the right goes to $\frac{\lambda}{\gamma^d} \to \infty$ or $\lambda > 0$. Also contradiction.

Recall that $c(d) = \frac{d(1-\beta\gamma)\widehat{x}_d}{p(\widehat{x}_d)} = \frac{d(1-\beta\gamma)\widehat{x}_d}{\beta \widehat{x}_d^2 + (1+\beta\gamma)\widehat{x}_d + \gamma}$. Since $\widehat{x}_d$ is bounded away from infinity and 0, it follows that $\frac{\widehat{x}_d}{\beta \widehat{x}_d^2 + (1+\beta\gamma)\widehat{x}_d + \gamma \widehat{x}_d}$ is bounded away from 0. Then $c(d)$ must go to infinity as d goes. □

Now we are ready to prove Theorem 7.27.

Proof of Theorem 7.27. Recall that $c(1) < 1$. By Lemma 7.29 and Lemma 7.30, there must exist a unique integer $\Delta_c \geqslant 2$ such that $c(\Delta_c) \geqslant 1$ and $c(\Delta_c - 1) < 1$. Again by Lemma 7.29, for any integer $1 \leqslant d < \Delta_c$, $c(d) < 1$ and for any integer $d \geqslant \Delta_c$, $c(d) \geqslant 1$. Since StrUnique$(\beta, \gamma, \lambda, d)$ holds if and only if $c(d) < 1$, the theorem follows. □

The proof of Theorem 7.28

In this section we deal with the case that (β, γ) is not in the unit square, that is, $\gamma > 1$. We will show that there exists a unique d_0 such that $c(d)$ takes its maximum at $d = d_0$, and $c(d)$ is increasing if $d < d_0$ and decreasing if $d > d_0$.

First, we claim that $\widehat{x}_d$ is decreasing in d under our assumption. Recall that $q(\widehat{x}_d) = \frac{\beta\widehat{x}_d+1}{\widehat{x}_d+\gamma} < 1$ under our assumption. Therefore, by (7.31),

$$\frac{\partial \widehat{x}_d}{\partial d} = \log q(\widehat{x}_d) \cdot \frac{p(\widehat{x}_d)\widehat{x}_d}{p(\widehat{x}_d) + d(1-\beta\gamma)\widehat{x}_d} < 0.$$

Moreover, as d goes to ∞, $\widehat{x}_d$ goes to 0 since $\widehat{x}_d = \lambda\left(\frac{\beta\widehat{x}_d+1}{\widehat{x}_d+\gamma}\right)^d < \frac{\lambda}{\gamma^d}$.

Define a new function

$$r(x) := \log q(x)(\beta x^2 - \gamma) - (1-\beta\gamma)x.$$

We will see the use of $r(x)$ in the proof of Lemma 7.32 and need the following technical lemma.

Lemma 7.31. *For* $x \in [0, \infty)$*, if* $\gamma > 1 > \beta \geqslant 0$*, then there exists a unique* x_c *such that* $r(x_c) = 0$*. Moreover* $r(x) > 0$ *when* $x < x_c$ *and* $r(x) < 0$ *when* $x > x_c$*.*

Proof. Clearly $r(x)$ is continuous for $x \geqslant 0$. It is easy to calculate that

$$r'(x) = 2\beta x \log q(x) - (1-\beta\gamma)\frac{x(2\beta x + (1+\beta\gamma))}{p(x)}.$$

Since $\log q(x) < -\log\gamma < 0$, $r'(x) < 0$ for all $x \geqslant 0$. Moreover, $r(0) = r\log r > 0$ and $r(x) \to -\infty$ as $x \to \infty$. There must exist x_c such that $r(x_c) = 0$, $r(x) > 0$ if $x < x_c$, and $r(x) < 0$ if $x > x_c$. □

Now we are ready to show the key lemma of this section.

Lemma 7.32. *Let* (β, γ, λ) *be anti-ferromagnetic. If* $\gamma > 1$*, then there exists a unique* $d_c = d_c(\beta, \gamma, \lambda) < \infty$ *such that* $c(d)$ *reaches its maximum at* d_c *for all positive* d*. Moreover,* $c(d)$ *is increasing for* $d < d_c$*, and decreasing for* $d > d_c$*.*

Proof. By the argument before, it is enough to show that there exists a unique $d_c = d_c(\beta, \gamma, \lambda)$ such that $s(d_c) = 0$, $s(d) > 0$ if $d < d_c$, and $s(d) < 0$ if $d > d_c$.

We rewrite $s(d)$ as follows,

$$\begin{aligned}
s(d) &= 1 - d\log q(\widehat{x}_d)\frac{\beta\widehat{x}_d^2-\gamma}{p(\widehat{x}_d)+d(1-\beta\gamma)\widehat{x}_d} \\
&= \frac{1}{p(\widehat{x}_d)+d(1-\beta\gamma)\widehat{x}_d}\left(p(\widehat{x}_d) - d\left(\log q(\widehat{x}_d)(\beta\widehat{x}_d^2-\gamma)-(1-\beta\gamma)\widehat{x}_d\right)\right) \\
&= \frac{p(\widehat{x}_d)-dr(\widehat{x}_d)}{p(\widehat{x}_d)+d(1-\beta\gamma)\widehat{x}_d}
\end{aligned}$$

where $r(x) = \log q(x)(\beta x^2-\gamma)-(1-\beta\gamma)x$ is defined above.

Now consider the function $t(x) = \dfrac{p(x)}{r(x)}$. By Lemma 7.31 there is a unique pole x_c for $t(x)$ on $x \geqslant 0$. On the continuous intervals $[0, x_c)$ and (x_c, ∞) of $t(x)$, we have

$$t'(x) = -\frac{\log q(x)}{\left(\log q(x)(\beta x^2-\gamma)-(1-\beta\gamma)x\right)^2}\left((1+\beta\gamma)\beta x^2 + 4\beta\gamma x + (1+\beta\gamma)\gamma\right)$$

Notice that $\log q(x) < \log(1/\gamma) < 0$ for $\gamma > 1$. We have $t(x)$ is increasing in x on each continuous interval for $x > 0$. Moreover $t(0) = \frac{1}{\log\gamma}$.

Depending on whether the discontinuous point x_c is achievable for $\widehat{x}_d$, there are two cases.

1. If $x_c \geqslant \widehat{x}_0$, then $\widehat{x}_d < x_c$ for all $d > 0$ as $\widehat{x}_d$ is decreasing in d. Furthermore $r(\widehat{x}_d) > 0$ for all $d > 0$ by Lemma 7.31 and $t(\widehat{x}_d) > 0$ for all $d > 0$ as well. On the other hand, $t(\widehat{x}_d)$ is strictly increasing in $\widehat{x}_d$ and hence strictly decreasing in d. Moreover $t(\widehat{x}_d)$ goes to $\frac{1}{\log\gamma} < \infty$ as d goes to ∞. Therefore d as a function intersects with $t(\widehat{x}_d)$ at a unique point $d_c = d_c(\beta,\gamma,\lambda) > 0$ such that $d_c = t(\widehat{x}_{d_c})$. Moreover $d < t(\widehat{x}_d)$ if $d < d_c$ and $d > t(\widehat{x}_d)$ if $d > d_c$. Since $r(\widehat{x}_d) > 0$, it implies that $s(d_c) = 0$, $s(d) > 0$ if $d < d_c$, and $s(d) < 0$ if $d > d_c$.

2. Otherwise $x_c < \widehat{x}_0$. There exists d_0 such that $\widehat{x}_{d_0} = x_c$ as $\widehat{x}_d$ goes to 0 as d goes to ∞. Then

$$s(d_0) = \frac{p(\widehat{x}_{d_0})}{p(\widehat{x}_{d_0})+d_0(1-\beta\gamma)\widehat{x}_{d_0}} > 0$$

   For $d \in [0, d_0]$, we have that $\widehat{x}_d \geqslant \widehat{x}_{d_0} = x_c$ and $r(\widehat{x}_d) \leqslant 0$ by Lemma 7.31. Hence $s(d) > 0$ for $d \in [0, d_0]$.

On the interval (d_0, ∞), it reduces to case (1). By a similar analysis, there exists a unique point $d_c = d_c(\beta, \gamma, \lambda) \in (d_0, \infty)$ such that $s(d_c) = 0$, $s(d) > 0$ if $d_0 < d < d_c$, and $s(d) < 0$ if $d > d_c$.

This completes our proof. □

The condition of $\gamma > 1$ in the above proof is crucial in order to show that $t(\widehat{x}_d)$ is decreasing in d. If $\gamma \leqslant 1$, such $t(\widehat{x}_d)$ may not decrease for all d, and the analysis would fail. In fact, since we have shown $s(d) > 0$ for all $d > 0$ if $\gamma \leqslant 1$, d and $t(\widehat{x}_d)$ do not intersect for any positive d.

Proof of Theorem 7.28. Recall that $c(d) = \frac{d(1-\beta\gamma)\widehat{x}_d}{p(\widehat{x}_d)}$. If $\gamma > 1$, as d goes to ∞, $\widehat{x}_d < 1/\gamma^d$ goes to 0, and therefore $c(d)$ goes to 0.

By Lemma 7.32, $c(d)$ achieve its unique maximum at $d_c > 0$. If $d_c < 1$, then as $c(1) < 1$, for all integer $d \geqslant 1$, $c(d) < 1$, and (β, γ, λ) is universally strictly unique. Otherwise $d_c \geqslant 1$, let $d_0 = \lfloor d_c \rfloor \geqslant 1$ and $d_1 = d_0 + 1$. Then $c(d)$ is increasing in $[1, d_0]$ and decreasing in $[d_1, \infty)$. If $c(d_0) < 1$ and $c(d_1) < 1$, then again (β, γ, λ) is universally strictly unique as for any integer $d \geqslant 1$, $c(d) < 1$.

Otherwise, $c(d_0) \geqslant 1$ or $c(d_1) \geqslant 1$ or both. Assume $c(d_0) \geqslant 1$ but $c(d_1) < 1$ first. Since $c(1) < 1$, $c(d_0) \geqslant 1$ and $c(d)$ is increasing in $[1, d_0]$, there exists a unique $\Delta_c \in [1, d_0]$ such that $c(\Delta_c) \geqslant 1$ and $c(\Delta_c - 1) < 1$. Moreover, let $\overline{\Delta_c} = d_0$. Then StrUnique$(\beta, \gamma, \lambda, d)$ holds for any integer $d \in [1, \Delta_c)$ or $d \in (\overline{\Delta_c}, \infty)$, and StrUnique$(\beta, \gamma, \lambda, d)$ fails for any integer $d \in [\Delta_c, \overline{\Delta_c}]$.

The case of $c(d_0) < 1$ but $c(d_1) \geqslant 1$ is similar. We pick $\Delta_c = d_1$. Since $c(d) \to 0$ as $d \to \infty$, there exists a $\overline{\Delta_c} \in [d_1, \infty)$ such that $c(\overline{\Delta_c}) \geqslant 1$ and $c(\overline{\Delta_c} + 1) < 1$ as required.

The last case is that $c(d_0) \geqslant 1$ and $c(d_1) \geqslant 1$. In this case, there exists $\Delta_c \in [1, d_0]$ such that $c(\Delta_c) \geqslant 1$ and $c(\Delta_c - 1) < 1$ and $\overline{\Delta_c} \in [d_1, \infty)$ such that $c(\overline{\Delta_c}) \geqslant 1$ and $c(\overline{\Delta_c} + 1) < 1$. As $c(d)$ is increasing in $[1, d_0]$ and decreasing in $[d_1, \infty)$, the theorem follows. □

7.7 Uniqueness Thresholds

In this section, we recast the uniqueness condition into various threshold forms. We have four parameters, (β, γ, λ) and the degree d. We may fix any three and discuss the threshold of the last one. Thresholds about d while fixing (β, γ, λ) has been shown in Section 7.6. In this section we fix any set of other three parameters. Due to the symmetry between β and γ, we only need to consider fixing either (β, λ) and d, or (β, γ) and d.

Thresholds of γ

The threshold of γ while fixing (β, λ) and d is summarized in the following theorem. For the hardness, we use results from [SS14].

Theorem 7.33. *For any $\beta \in [0, 1), \lambda > 0$, and an integer $\Delta \geqslant 2$, exactly one of the following two cases is true:*

- *for any $\gamma \in [0, 1/\beta)$, there exists an FPTAS for #Δ-2*Spin*(β, γ, λ);*
- *there exists a critical threshold $\gamma_c = \gamma_c(\beta, \lambda, \Delta)$ such that*
  1. *if $\gamma \in (\gamma_c, 1/\beta)$, then there exists an FPTAS for #Δ-2*Spin*(β, γ, λ);*
  2. *if $\gamma \in [0, \gamma_c)$, then there is no FPRAS for #Δ-2*Spin*(β, γ, λ) unless* **NP**=**RP**.

Moreover, $\gamma_c^\infty = \lim_{d \to \infty} \gamma_c(\beta, \lambda, d) \geqslant 1$ exists and $\gamma_c^\infty < 1/\beta$. It holds that

1. *if $\gamma \in (\gamma_c^\infty, \frac{1}{\beta})$, then there is an FPTAS for #2*Spin*(β, γ, λ);*
2. *if $\gamma \in [0, \gamma_c^\infty)$, then there is no FPRAS for #Δ_c-2*Spin*(β, γ, λ) unless* **NP**=**RP***, where*

$$\Delta_c = \arg\max_{d \geqslant 1} \mathrm{Ctr}(\beta, \gamma, \lambda, d).$$

We first show a lemma.

Lemma 7.34. *Let $0 \leqslant \beta < 1$ and $\lambda > 0$. For any integer $\Delta \geqslant 2$, either* StrUnique$(\beta, \gamma, \lambda, d)$ *holds for all $1 \leqslant d < \Delta$ and $\gamma \in [0, 1/\beta)$, or there exists a critical threshold $\gamma_c = \gamma_c(\beta, \lambda, \Delta)$ such that* StrUnique$(\beta, \gamma, \lambda, d)$ *holds for all $1 \leqslant d < \Delta$ if and only if $\gamma \in (\gamma_c, \frac{1}{\beta})$.*

Proof. Let $f_\gamma(x) := \lambda \left(\frac{\beta x + 1}{x + \gamma}\right)^d$. Notice that the definition is the same as that of $f_d(x)$, but we want to make clear the dependence on γ, instead of that on d. Let $\widehat{x}_\gamma$ be the unique positive fixed point of $f_\gamma(x)$ such that $\widehat{x}_\gamma = \lambda \left(\frac{\beta \widehat{x}_\gamma + 1}{\widehat{x}_\gamma + \gamma}\right)^d$.

We first claim that $|f'_\gamma(\widehat{x}_\gamma)|$ is decreasing in γ. For any $\gamma' > \gamma$, let $\widehat{x}_{\gamma'}$ be the unique fixed point of $f_{\gamma'}(x)$, that is, $\widehat{x}_{\gamma'} = \lambda\left(\frac{\beta\widehat{x}_{\gamma'}+1}{\widehat{x}_{\gamma'}+\gamma'}\right)^d$. We claim that $\widehat{x}_{\gamma'} < \widehat{x}_\gamma$. Assume for contradiction that $\widehat{x}_{\gamma'} \geqslant \widehat{x}_\gamma$. Since for any anti-ferromagnetic (β, γ, λ), the function $f_\gamma(x)$ is monotonically decreasing in x, we have that

$$\widehat{x}_\gamma = \lambda\left(\frac{\beta\widehat{x}_\gamma + 1}{\widehat{x}_\gamma + \gamma}\right)^d \geqslant \lambda\left(\frac{\beta\widehat{x}_{\gamma'} + 1}{\widehat{x}_{\gamma'} + \gamma}\right)^d$$
$$> \lambda\left(\frac{\beta\widehat{x}_{\gamma'} + 1}{\widehat{x}_{\gamma'} + \gamma'}\right)^d = \widehat{x}_{\gamma'}.$$

Contradiction.

Therefore $\widehat{x}_{\gamma'} < \widehat{x}_\gamma$, which implies that

$$\lambda\left(\beta + \frac{(1-\beta\gamma')}{\widehat{x}_{\gamma'} + \gamma'}\right)^d = \widehat{x}_{\gamma'} < \widehat{x}_\gamma = \lambda\left(\beta + \frac{(1-\beta\gamma)}{\widehat{x}_\gamma + \gamma}\right)^d.$$

So we have that $\frac{(1-\beta\gamma')}{\widehat{x}_{\gamma'}+\gamma'} < \frac{(1-\beta\gamma)}{\widehat{x}_\gamma+\gamma}$. For $\widehat{x}_{\gamma'} < \widehat{x}_\gamma$, it also holds that $\frac{\widehat{x}_{\gamma'}}{\beta\widehat{x}_{\gamma'}+1} < \frac{\widehat{x}_\gamma}{\beta\widehat{x}_\gamma+1}$. Multiplying above two inequalities together, we have that

$$|f'_{\gamma'}(\widehat{x}_{\gamma'})| = \frac{d(1-\beta\gamma')\widehat{x}_{\gamma'}}{(\beta\widehat{x}_{\gamma'}+1)(\widehat{x}_{\gamma'}+\gamma')} < \frac{d(1-\beta\gamma)\widehat{x}_\gamma}{(\beta\widehat{x}_\gamma+1)(\widehat{x}_\gamma+\gamma)} = |f'_\gamma(\widehat{x}_\gamma)|.$$

Next we show that for any $0 \leqslant \beta < 1$, $\lambda > 0$, and integer $d \geqslant 1$, $|f'_\gamma(\widehat{x}_\gamma)|$ goes to 0 as γ goes to $1/\beta$. Notice that $\widehat{x}_\gamma \geqslant \lambda\beta^d$. It implies that $\frac{d\widehat{x}_\gamma}{(\beta\widehat{x}_\gamma+1)(\widehat{x}_\gamma+\gamma)} \leqslant \frac{d}{\lambda\beta^{d+1}}$. Hence $|f'_\gamma(\widehat{x}_\gamma)| = \frac{d(1-\beta\gamma)\widehat{x}_\gamma}{(\beta\widehat{x}_\gamma+1)(\widehat{x}_\gamma+\gamma)} \leqslant \frac{d(1-\beta\gamma)}{\lambda\beta^{d+1}}$ goes to 0 as γ goes to $1/\beta$.

Since $|f'_\gamma(\widehat{x}_\gamma)|$ is decreasing in γ, if $|f'_0(\widehat{x}_0)| < 1$, then for all $\gamma \in [0, 1/\beta)$, $|f'_\gamma(\widehat{x}_\gamma)| < 1$. In this case let $\gamma_d = -1$. Otherwise $|f'_0(\widehat{x}_0)| \geqslant 1$. Since $|f'_\gamma(\widehat{x}_\gamma)|$ goes to 0 as γ goes to $1/\beta$, there exists a unique γ_d such that $|f'_{\gamma_d}(\widehat{x}_{\gamma_d})| = 1$, and $|f'_\gamma(\widehat{x}_\gamma)| < 1$ for all $\gamma \in (\gamma_d, 1/\beta)$.

If $\gamma_d = -1$ for all $1 \leqslant d < \Delta$, then $\mathrm{StrUnique}(\beta, \gamma, \lambda, d)$ holds for all $1 \leqslant d < \Delta$ and $\gamma \in [0, 1/\beta)$. Otherwise let $\gamma_c = \gamma_c(\beta, \lambda, \Delta) = \max_{1 \leqslant d < \Delta} \gamma_d \geqslant 0$. Then $\mathrm{StrUnique}(\beta, \gamma, \lambda, d)$ holds for all $1 \leqslant d < \Delta$ if and only if $\gamma \in (\gamma_c, \frac{1}{\beta})$. □

Now we are ready to prove Theorem 7.33.

Proof of Theorem 7.33. All claims regarding hardness are due to Sly and Sun [SS14].

The first part of Theorem 7.33 follows from Lemma 7.34 and Theorem 7.25.

For the second part, the existence of Δ_c is guaranteed by Lemma 7.32. We want to apply Theorem 7.26. We only need to show that $\gamma_c^\infty = \lim_{d\to\infty} \gamma_c(\beta, \lambda, d)$ exists, and $\gamma_c^\infty \in [1, 1/\beta)$. It is easy to see that $\gamma_c(\beta, \lambda, d)$ is non-decreasing in d and $\gamma_c(\beta, \lambda, d) \leqslant 1/\beta$ by the definition of $\gamma_c(\beta, \lambda, d)$. Hence γ_c^∞ exists. Assume for contradiction that $\gamma_c^\infty < 1$. Then there exists $\gamma \in (\gamma_c^\infty, 1]$ such that (β, γ, λ) is universally strictly unique, contradicting to Lemma 7.5.

We still need to show that $\gamma_c^\infty < 1/\beta$. It is sufficient to show that there exists a $1 \leqslant \gamma < 1/\beta$ such that (β, γ, λ) is universally strictly unique. First pick an arbitrary $\gamma_0 \in (1, 1/\beta)$. By Theorem 7.28, there exists a $\overline{\Delta_c}$ such that for any integer $d > \overline{\Delta_c}$, StrUnique$(\beta, \gamma_0, \lambda, d)$ holds. Let $\gamma_1 = \max\left\{\frac{1}{\beta}\left(\frac{\overline{\Delta_c}}{\overline{\Delta_c}+2}\right)^2, \gamma_0\right\}$. Clearly $\gamma_1 < 1/\beta$. For any $d > \overline{\Delta_c}$, as $\gamma_1 > \gamma_0$, StrUnique$(\beta, \gamma_1, \lambda, d)$ holds due to the monotonicity of $|f'_\gamma(\widehat{x}_\gamma)|$ showed in the proof of Lemma 7.34. Moreover, by our choice $\frac{1-\sqrt{\beta\gamma_1}}{1+\sqrt{\beta\gamma_1}} < \frac{1}{\overline{\Delta_c}+1}$. For any $1 \leqslant d \leqslant \overline{\Delta_c}$,

$$|f'_d(\widehat{x}_d)| = \frac{d(1-\beta\gamma_1)\widehat{x}_d}{(\beta\widehat{x}_d + 1)(\widehat{x}_d + \gamma_1)} \leqslant \frac{d\left(1-\sqrt{\beta\gamma_1}\right)}{1+\sqrt{\beta\gamma_1}} \leqslant \frac{d}{\overline{\Delta_c}+1} < 1.$$

Hence for any $1 \leqslant d \leqslant \overline{\Delta_c}$, StrUnique$(\beta, \gamma_1, \lambda, d)$ holds as well. To sum up, $(\beta, \gamma_1, \lambda)$ is universally strictly unique. The theorem follows. □

Thresholds of λ

Here we discuss thresholds about λ while fixing (β, γ) and Δ. This setup is closer to studies about the hardcore gas model. Cases are more complicated than those of Theorem 7.33. We need to distinguish between hard constraints ($\beta = 0$) and soft constraints ($\beta > 0$).

Hard constraints

We deal with hard constraints first. Let $\lambda_c(0, \gamma, \Delta) := \min_{1\leqslant d<\Delta} \frac{\gamma^{d+1} d^d}{(d-1)^{d+1}}$. In particular, for any $\gamma > 0$, $\lambda_c(0, \gamma, 2) = \infty$. Then we have the following lemma.

Lemma 7.35. *Let $\beta = 0$, $\gamma > 0$, and $\Delta \geqslant 2$ be an integer. There exists a critical threshold $\lambda_c = \lambda_c(0, \gamma, \Delta)$ such that* StrUnique$(\beta, \gamma, \lambda, d)$ *holds for all $1 \leqslant d < \Delta$ if and only if $\lambda \in (0, \lambda_c)$.*

Proof. For $\Delta = 2$, it is easy to verify that $c(1) < 1$ for any anti-ferromagnetic (β, γ, λ) . Hence $\mathrm{StrUnique}(0, \gamma, \lambda, 1)$ always holds and $\lambda_c = \infty$. In the following we assume that $\Delta \geqslant 3$ and $2 \leqslant d < \Delta$.

As $\beta = 0$, $|f'_d(\widehat{x}_d)| = \frac{d(1-\beta\gamma)\widehat{x}_d}{(\beta\widehat{x}_d+1)(\widehat{x}_d+\gamma)} = \frac{d\widehat{x}_d}{\widehat{x}_d+\gamma}$, and $\mathrm{StrUnique}(0, \gamma, \lambda, d)$ holds if and only if $\widehat{x}_d < \frac{\gamma}{d-1}$. Recall that $\widehat{x}_d = \lambda \left(\frac{1}{\widehat{x}_d+\gamma}\right)^d$. Then $\widehat{x}_d < \frac{\gamma}{d-1}$ if and only if

$$\lambda = \widehat{x}_d(\widehat{x}_d + \gamma)^d < \frac{\gamma^{d+1} d^d}{(d-1)^{d+1}}.$$

Hence $\mathrm{StrUnique}(0, \gamma, \lambda, d)$ holds for all $1 \leqslant d < \Delta$ if and only if $\lambda < \lambda_c = \lambda_c(0, \gamma, \Delta) = \min_{1 \leqslant d < \Delta} \frac{\gamma^{d+1} d^d}{(d-1)^{d+1}}$. □

Now we are ready to state and prove our theorem.

Theorem 7.36. *Let $\beta = 0$, $\gamma > 0$, and $\Delta \geqslant 2$ be an integer. There exists a critical threshold $\lambda_c = \lambda_c(0, \gamma, \Delta)$ such that*

1. *if $\lambda \in (0, \lambda_c)$, then there exists an FPTAS for #Δ-2Spin$(0, \gamma, \lambda)$;*
2. *if $\lambda \in (\lambda_c, \infty)$, then there is no FPRAS for #Δ-2Spin$(0, \gamma, \lambda)$ unless* **NP**=**RP**.

Moreover,

1. *if $\gamma \leqslant 1$, then for any $\lambda > 0$ then there is no FPRAS for #$\Delta_{\gamma,\lambda}$-2Spin$(0, \gamma, \lambda)$ unless* **NP**=**RP** *where $\Delta_{\gamma,\lambda}$ is a sufficiently large integer depending on γ and λ;*
2. *if $\gamma > 1$, then there exists a critical λ_c^∞ such that*
   a) *if $\lambda \in (0, \lambda_c^\infty)$, then there exists an FPTAS for #2Spin$(0, \gamma, \lambda)$;*
   b) *if $\lambda \in (\lambda_c^\infty, \infty)$, then there is no FPRAS for #Δ_c-2Spin$(0, \gamma, \lambda)$ unless* **NP**=**RP** *where*

$$\Delta_c = \arg\max_{d \geqslant 1} \mathrm{Ctr}(0, \gamma, \lambda, d).$$

Proof. Again, all hardness statements are due to Sly and Sun [SS14].

The first part of Theorem 7.36 follows from Lemma 7.35 and Theorem 7.25.

For the second part, if $\gamma \leqslant 1$, then $\frac{\gamma^{d+1} d^d}{(d-1)^{d+1}}$ is strictly decreasing in d and goes to 0. Hence there is no λ such that $(0, \gamma, \lambda)$ is universally strictly unique. By Lemma 7.30, if $\gamma \leqslant 1$, then there is a sufficiently large integer $\Delta_{\gamma,\lambda}$ such that $\text{StrUnique}(0, \gamma, \lambda, \Delta_{\gamma,\lambda})$ fails.

Otherwise assume $\gamma > 1$. The existence of Δ_c is guaranteed by Lemma 7.32. To get tractability results, we want to apply Lemma 7.35 and Theorem 7.26. Let $l(d) = \frac{\gamma^{d+1} d^d}{(d-1)^{d+1}}$ for $d > 1$. Then

$$l'(d) = l(d) \left(\log \gamma + \log \left(\frac{d}{d-1} \right) - \frac{2}{d-1} \right).$$

The function $-\log\left(\frac{d}{d-1}\right) + \frac{2}{d-1}$ is decreasing for all $d > 1$. If $\log \gamma \geqslant -\log\left(\frac{2}{2-1}\right) + \frac{2}{2-1} = -\log 2 + 2$, that is, $\gamma \geqslant e^2/2 \approx 3.69453$, then $l'(d) \geqslant 0$ for all $d \geqslant 2$. Recall that $l(1) = \infty$, we have that $\lambda_c^\infty = \min_{d \geqslant 1} l(d) = l(2) = 4\gamma^3$ where the minimum is taken with respect to integers d.

Otherwise, $\gamma < e^2/2$ and there exists a unique $d_0 > 2$ such that $\log \gamma = -\log\left(\frac{d_0}{d_0-1}\right) + \frac{2}{d_0-1}$. The function $l(d)$ takes its minimum at this point $l'(d_0) = 0$ in $d > 1$. Let $D_0 = \lfloor d_0 \rfloor$ and $D_1 = D_0 + 1$. Then $\lambda_c^\infty = \min_{d \geqslant 1} l(d) = \min\{l(D_0), l(D_1)\}$, where the first minimum is taken with respect to integers d. $\square$

Soft constraints

Next we deal with soft constraints. The main result is summarized as follows.

Theorem 7.37. *Let $0 < \beta \leqslant \gamma$, $\beta\gamma < 1$, and $\Delta \geqslant 2$ be an integer. We have the following cases:*

1. *if $\sqrt{\beta\gamma} > \frac{\Delta-2}{\Delta}$, then for any $\lambda > 0$, then there exists an FPTAS for #Δ-2SPIN(β, γ, λ);*

2. *if $\sqrt{\beta\gamma} \leqslant \frac{\Delta-2}{\Delta}$, then there exist two critical thresholds $\lambda_c = \lambda_c(\beta, \gamma, \Delta)$ and $\overline{\lambda}_c = \overline{\lambda}_c(\beta, \gamma, \Delta)$ such that*

    a) *if $\lambda \in (0, \lambda_c) \cup (\overline{\lambda}_c, \infty)$, there exists an FPTAS for #Δ-2SPIN(β, γ, λ);*

    b) *if $\lambda \in (\lambda_c, \overline{\lambda}_c)$, there is no FPRAS for #Δ-2SPIN(β, γ, λ) unless* **NP**=**RP**.

Moreover,

1. *if* $\gamma \leqslant 1$*, then for any* $\lambda > 0$ *then there is no FPRAS for* $\#\Delta_{\gamma,\lambda}\text{-}2\textsc{Spin}(0,\gamma,\lambda)$ *unless* **NP**=**RP** *where* $\Delta_{\gamma,\lambda}$ *is a sufficiently large integer depending on* γ *and* λ*;*

2. *if* $\gamma > 1$*, then there exists a critical* λ_c^{∞} *such that*

   a) *if* $\lambda \in (0, \lambda_c^{\infty})$*, then there exists an FPTAS for* $\#2\textsc{Spin}(\beta,\gamma,\lambda)$*;*

   b) *if* $\lambda \in (\lambda_c^{\infty}, \infty)$*, then there is no FPRAS for* $\#\Delta_c\text{-}2\textsc{Spin}(0,\gamma,\lambda)$ *unless* **NP**=**RP** *where*

$$\Delta_c = \arg\max_{d \geqslant 1} \mathrm{Ctr}(\beta,\gamma,\lambda,d).$$

We first observe that if $\sqrt{\beta\gamma} > \frac{\Delta-2}{\Delta}$, then the uniqueness holds for any $\lambda > 0$.

Lemma 7.38. *Let* $0 \leqslant \beta \leqslant \gamma$*,* $\beta\gamma < 1$*, and* $\Delta \geqslant 2$ *be an integer. If* $\sqrt{\beta\gamma} > \frac{\Delta-2}{\Delta}$*, then for any* $\lambda > 0$ *and any integer* $1 \leqslant d < \Delta$*,* $\mathrm{StrUnique}(\beta,\gamma,\lambda,d)$ *holds.*

Proof. We note that $|f_d'(\widehat{x}_d)| = \frac{d(1-\beta\gamma)\widehat{x}_d}{(\beta\widehat{x}_d+1)(\widehat{x}_d+\gamma)}$, as a rational function in $\widehat{x}_d$, is not monotone. It achieves its maximum value at $\widehat{x}_d = \sqrt{\gamma/\beta}$. Therefore, if for any $1 \leqslant d < \Delta$, $\frac{d(1-\beta\gamma)\sqrt{\gamma/\beta}}{(\beta\sqrt{\gamma/\beta}+1)(\sqrt{\gamma/\beta}+\gamma)} < 1$, then $\mathrm{StrUnique}(\beta,\gamma,\lambda,d)$ holds for any $\lambda > 1$. In fact, the condition is equivalent to $\sqrt{\beta\gamma} > \frac{d-1}{d+1}$. If $\sqrt{\beta\gamma} > \frac{\Delta-2}{\Delta}$, then for all integers $1 \leqslant d < \Delta$, $\sqrt{\beta\gamma} > \frac{d-1}{d+1}$, and hence $\mathrm{StrUnique}(\beta,\gamma,\lambda,d)$ holds for any $\lambda > 0$. □

Otherwise $\sqrt{\beta\gamma} \leqslant \frac{d-1}{d+1}$ (denote $d = \Delta - 1$) and $d \geqslant \frac{1+\sqrt{\beta\gamma}}{1-\sqrt{\beta\gamma}}$. We have two thresholds, instead of just one as in Lemma 7.35. We need to do some technical preparation first. If $\sqrt{\beta\gamma} \leqslant \frac{d-1}{d+1}$, then the equation $d(1-\beta\gamma)x = (\beta x + 1)(x + \gamma)$ has two real roots $x_1(d) \leqslant x_2(d)$. To be specific,

$$\begin{aligned} x_1(d) &= \frac{1}{2\beta}\left(-1-\beta\gamma+d(1-\beta\gamma)-\sqrt{(-1-\beta\gamma+d(1-\beta\gamma))^2-4\beta\gamma}\right), \\ x_2(d) &= \frac{1}{2\beta}\left(-1-\beta\gamma+d(1-\beta\gamma)+\sqrt{(-1-\beta\gamma+d(1-\beta\gamma))^2-4\beta\gamma}\right). \end{aligned} \tag{7.32}$$

Moreover, $x_1(d) + x_2(d) = \frac{d(1-\beta\gamma)-(1+\beta\gamma)}{\beta} \geqslant \frac{(1+\sqrt{\beta\gamma})^2-(1+\beta\gamma)}{\beta} > 0$ and $x_1(d)x_2(d) = \gamma/\beta$. Hence both $x_1(d)$ and $x_2(d)$ are positive. It is easy to see that $x_2(d)$ is monotonically increasing in d and goes to ∞ as d grows. As $x_1(d)x_2(d) = \gamma/\beta$, $x_1(d)$ is monotonically decreasing in d and goes to 0 as d grows.. Let $\lambda_i(d) = x_i(d)\left(\frac{x_i(d)+\gamma}{\beta x_i(d)+1}\right)^d$, for $i = 1, 2$. We need the following technical lemma.

Lemma 7.39. *If $\gamma \leqslant 1$, then $\lambda_1(d)$ is decreasing in d and goes to 0 as d grows. If $\gamma > 1$, then there exists a unique integer d_0 such that $\lambda_1(d)$ takes its minimum at $d = d_c$ among integers $d \geqslant \frac{1+\sqrt{\beta\gamma}}{1-\sqrt{\beta\gamma}}$. The function $\lambda_2(d)$ is increasing in d and goes to ∞ as d grows.*

Proof. It is easy to calculate that

$$\begin{aligned}\frac{\lambda_i'(d)}{\lambda_i(d)} &= \frac{x_i'(d)}{x_i(d)} + \log\frac{x_i(d)+\gamma}{\beta x_i(d)+1} + \frac{d(1-\beta\gamma)x_i'(d)}{(x_i(d)+\gamma)(\beta x_i(d)+1)} \\ &= \frac{2x_i'(d)}{x_i(d)} + \log\frac{x_i(d)+\gamma}{\beta x_i(d)+1},\end{aligned} \tag{7.33}$$

as $d(1-\beta\gamma)x_i(d) = (x_i(d)+\gamma)(\beta x_i(d)+1)$. Since $x_2(d) \geqslant x_1(d)$ and $\gamma \geqslant \beta$, we have that $x_2(d) \geqslant \sqrt{\frac{\gamma}{\beta}} \geqslant \frac{1-\gamma}{1-\beta}$. Hence $\frac{x_2(d)+\gamma}{\beta x_2(d)+1} \geqslant 1$. Together with $x_2'(d) > 0$, from (7.33) we have that $\lambda_2'(d) > 0$.

Moreover we can calculate that

$$\frac{x_1'(d)}{x_1(d)} = -\frac{1-\beta\gamma}{\sqrt{((1-\beta\gamma)d-1-\beta\gamma)^2-4\beta\gamma}},$$

and

$$\frac{\lambda_1''(d)}{\lambda_1(d)} = \frac{(1-\beta\gamma)^3(d^2-1)}{d\left(((1-\beta\gamma)d-1-\beta\gamma)^2-4\beta\gamma\right)^{3/2}} + \left(\frac{2x_1'(d)}{x_1(d)} + \log\frac{x_1(d)+\gamma}{\beta x_1(d)+1}\right)^2 > 0.$$

Hence $\lambda_1'(d)$ is increasing goes from $-\infty$ to $\log\gamma$ as d goes from $\frac{1+\sqrt{\beta\gamma}}{1-\sqrt{\beta\gamma}}$ to ∞.

If $\gamma \leqslant 1$, then $\lambda_1'(d) < 0$ for all $d \geqslant \frac{1+\sqrt{\beta\gamma}}{1-\sqrt{\beta\gamma}}$. Hence $\lambda_1(d)$ is decreasing in d.

If $\gamma > 1$, then there exists a unique d_0 such that $\lambda_1'(d_0) = 0$, $\lambda_1'(d) < 0$ if $\frac{1+\sqrt{\beta\gamma}}{1-\sqrt{\beta\gamma}} \leqslant d < d_0$, and $\lambda_1(d)' > 0$ if $d > d_0$. Hence $\lambda_1(d)$ takes its minimum at $d = d_0$ in the range $\left[\frac{1+\sqrt{\beta\gamma}}{1-\sqrt{\beta\gamma}}, \infty\right)$. Notice that we only care about integers. Let $D_0 = \max\left\{\left\lceil\frac{1+\sqrt{\beta\gamma}}{1-\sqrt{\beta\gamma}}\right\rceil, \lfloor d_0 \rfloor\right\}$ and $D_1 = D_0 + 1$. Then $d_c = D_i$ where $i \in \{0,1\}$ such that $\lambda_1(D_i) \leqslant \lambda_1(D_{1-i})$. □

Next we analyze the uniqueness condition for a single integer d.

Lemma 7.40. *Let $0 < \beta \leqslant \gamma$, $\beta\gamma < 1$, and $d \geqslant 1$ be an integer. If $\sqrt{\beta\gamma} \leqslant \frac{d-1}{d+1}$, then* StrUnique$(\beta, \gamma, \lambda, d)$ *holds if and only if $\lambda \in (0, \lambda_1(d)) \cup (\lambda_2(d), \infty)$.*

Proof. Recall that under our assumption the equation $\frac{d(1-\beta\gamma)x}{(\beta x+1)(x+\gamma)} = 1$ has two positive roots $x_1(d)$ and $x_2(d)$ as in (7.32). It then holds that $|f_d'(\widehat{x}_d)| = \frac{d(1-\beta\gamma)\widehat{x}_d}{(\beta\widehat{x}_d+1)(\widehat{x}_d+\gamma)} < 1$ if and only if $\widehat{x}_d < x_1(d)$ or $\widehat{x}_d > x_2(d)$. Note that $x\left(\frac{x+\gamma}{\beta x+1}\right)^d$ is monotonically increasing in x for any fixed d. Thus $\widehat{x}_d < x_1(d)$ if and only if

$$\lambda = \widehat{x}_d\left(\frac{\widehat{x}_d+\gamma}{\beta\widehat{x}_d+1}\right)^d < x_1(d)\left(\frac{x_1(d)+\gamma}{\beta x_1(d)+1}\right)^d = \lambda_1(d),$$

and $\widehat{x}_d > x_2(d)$ if and only if

$$\lambda = \widehat{x}_d\left(\frac{\widehat{x}_d+\gamma}{\beta\widehat{x}_d+1}\right)^d > x_2(d)\left(\frac{x_2(d)+\gamma}{\beta x_2(d)+1}\right)^d = \lambda_2(d).$$

Therefore, $|f_d'(\widehat{x}_d)| < 1$ if and only if $\lambda < \lambda_1(d)$ or $\lambda > \lambda_2(d)$. □

Then we have the following two lemmas.

Lemma 7.41. *Let* $0 < \beta \leqslant \gamma$, $\beta\gamma < 1$, *and* $\Delta \geqslant 2$ *be an integer. If* $\sqrt{\beta\gamma} \leqslant \frac{\Delta-2}{\Delta}$, *then there exist two critical thresholds* $\lambda_c = \lambda_c(\beta,\gamma,\Delta)$ *and* $\overline{\lambda}_c = \overline{\lambda}_c(\beta,\gamma,\Delta) = \lambda_2(\Delta-1)$ *such that* $\mathrm{StrUnique}(\beta,\gamma,\lambda,d)$ *holds for all integers* $1 \leqslant d < \Delta$ *if and only if* $\lambda \in (0,\lambda_c) \cup (\overline{\lambda}_c,\infty)$.

Proof. Let $\overline{\Delta}$ be the largest integer such that $\overline{\Delta} < \frac{1+\sqrt{\beta\gamma}}{1-\sqrt{\beta\gamma}}$ and therefore $\sqrt{\beta\gamma} > \frac{\overline{\Delta}-1}{\overline{\Delta}+1}$. Since $\sqrt{\beta\gamma} \leqslant \frac{\Delta-2}{\Delta}$, $\Delta - 1 > \overline{\Delta}$. Moreover, we have that $\sqrt{\beta\gamma} > \frac{d-1}{d+1}$ for all $1 \leqslant d \leqslant \overline{\Delta}$ and $\sqrt{\beta\gamma} \leqslant \frac{d-1}{d+1}$ for all $\overline{\Delta} < d < \Delta$. By Lemma 7.38, for all $\lambda > 0$, $\mathrm{StrUnique}(\beta,\gamma,\lambda,d)$ holds for all $d \in [1,\overline{\Delta}]$.

For any integer $d \in (\overline{\Delta},\Delta)$, we have that $\sqrt{\beta\gamma} \leqslant \frac{d-1}{d+1}$. Hence by Lemma 7.40, StrUnique (β,γ,λ,d) holds if and only if $\lambda \in (0,\lambda_1(d)) \cup (\lambda_2(d),\infty)$. Let

$$\lambda_c = \lambda_c(\beta,\gamma,\Delta) = \min_{\overline{\Delta}<d<\Delta} \lambda_1(d),$$
$$\overline{\lambda}_c = \overline{\lambda}_c(\beta,\gamma,\Delta) = \max_{\overline{\Delta}<d<\Delta} \lambda_2(d) = \lambda_2(\Delta-1),$$

where min and max are taken over integers, and we get the second line since $\lambda_2'(d) > 0$ by Lemma 7.39. It holds that $\mathrm{StrUnique}(\beta,\gamma,\lambda,d)$ holds for any integer $d \in [1,\Delta)$ if and only if $\lambda \in (0,\lambda_c) \cup (\overline{\lambda}_c,\infty)$. □

We remark that if $\beta = \gamma < 1$, that is, the system is an Ising model, then $\lambda_1(d)\lambda_2(d) = 1$. Moreover, by Lemma 7.39, $\lambda_1(d)$ is decreasing, and hence $\lambda_c = \lambda_1(\Delta - 1)$. We then get that $\lambda_c \cdot \overline{\lambda}_c = 1$. In this case, $\mathrm{StrUnique}(\beta, \gamma, \lambda, d)$ holds for all integers $1 \leqslant d < \Delta$ if and only if $|\log \lambda| > \log \overline{\lambda}_c$. In particular, we have the following lemma for the Ising model, which will be useful in Section 9.6.

Lemma 7.42. *Let* $0 < \beta < 1$ *and* $\Delta = \left\lfloor \frac{1+\beta}{1-\beta} \right\rfloor + 1$ *be an integer. Then for any* $C > 1$*, there exists* $\lambda_C > 0$ *such that* $\lambda_C > 1$*,* $1/C < \lambda_C < C$*, and* $\mathrm{Ctr}(\beta, \beta, \lambda_C, \Delta) > 1$*.*

Proof. We claim that for any integer d such that $\beta < \frac{d-1}{d+1}$, $\mathrm{Ctr}(\beta, \beta, 1, d) > 1$. The lemma follows from continuity and the fact that $\Delta > \frac{1+\beta}{1-\beta}$.

To show the claim, notice that if $\lambda = 1$, then the fixed point $\widehat{x}_d = 1$ for any $d > 0$. Hence $\mathrm{Ctr}(\beta, \beta, 1, d) = \frac{d(1-\beta^2)}{(1+\beta)^2} = \frac{d(1-\beta)}{1+\beta} > 1$ if $d > \frac{1+\beta}{1-\beta}$. □

At last we discuss universally strictly uniqueness.

Lemma 7.43. *If* $\beta > 0$ *and* $\gamma > 1$*, there exists a constant* $\lambda_c^\infty = \lambda_c(\beta, \gamma)$ *such that* (β, γ, λ) *is universally strictly unique if and only if* $\lambda \in (0, \lambda_c^\infty)$*.*

Proof. Let $\overline{\Delta}$ be the largest integer such that $\overline{\Delta} < \frac{1+\sqrt{\beta\gamma}}{1-\sqrt{\beta\gamma}}$. Then by Lemma 7.38, for any integer $1 \leqslant d \leqslant \overline{\Delta}$ and $\lambda > 0$, $\mathrm{StrUnique}(\beta, \gamma, \lambda, d)$ holds. By Lemma 7.40, for any integer $d > \overline{\Delta}$, $\mathrm{StrUnique}(\beta, \gamma, \lambda, d)$ holds if and only if $\lambda \in (0, \lambda_1(d)) \cup (\lambda_2(d), \infty)$. By Lemma 7.39, $\lambda_2(d)$ goes to infinity as d grows. Since $\gamma > 1$, then again by Lemma 7.39, $\lambda_c^\infty = \min_{d > \overline{\Delta}} \lambda_1(d) = \lambda_1(d_c)$. Therefore, (β, γ, λ) is universally strictly unique if and only if $\lambda \in (0, \lambda_c^\infty)$. □

Now we are ready to show Theorem 7.37.

Proof of Theorem 7.37. Once again, all hardness statements are due to Sly and Sun [SS14].

Statements about FPTAS in the first part of Theorem 7.37 follows from Lemma 7.41 and Lemma 7.25. Statements about FPTAS in the second part follows from Lemma 7.5, Lemma 7.43, and Theorem 7.26. □

7.8 Correlation Decay at the Threshold

In this section, we show that correlation decays at the rate of $O(\ell^{-1/2})$ when (β, γ, λ) and Δ are right at the uniqueness threshold. That is,

$$f_d'(\widehat{x}_d) = \frac{d(\beta\gamma - 1)\widehat{x}_d}{(\beta\widehat{x}_d + 1)(\widehat{x}_d + 1)} = -1.$$

where $\widehat{x}_d$ is the unique solution such that $\widehat{x}_d = f_d(\widehat{x}_d)$. Here we are still working under the assumption that $\gamma > 0$, $\beta \in [0, \gamma]$, $\beta\gamma < 1$, $\lambda > 0$, and d is an integer. Derivatives of $f_d(x)$ are

$$\begin{aligned}
f_d'(x) &= \frac{d(\beta\gamma - 1)f_d(x)}{(\beta x + 1)(x + \gamma)}, \\
f_d''(x) &= f_d'(x) \cdot \left(\frac{(d-1)\beta}{\beta x + 1} - \frac{d+1}{x + \gamma} \right), \\
f_d'''(x) &= f_d''(x) \left(\frac{(d-1)\beta}{\beta x + 1} - \frac{d+1}{x + \gamma} \right) - f_d'(x) \cdot \left(\frac{(d-1)\beta^2}{(\beta x + 1)^2} - \frac{d+1}{(x + \gamma)^2} \right) \\
&= \frac{f_d''(x)^2}{f_d'(x)} - f_d'(x) \cdot \left(\frac{(d-1)\beta^2}{(\beta x + 1)^2} - \frac{d+1}{(x + \gamma)^2} \right).
\end{aligned}$$

Hence we have that

$$f_d''(\widehat{x}_d) = \frac{1}{\widehat{x}_d} + \frac{\beta}{\beta\widehat{x}_d + 1} + \frac{1}{\widehat{x}_d + \gamma}, \tag{7.34}$$

$$f_d'''(\widehat{x}_d) = -f_d''(\widehat{x}_d)^2 - \frac{\beta}{\widehat{x}_d(\beta\widehat{x}_d + 1)} - \frac{1}{\widehat{x}_d(\widehat{x}_d + \gamma)} - \frac{\beta^2}{(\beta\widehat{x}_d + 1)^2} - \frac{1}{(\widehat{x}_d + \gamma)^2}, \tag{7.35}$$

where in the last equation we use the fact that

$$d\left(\frac{\beta}{\beta\widehat{x}_d + 1} - \frac{1}{\widehat{x}_d + \gamma} \right) = -\frac{1}{\widehat{x}_d}, \tag{7.36}$$

since $f_d'(\widehat{x}_d) = -1$. We will use $\Phi_*(x)$ as our potential function, that is

$$\varphi_*(x) = \Phi_*'(x) = \frac{1}{\sqrt{x(\beta x + 1)(x + \gamma)}}.$$

Recall that we pick this potential function such that

$$\frac{\varphi'(\widehat{x}_d)}{\varphi(\widehat{x}_d)} = -\frac{f_d''(\widehat{x}_d)}{2}. \tag{7.37}$$

We consider the recursion with this potential function, that is $g_d(y) := \Phi(f_d(\Phi^{-1}(y)))$. We see that $\widehat{y}_d = \Phi^{-1}(\widehat{x}_d)$ is the unique fixed point of $g_d(y)$. Derivatives of $g_d(y)$ of the first three orders are

$$\begin{aligned}
g_d'(y) &= \frac{\varphi(f_d(x))}{\varphi(x)} f_d'(x), \\
g_d''(y) &= \frac{g_d'(y)}{\varphi(x)} \left(\frac{\varphi'(f_d(x)) f_d'(x)}{\varphi(f_d(x))} + \frac{f_d''(x)}{f_d'(x)} - \frac{\varphi'(x)}{\varphi(x)} \right), \\
g_d'''(y) &= \frac{g_d''(y)^2}{g_d'(y)} - g_d''(y) \frac{\varphi'(x)}{\varphi(x)} + \frac{g_d'(y)}{\varphi(x)} \left(\frac{f_d'''(x)}{f_d'(x)} - \left(\frac{f_d''(x)}{f_d'(x)} \right)^2 - \frac{\varphi''(x)}{\varphi(x)} + \left(\frac{\varphi'(x)}{\varphi(x)} \right)^2 \right. \\
&\quad \left. + \frac{\varphi''(f_d(x)) f_d'(x)^2}{\varphi(f_d(x))} + \frac{\varphi'(f_d(x)) f_d''(x)}{\varphi(f_d(x))} - \left(\frac{\varphi'(f_d(x)) f_d'(x)}{\varphi(f_d(x))} \right)^2 \right),
\end{aligned}$$

where $x = \Phi^{-1}(y)$. Plugging in $\widehat{y}_d$ we get that

$$g_d'(\widehat{y}_d) = f_d'(\widehat{x}_d) = -1, \qquad g_d''(\widehat{y}_d) = 0, \tag{7.38}$$

and hence

$$\begin{aligned}
g_d'''(\widehat{y}_d) &= -\frac{1}{\varphi(\widehat{x}_d)} \left(-f_d'''(\widehat{x}_d) - f_d''(\widehat{x}_d)^2 + \frac{\varphi'(\widehat{x}_d) f_d''(\widehat{x}_d)}{\varphi(\widehat{x}_d)} \right) && \text{(by (7.38))} \\
&= -\frac{1}{\varphi(\widehat{x}_d)} \left(-f_d'''(\widehat{x}_d) - \frac{3 f_d''(\widehat{x}_d)^2}{2} \right) && \text{(by (7.37))} \\
&= -\frac{1}{2\varphi(\widehat{x}_d)} \left(\frac{\beta^2}{(\beta\widehat{x}_d + 1)^2} + \frac{1}{(\widehat{x}_d + \gamma)^2} - \frac{2\beta}{(\beta\widehat{x}_d + 1)(\widehat{x}_d + \gamma)} - \frac{1}{\widehat{x}_d^2} \right) \\
& && \text{(by (7.34) and (7.35))} \\
&= -\frac{1}{2\varphi(\widehat{x}_d)} \left(\left(\frac{\beta}{\beta\widehat{x}_d + 1} - \frac{1}{\widehat{x}_d + \gamma} \right)^2 - \frac{1}{\widehat{x}_d^2} \right) \\
&= \frac{1}{2\varphi(\widehat{x}_d)\widehat{x}_d^2} \left(1 - \frac{1}{d^2} \right). && \text{(by (7.36))}
\end{aligned}$$

It is easy to see that $\frac{1}{\varphi(x)x^2}$ is decreasing in x. Let $C := \frac{1}{2\varphi(\lambda\beta^d)\lambda^2\beta^{2d}}\left(1-\frac{1}{d^2}\right) > 0$ and $C' := \frac{\gamma^{2d}}{2\varphi(\lambda/\gamma^d)\lambda^2}\left(1-\frac{1}{d^2}\right) > 0$ be two constants. Since $\widehat{x}_d \in (\lambda\beta^d, \lambda/\gamma^d)$, we have that

$$C' < g_d'''(\widehat{x}_d) < C. \tag{7.39}$$

The Taylor expansion $g_d(y)$ at $y = \widehat{y}_d$ with the Peano reminder is

$$\begin{aligned} g_d(y) &= g_d(\widehat{y}_d) + g_d'(\widehat{y}_d)(y-\widehat{y}_d) + \frac{g_d''(\widehat{y}_d)}{2}(y-\widehat{y}_d)^2 + \frac{g_d'''(\widehat{y}_d)}{6}(y-\widehat{y}_d)^3 + h(y)(y-\widehat{y}_d)^4 \\ &= 2\widehat{y}_d - y + \frac{g_d'''(\widehat{y}_d)}{6}(y-\widehat{y}_d)^3 + h(y)(y-\widehat{y}_d)^4, \end{aligned} \tag{7.40}$$

where $h(y)$ is a continuous function such that $\lim_{y\to\widehat{y}_d} h(y) = 0$.

We will now consider an infinite d-ary tree $\widehat{\mathbb{T}}_d$ with arbitrary but uniform boundary condition. That is to say, we have $y_0 \geqslant 0$ as the starting point of our recursion and $y_{k+1} = g_d(y_k)$. If $y_k > \widehat{y}_d$, then $y_{k+1} < \widehat{y}_d$. Similarly for the case $y_k < \widehat{y}_d$. Moreover, if for some integer $k_0 > 0$, $|y_{k_0} - \widehat{y}_d| < \delta$, then for all $k \geqslant k_0$, $|y_k - \widehat{y}_d| < \delta$.

Theorem 7.44. $|y_n - \widehat{y}_d| = \Omega(1/\sqrt{n})$.

Proof. Let $\varepsilon_n = |y_n - \widehat{y}_d|$. Notice that $\lim_{n\to\infty} \varepsilon_n = 0$. We first show that $\varepsilon_n = O(1/\sqrt{n})$. By (7.40), it is easy to see that

$$\begin{aligned} \varepsilon_{n+1} &= |y_{n+1} - \widehat{y}_d| = |g_d(y_n) - \widehat{y}_d| \\ &= \left| 2\widehat{y}_d - y_n + \frac{g_d'''(\widehat{y}_d)}{6}(y_n-\widehat{y}_d)^3 + h(y_n)(y_n-\widehat{y}_d)^4 - \widehat{y}_d \right| \\ &= \varepsilon_n \left| 1 - \varepsilon_n^2 \left(\frac{g_d'''(\widehat{y}_d)}{6} \pm h(y_n)\varepsilon_n \right) \right|. \end{aligned} \tag{7.41}$$

As $\lim_{n\to\infty} h(y_n) = 0$ and $\lim_{n\to\infty} \varepsilon_n = 0$, there exists a constant N_0 such that for all $n \geqslant N_0$,

$$|h(y_n)\varepsilon_n| < \frac{C'}{12} \text{ and } \varepsilon_n^2 < \frac{12}{2C + C'}.$$

Hence for any $n \geqslant N_0$, we can use (7.39) to simplify (7.41) and get

$$\begin{aligned}\varepsilon_{n+1} &= \varepsilon_n \left(1 - \varepsilon_n^2 \left(\frac{g_d'''(\widehat{y}_d)}{6} \pm h(y_n)\varepsilon_n\right)\right) \\ &< \varepsilon_n(1 - C_1\varepsilon_n^2),\end{aligned}$$

where $C_1 = \frac{C'}{12}$. Let $A_{N_0} = \frac{1}{\sqrt{C/6+C_1}}$, and for any $n \geqslant N_0$, $A_{n+1} = A_n(1 - C_1 A_n^2)$. Notice that $A_{N_0}^2 = \frac{12}{2C+C'} > \varepsilon_{N_0}^2$. This will be our induction bases. It is easy to see by induction that $\varepsilon_n \leqslant A_n$ for all $n \geqslant N_0$. Moreover, by induction, we can verify that for any $n \geqslant N_0$, $A_n \leqslant \frac{1}{\sqrt{2C_1(n-N_0)+C/6+C_1}}$. Therefore $\varepsilon_n = O(1/\sqrt{n})$.

Next we lower bound ε_n. Again we use (7.39) and (7.41). It holds that, for all $n \geqslant N_0$,

$$\begin{aligned}\varepsilon_{n+1} &= \varepsilon_n \left(1 - \varepsilon_n^2 \left(\frac{g_d'''(\widehat{x}_d)}{6} \pm h(x_n)\varepsilon_n\right)\right) \\ &> \varepsilon_n(1 - C_2\varepsilon_n^2),\end{aligned}$$

where $C_2 = C/6 + C_1$. We now pick another constant $N_1 > N_0$ such that for all $n \geqslant N_1$,

$$\varepsilon_n < \frac{1}{\sqrt{2C_2}}.$$

We define $B_{N_1} = \varepsilon_{N_1}$ and for all $n \geqslant N_1$, $B_{n+1} = B_n(1 - C_2 B_n)$. By induction we can show that $\varepsilon_n \geqslant B_n$. Moreover, we can verify by induction that $B_n \geqslant \left(4C_2(n - N_1) + \frac{1}{\varepsilon_{N_1}^2}\right)^{-1/2}$. This finishes our proof. □

7.9 Concluding and Bibliographic Remarks

Restrepo et al. [RST+13] showed a broader region beyond the uniqueness threshold for which the hardcore model exhibits strong spatial mixing in the grid lattice $\mathbb{Z}^2$. Their starting point is a special case of the potential function introduced in Section 7.4, and then they made some "educated guess" and solve certain optimization problems by intensive computer search. In the end, their potential function depends on specific structures of grid lattices. Under the name of message-passing, essentially the same potential method was also used by Sinclair et al. [SST12] to give an FPTAS for anti-ferromagnetic Ising models in bounded degree graphs below the uniqueness threshold. Their potential function is also a special case of the one introduced in Section 7.4.

The potential function introduced in Section 7.5 is from [LLY13, GLLY15]. The same potential function has been employed quite a few times in other situations. Notable examples include showing strong spatial mixing of 2-spin systems for graphs with bounded connective constant [SSY13, SSŠY15]. Moreover, its special case of $\beta = 0$ and $\gamma = 1$ was used to design approximate counting algorithms for monotone CNFs [LL15b].

When we choose the potential function in Section 7.5, the simplification of $f''_d(\widehat{x}_d)$ in (7.20) is not unique. Our particular choice is guided by the hope of eliminating the degree d. Different choices of simplifications resulted in different potential functions. For example, we can use (7.18) to rewrite $f''_d(\widehat{x}_d)$ as

$$f''_d(\widehat{x}_d) = \frac{1}{\widehat{x}_d} + \frac{1}{\widehat{x}_d + \gamma} + \frac{\beta}{\beta\widehat{x}_d + 1} = \frac{d+1}{d\widehat{x}_d} + \frac{2\beta}{\beta\widehat{x}_d + 1}.$$

Plugging this into (7.19) gives us that

$$(\log(\varphi(\widehat{x}_d)))' = -\frac{d+1}{2d\widehat{x}_d} - \frac{\beta}{\beta\widehat{x}_d + 1}.$$

If we impose the equation above to hold for all x, we can solve that

$$\varphi_1(x) = \frac{C}{x^{\frac{d+1}{2d}}(\beta x + 1)},$$

for some arbitrary constant C. This is the potential function used in [LLY12].

Weitz [Wei06] showed that for the hardcore model strong spatial mixing on an infinite Δ-regular tree implies the strong spatial mixing on graphs of maximum degree at most Δ (Theorem 2.3 in [Wei06]). Sinclair et al. [SST12] showed similar results for the anti-ferromagnetic Ising model (Theorem 2.8 in [SST12]). For the hardcore model, we have that $\beta = 0$, $\gamma < 1$. For the anti-ferromagnetic Ising mode, we have that $\beta = \gamma < 1$. Hence both cases satisfy the condition of Theorem 7.27. They can be seen as special cases of Theorem 7.27 combined with Theorem 7.25. In fact, they are consequences of the following fact.

Proposition 7.45. *For $0 \leqslant \beta, \gamma \leqslant 1$, strong spatial mixing on infinite Δ-regular tree implies strong spatial mixing and FPTAS of $Z_{\beta,\gamma,\lambda}(G)$ in graphs of maximum degree at most Δ.*

We may use some of our results to prove it, but it can be shown straightforwardly.

Proof. As noted before the FPTAS follows from strong spatial mixing. Given a graph G and vertex v, we construct the SAW tree $T = T_{SAW}(G, v)$. In T, for each vertex of k children with $k < \Delta - 1$, we can attach $\Delta - 1 - k$ dummy vertices as its children. Instead of fixing their spins, we fix marginal probabilities or distributions that will be used in the recursion of these dummy vertices. We want to fix it so that it has no effect on its parent in the recursion. Therefore for each dummy vertex v, we set $R_v = \frac{p_v}{1-p_v} = \frac{1-\gamma}{1-\beta}$ so that $\frac{\beta R_v+1}{R_v+\gamma} = 1$. As $0 \leqslant \beta, \gamma \leqslant 1$, we have that $R_v \geqslant 0$, and therefore $p_v = \frac{R_v}{1+R_v}$ satisfies $0 \leqslant p_v \leqslant 1$ and is a valid probability. Suppose that dummy vertices are $k+1$-th to $\Delta - 1$-th children of the parent. Then this choice satisfy our needs as

$$\begin{aligned} F_\Delta(R_1, \ldots, R_\Delta) &= \lambda \prod_{i=1}^{\Delta-1} \frac{\beta R_i + 1}{R_i + \gamma} = \lambda \prod_{i=1}^{k} \frac{\beta R_i + 1}{R_i + \gamma} \\ &= F_k(R_1, \ldots, R_k) \end{aligned}$$

where R_i is the ratio at the i-th child. Then we have appended T into a tree where the recursion of each step is with respect to $\Delta - 1$ children. Strong spatial mixing in Δ-regular trees implies the required correlation decay of such recursions. □

This appending method is used in both [Wei06] and [SST12]. Essentially the same idea has also appeared in the proof of Theorem 7.22. For the hardcore model dummy children are fixed to be unoccupied and for the anti-ferromagnetic Ising model dummy children are fixed with uniform distributions over the two spins. In both cases, the dummy children have no effect on their parent. However, it is easy to see that if $\gamma > 1 > \beta$, in order to have no effect on the parent, $R_v = \frac{1-\gamma}{1-\beta} < 0$ and therefore R_v does not induce a valid distribution. In fact, when $\gamma > 1$, by Theorem 7.28 the claim of Proposition 7.45 is no longer true.

Chapter 8

Phase Transitions and Computational Hardness

In this chapter, we give some complementing hardness results about anti-ferromagnetic 2-spin systems. Beyond the uniqueness threshold, #2SPIN(β, γ, λ) has no FPRAS unless **NP** = **RP**. We will then study anti-ferromagnetic 2-spin systems in bipartite graphs.

8.1 Phase Transitions in Anti-Ferromagnetic Systems

Recall that if $\mathrm{Ctr}(\beta, \gamma, \lambda, \Delta) > 1$, then the Gibbs measure in infinite $(\Delta+1)$-regular tree $\mathbb{T}_{\Delta+1}$ or infinite Δ-ary tree $\widehat{\mathbb{T}}_\Delta$ is not unique. There is a long line of research [DFJ02, MWW09, Sly10, GGS+14, GŠV12, SS14] studying the relationship between the phase transition of the uniqueness of Gibbs measures in $\widehat{\mathbb{T}}_\Delta$ and the approximation complexity of #Δ-2SPIN(β, γ, λ). The upshot [SS14, GŠV12] is that if $\mathrm{Ctr}(\beta, \gamma, \lambda, \Delta) > 1$, then #$\Delta$-2SPIN$(\beta, \gamma, \lambda)$ has no FPRAS unless **NP** = **RP**.

The core of the hardness reduction in all papers above except [SS14] is the analysis of gadgets based on random regular graphs. Most notably is the construction by Sly [Sly10]. For integers r, n, let $\mathcal{G}_n^r$ be the following graph distribution:

1. $\mathcal{G}_n^r$ is supported on bipartite graphs. The two parts of the bipartite graph are labelled by $+, -$ and each is partitioned as $V^\pi := U^\pi \cup W^\pi$ where $|U^\pi| = n$, $|W^\pi| = r$ for $\pi = \{+, -\}$. U denotes the set $U^+ \cup U^-$ and similarly W denotes the set $W^+ \cup W^-$.

2. To sample $G \sim \mathcal{G}_n^r$, sample uniformly and independently Δ matchings: (i) $(\Delta - 1)$ random perfect matchings between $U^+ \cup W^+$ and $U^- \cup W^-$, (ii) a random perfect matching between U^+ and U^-. The edge set of G is the union of the Δ matchings. Thus, vertices in U have degree Δ, while vertices in W have degree $\Delta - 1$.

The case $r = 0$ will also be important, in which case we denote the distribution as $\mathcal{G}_n$. Note that $\mathcal{G}_n$ is supported on bipartite Δ-regular graphs. Strictly speaking, $\mathcal{G}_n^r, \mathcal{G}_n$ are supported on multi-graphs, but it is well known that every statement that holds asymptotically almost surely on these spaces continues to hold asymptotically almost surely conditioned on the event that the graph is simple [JLR00].

For positive integers Δ, t and n where n is even and is at least $2t$, let T^- and T^+ be disjoint vertex sets of size t and let V^- be a size-$n/2$ superset of T^- and V^+ be a size-$n/2$ superset of T^+ which is disjoint from V^-. Let $T = T^- \cup T^+$ and $V(t, n) = V^- \cup V^+$. Let $\mathcal{G}(t, n, \Delta)$ be the set of bipartite graphs with vertex partition (V^-, V^+) in which every vertex has degree at most Δ and every vertex in T has degree at most $\Delta - 1$. We refer to the vertices in T as "terminals". Vertices in T^+ are "positive terminals" and vertices in T^- are "negative terminals".

The construction goes as follows. For constants $0 < \theta, \psi < 1/8$, let

$$m' := (\Delta - 1)^{\lfloor \theta \log_{\Delta-1} n \rfloor + 2\lfloor \frac{\psi}{2} \log_{\Delta-1} n \rfloor}.$$

Note that $m' = o(n^{1/4})$. First, sample G from the distribution $\mathcal{G}_n^{m'}$ conditioning on G being simple. Next, for $\pi \in \{+, -\}$, attach t disjoint $(\Delta - 1)$-ary trees of even depth ℓ (with $t = (\Delta - 1)^{\lfloor \theta \log_{\Delta-1} n \rfloor}$ and $\ell = 2\lfloor \frac{\psi}{2} \log_{\Delta-1} n \rfloor$) to W^π, so that every vertex in W^π is a leaf of exactly one tree (this is possible since $m' = |W| = t(\Delta - 1)^\ell$). Denote by T^π the roots of the trees, so that $|T^\pi| = t$. The trees do not share common vertices with the graph G, apart from the vertices in W. The final graph $\widetilde{G}$ is the desired gadget, where the terminals T are the roots of the trees. We denote the family of graphs that can be constructed this way by $\widetilde{\mathcal{G}}(t, n(t), \Delta)$. Note that the size of the construction is $(2 + o(1))n$ which is bounded above by a polynomial in t when Δ is a fixed constant. Moreover, any $\widetilde{G}$ drawn from $\widetilde{\mathcal{G}}(t, n(t), \Delta)$ is bipartite. The terminals of $\widetilde{G}$ are $T(G) = T^+ \cup T^-$, and T^+ and T^- are on distinct partitions of the bipartite graph.

The key property of Sly's gadget is what we call *nearly-independent phase-correlated spins.*

When the gadget G is drawn from $\mathcal{G}(t, n, \Delta)$, we use the notation $T(G)$ to refer to the set of terminals. Each configuration $\sigma\colon V(t,n) \to \{0,1\}$ is assigned a unique phase $Y(\sigma) \in \{-,+\}$. Roughly in our applications of the definitions below the phase of a configuration σ is π if V^π contains more vertices with spin 1 than does $V^{-\pi}$.

Consider the two extremal semi-translation-invariant Gibbs measures corresponding to the all 1's and all 0's boundary conditions. Let $0 < q^- < q^+ < 1$ be the two marginal probabilities of the root having spin 0 in these two measures. More precisely, one can define q^+, q^- as follows. For $s \in \{0,1\}$, let $q_{\ell,s}$ denote the probability that the root is assigned spin 0 in the Δ-ary tree of depth ℓ in the Gibbs distribution where the leaves are fixed to spin s. In standard terminology, fixing the configuration on the leaves to all 0's or all 1's is most commonly referred to as the $+,-$ boundary conditions, respectively (and hence the indices $+,-$ in our notation of q^+, q^-). It is not hard to show that $q_{2\ell,0}$ is decreasing in ℓ, while $q_{2\ell,1}$ is increasing. Let $q^+ := \lim_{\ell\to\infty} q_{2\ell,0}$ and let $q^- := \lim_{\ell\to\infty} q_{2\ell,1}$. The two quantities q^+ and q^- satisfy the standard tree recursion in the following sense. Let $r^\pi = \frac{q^\pi}{1-q^\pi}$ for $\pi \in \{-,+\}$, and $f_\Delta(x) = \lambda\left(\frac{\beta x+1}{x+\gamma}\right)^\Delta$. Then $r^+ = f_\Delta(r^-)$ and $r^- = f_\Delta(r^+)$

We define measures Q^+ and Q^-. Fix $0 < q^- < q^+ < 1$. For any positive integer t,

- Q^+ is the distribution on configurations $\tau\colon T \to \{0,1\}$ such that, for every $v \in T^+$, $\tau(v) = 0$ independently with probability q^+ and, for every $v \in V^-$, $\tau(v) = 0$ independently with probability q^-;
- Q^- is the distribution on configurations $\tau\colon T \to \{0,1\}$ such that, for every $v \in T^-$, $\tau(v) = 0$ independently with probability q^+ and, for every $v \in T^+$, $\tau(v) = 0$ independently with probability q^-.

To prove the hardness we need a gadget where the spins of the terminals are drawn from distributions close to Q^+ or Q^- conditioned on the phase $+$ or $-$.

Definition 8.1. *A tuple of parameters* $(\beta, \gamma, \lambda, \Delta)$ supports nearly-independent phase-correlated spins *if there are efficiently-approximable values* $0 < q^- < q^+ < 1$ *such that the following is true. There are functions* $n(t,\varepsilon)$*,* $m(t,\varepsilon)$*, and* $f(t,\varepsilon)$*, each of which is bounded from above by a polynomial in* t *and* ε^{-1}*, and for every* t *and* ε *there is a distribution on graphs in* $\mathcal{G}(t, n(t,\varepsilon), \Delta)$

such that a gadget $G = (V, E)$ *with terminals* T *can be drawn from the distribution within* $m(t, \varepsilon)$ *time, and the probability that the following inequalities hold is at least* $3/4$*:*

1. *The phases are roughly balanced, i.e.,*

$$\Pr_{G;\beta,\gamma,\lambda}(Y(\sigma) = +) \geqslant \frac{1}{f(t,\varepsilon)} \text{ and } \Pr_{G;\beta,\gamma,\lambda}(Y(\sigma) = -) \geqslant \frac{1}{f(t,\varepsilon)}. \tag{8.1}$$

2. *For a configuration* $\sigma\colon V \to \{0, 1\}$ *and any* $\tau\colon T \to \{0, 1\}$*,*

$$\left|\frac{\Pr_{G;\beta,\gamma,\lambda}(\sigma|_T = \tau \mid Y(\sigma) = +)}{Q^+(\tau)} - 1\right| \leqslant \varepsilon \text{ and } \left|\frac{\Pr_{G;\beta,\gamma,\lambda}(\sigma|_T = \tau \mid Y(\sigma) = -)}{Q^-(\tau)} - 1\right| \leqslant \varepsilon. \tag{8.2}$$

In fact, given a gadget with the above property, one can construct a gadget where the phases are (nearly) uniformly distributed as detailed in the following definition.

Definition 8.2. *We say that the tuple of parameters* $(\beta, \gamma, \lambda, \Delta)$ *supports* balanced *nearly-independent phase-correlated spins if Definition 8.1 holds with (8.1) replaced by:*

$$\Pr_{G;\beta,\gamma,\lambda}(Y(\sigma) = +) \geqslant \frac{1-\varepsilon}{2} \text{ and } \Pr_{G;\beta,\gamma,\lambda}(Y(\sigma) = -) \geqslant \frac{1-\varepsilon}{2}, \tag{8.3}$$

where ε *is quantified as in Definition 8.1.*

Balanced phases are very important in later AP-reductions. Later in Lemma 8.7 we will need to assume the existence of a gadget with balanced phases. The following lemma shows that for essentially all 2-spin systems, Definition 8.1 implies Definition 8.2.

Lemma 8.3. *If the parameter tuple* $(\beta, \gamma, \lambda, \Delta)$ *with* $\beta\gamma \neq 1$ *supports nearly-independent phase-correlated spins, then it supports balanced nearly-independent phase-correlated spins.*

Sly [Sly10, Theorem 2.1] showed that for every $\Delta \geqslant 3$, there exists ε_Δ such that the hard-core model $(1, 0, \lambda, \Delta)$ supports nearly-independent phase-correlated spins for any λ satisfying $\lambda_c(\widehat{\mathbb{T}}_\Delta) < \lambda < \lambda_c(\widehat{\mathbb{T}}_\Delta) + \varepsilon_\Delta$. This region is a small interval just above the uniqueness threshold. In the same paper Sly also showed that $(1, 0, 1, 6)$ supports nearly-independent phase-correlated spins. The quantities defining measures $Q^\pm$ in Definition 8.1 are exactly the marginal probabilities $q^\pm$ on $\widehat{\mathbb{T}}_\Delta$.

Galanis et al. [GGS+14] extended the applicable region of Sly's gadget for the hard-core model to all $\lambda > \lambda_c(\mathbb{T}_\Delta)$ for $\Delta = 3$ and $\Delta \geqslant 6$. The gap of $\Delta = 4$ and $\Delta = 5$ is later closed [GŠV12]. In the later paper Galanis et al. [GŠV12] also verified Sly's gadget for parameters $(\beta, \gamma, \lambda, \Delta)$ such that $\beta\gamma < 1$, $\sqrt{\beta\gamma} \geqslant \frac{\sqrt{\Delta-1}-1}{\sqrt{\Delta-1}+1}$, and (β, γ, λ) is in the non-uniqueness region of infinite tree $\mathbb{T}_\Delta$. Using techniques from [GŠV14] the applicable region of Sly's gadget is extended to the entire non-uniqueness region for all 2-spin anti-ferromagnetic models [CGG+14].

Lemma 8.4. *For all* $\Delta \geqslant 3$*, all* $\beta, \gamma, \lambda > 0$ *where* $\beta\gamma < 1$*, if* $\mathrm{Ctr}(\beta, \gamma, \lambda, \Delta) > 1$ *then the tuple of parameters* $(\beta, \gamma, \lambda, \Delta)$ *supports nearly-independent phase-correlated spins.*

Roughly speaking, if a set of parameters $(\beta, \gamma, \lambda, \Delta)$ supports balanced nearly-independent phase-correlated spins, then #SAT $\leqslant_{\mathrm{AP}}$ #Δ-2SPIN(β, γ, λ).

8.2 Spin Systems in Bipartite Graphs

As we have seen, for example in Theorem 6.17, counting problems can usually be classified into tractable and hard classes. Interestingly, in approximate counting, there has emerged a third distinct class of natural problems, which seems to be of intermediate complexity. It is conjectured [DGGJ03] that the problems in this class do not have an FPRAS but that they are not as hard as #SAT to approximate. A canonical problem in this class has been identified, which is to count the number of independent sets in a bipartite graph (#BIS). Despite many attempts, nobody has found an FPRAS for #BIS nor an AP-reduction from #SAT to #BIS. The conjecture is that neither exists. Mossel et al. [MWW09] showed that the Gibbs sampler for sampling independent sets in bipartite graphs mixes slowly even if degrees are at most 6. Another interesting attempted Markov Chain for #BIS by Ge and Stefankovic [GŠ12] was also shown later to be slowly mixing by Goldberg and Jerrum [GJ12b].

Name #BIS

Instance A bipartite graph B.

Output The number of independent sets in B.

#BIS plays an important role in classifying counting problems with respect to approximation. A trichotomy theorem is shown for the complexity of approximately solving unweighted Boolean counting CSPs, where in addition to problems that are solvable by FPRASes and those that are AP-reducible from #SAT, there is the intermediate class of problems which are equivalent to #BIS [DGJ10]. Many counting problems are shown to be #BIS-hard and hence are conjectured to have no FPRAS [BDG+13, CDG+15], including estimating the partition function of the ferromagnetic Potts model [GJ12a]. Moreover, under AP-reductions #BIS is complete in a logically defined class of problems, called $\#RH\Pi_1$, to which an increasing variety of problems have been shown to belong. Other typical complete problems in $\#RH\Pi_1$ include counting the number of downsets in a partially ordered set [DGGJ03] and computing the partition function of the ferromagnetic Ising model with local external fields [GJ07].

Motivated by #BIS, in the rest of this chapter, we will focus on 2-spin systems over bounded or unbounded degree bipartite graphs parametrized by a tuple (β, γ, λ). For efficiently approximable non-negative real numbers β, γ, λ and a positive integer Δ, we define the problem of computing the partition function of the 2-spin system (β, γ) with external field λ on bipartite graphs of bounded degree Δ, as follows.

Name #(Δ-)BI-(M-)2SPIN(β, γ, λ)

Instance A bipartite (multi)graph $B = (V, E)$ (with degree bound Δ).

Output The quantity $Z_{\beta,\gamma,\lambda}(B)$.

Notice that we in fact introduced four problems, depending on whether there is a degree bound and whether the graph is simple.

We found the notion of non-uniform external field useful in the reductions. The following problems are introduced as intermediate problems. We also introduce its multigraph version, but as intermediate problems we do not need the bounded degree variant.

Name #BI-(M-)NONUNIFORM-2SPIN(β, γ, λ)

Instance A bipartite (multi)graph $B = (V, E)$ and a subset $U \subseteq V$.

Output The quantity

$$Z_{\beta,\gamma,\lambda}(B;U) = \sum_{\sigma:V\to\{0,1\}^{|V|}} \lambda^{\sum_{v\in U} 1-\sigma(v)} \prod_{(v,u)\in E} \beta^{(1-\sigma(v))(1-\sigma(u))}\gamma^{\sigma(v)\sigma(u)}.$$

It turns out that the notion of nearly-independent phase-correlated spins (Definitions 8.1 and 8.2) is not sufficient to imply #BIS-hardness. In order to study spin systems in bipartite graphs, we want the gadget to satisfy another property, which we call *symmetry breaking*.

Definition 8.5. *We say that a tuple of parameters* $(\beta,\gamma,\lambda,\Delta)$ supports symmetry-breaking *if there is a bipartite graph* H *whose vertices have degree at most* Δ *which has a distinguished degree-*1 *vertex* v_H *such that* $\Pr_{H;\beta,\gamma,\lambda}(\sigma_{v_H} = 0) \notin \{0, \lambda/(1+\lambda), 1\}$.

We will prove in Section 8.4 that symmetry breaking holds for all 2-spin models except for two cases.

Lemma 8.6. *Assume* $\Delta \geqslant 3$*. The parameters* $(\beta,\gamma,\lambda,\Delta)$ *support symmetry breaking unless* (i) $\beta\gamma = 1$*, or* (ii) $\beta = \gamma$ *and* $\lambda = 1$*.*

Lemma 8.7. *Suppose a set of parameters* $(\beta,\gamma,\lambda,\Delta)$ *with* $\beta\gamma \neq 1$ *and* $\Delta \geqslant 3$ *supports balanced nearly-independent phase-correlated spins and symmetry-breaking. Then* #Δ-Bi-2Spin(β,γ,λ) *is* #BIS*-hard to approximate.*

Let $\mathrm{Ctr}(\beta,\gamma,\lambda,\infty) := \max_{d\geqslant 1} \mathrm{Ctr}(\beta,\gamma,\lambda,d)$. By Lemma 7.30, if $\beta,\gamma < 1$, then $\mathrm{Ctr}(\beta,\gamma,\lambda,\infty) = \infty$. Combining Lemma 8.7 with Lemmas 8.4, 8.3, and 8.6, we get the following theorem.

Theorem 8.8. *Let* (β,γ,λ) *be a set of parameters such that* $\beta\gamma < 1$ *and* $\Delta \geqslant 3$ *be an integer or* $\Delta = \infty$*. Then* #Δ-Bi-2Spin(β,γ,λ) *is* #BIS*-hard to approximate unless* $\mathrm{Ctr}(\beta,\gamma,\lambda,d) \leqslant 1$ *for all integers* $d \in [1,\Delta)$ *or* $\beta = \gamma$ *and* $\lambda = 1$*.*

We note that if $\mathrm{Ctr}(\beta,\gamma,\lambda,d) < 1$ for all integers $d \in [1,\Delta)$, then #Δ-2Spin(β,γ,λ) has an FPTAS by Theorem 7.25 or Theorem 7.26 and so does #Δ-Bi-2Spin(β,γ,λ). If $\beta = \gamma < 1$ and $\lambda = 1$, then we can reduce #Bi-2Spin$(\beta,\beta,1)$ to ferromagnetic Ising models without external fields, which has an FPTAS by Jerrum and Sinclair [JS93]. Hence #Bi-2Spin$(\beta,\beta,1)$ also has an FPTAS. Details can be found in Corollary 8.10.

8.3 Balancing Nearly-Independent Phase-Correlated Spins

In this section we prove Lemma 8.3 that a gadget with nearly-independent phases can be used to construct a gadget with balanced phases.

Before proving the lemma, we introduce some notations. Let q^+ and q^- be the quantities from Definition 8.1. Let

$$M := \begin{bmatrix} \beta & 1 \\ 1 & \gamma \end{bmatrix} \quad \text{and} \quad M^+ := \begin{bmatrix} q^+ & 1-q^+ \\ q^- & 1-q^- \end{bmatrix}.$$

The two columns of M^+ correspond to spin 0 and spin 1. The first row corresponds to the distribution induced on a positive terminal from Q_t^+ and the second to the distribution induced from Q_t^-. Similarly the first row also corresponds to the distribution induced on a negative terminal from Q_t^- and the second to the distribution induced from Q_t^+. Notice that $\det(M^+) = q^+ - q^- > 0$.

When the parameters β, γ and λ are clear from the context, we make the notation more concise, by referring to the partition function as Z_G rather than as $Z_{\beta,\gamma,\lambda}(G)$. Also, given a configuration $\sigma\colon V(G) \to \{0,1\}$ and a subset S of $V(G)$, we often use the notation σ_S to denote the restriction $\sigma|_S$. For a gadget G drawn from $\mathcal{G}(t, n(t,\varepsilon), \Delta)$, let Z_G^π be the contribution of phase $\pi \in \{-,+\}$ to the partition function Z_G. Moreover, for a subset $S \subseteq T(G)$, suppose $\tau_S\colon T(G) \to \{0,1\}$ is a configuration on terminals in S. Let $Z_G^\pi(\tau_S)$ be the contribution of configurations that are consistent with τ_S and belong to phase π, that is,

$$Z_G^\pi(\tau_S) = \sum_{\substack{\sigma\colon Y(\sigma)=\pi \\ \sigma_S=\tau_S}} w(\sigma),$$

where $w(\sigma)$ is the weight of configuration σ defined in (7.1). It is easy to see that for $\pi \in \{-,+\}$,

$$\Pr_{G;\beta,\gamma,\lambda}(Y(\sigma)=\pi) = \frac{Z_G^\pi}{Z_G},$$

and

$$\Pr_{G;\beta,\gamma,\lambda}(\sigma_{T(G)} = \tau_{T(G)} \mid Y(\sigma)=\pi) = \frac{Z_G^\pi(\tau_{T(G)})}{Z_G^\pi}.$$

We are now prepared to prove the lemma which is the focus of this section.

Proof of Lemma 8.3. Let ε satisfy $0 < \varepsilon < 1$. By assumption we may draw a gadget G from $\mathcal{G}(t+t', n(t+t', \varepsilon'), \Delta)$ such that it satisfies (8.1) and (8.2) with probability at least 3/4, where t' and ε' will be specified later. Assume G does. Otherwise the construction fails.

We consider first the anti-ferromagnetic case $\beta\gamma < 1$. We construct a gadget K such that K satisfies (8.3) and (8.2). We make two copies of G, say G_1 and G_2. Let the terminals of G_i be $T(G_i) = T^+(G_i) \cup T^-(G_i)$ for $i = 1, 2$. For each $\pi \in \{-, +\}$, we add a set of edges that form a perfect matching between t' terminals in $T^\pi(G_1)$ and t' terminals in $T^\pi(G_2)$. Denote by P the edges of the two perfect matchings.K is the resulting graph. Denote by C_i the vertices of G_i that are endpoints of P. The terminals of K are those $2t$ terminal nodes in $T(G_1)$ that are still unmatched, that is $T^\pi(K) = T^\pi(G_1)\backslash C_1$ for $\pi \in \{-, +\}$, and $T(K) = T^+(K) \cup T^-(K)$. Denote by I the terminals of G_2 that are unmatched.

We define the phase of K to be the phase of G_1, that is, K is said to have phase $+$ or $-$ if and only if G_1 has the same $+$ or $-$ phase regardless of the phase of G_2. Let (π_1, π_2) be a vector denoting the phases of G_1 and G_2 where $\pi_1, \pi_2 \in \{-, +\}$. Then the $+$ phase of K corresponds to the vector $\{(+, +), (+, -)\}$ and the $-$ phase corresponds to $\{(-, +), (-, -)\}$. For two configurations τ_V and τ_U with $V \cap U = \emptyset$, let (τ_V, τ_U) be the joint configuration on $V \cup U$. Then we have the following:

$$\begin{aligned} Z_K^\pi(\tau_{T(K)}) &= \sum_{\tau_{C_1}, \tau_{C_2}, \tau_I} Z_{G_1}^\pi(\tau_{T(K)}, \tau_{C_1}) \left(Z_{G_2}^+(\tau_{C_2}, \tau_I) w(\tau_{C_1}, \tau_{C_2}) + Z_{G_2}^-(\tau_{C_2}, \tau_I) w(\tau_{C_1}, \tau_{C_2}) \right) \\ &= \sum_{\tau_{C_1}, \tau_{C_2}} Z_{G_1}^\pi(\tau_{T(K)}, \tau_{C_1}) w(\tau_{C_1}, \tau_{C_2}) \left(Z_{G_2}^+(\tau_{C_2}) + Z_{G_2}^-(\tau_{C_2}) \right), \end{aligned} \tag{8.4}$$

where $w(\tau_{C_1}, \tau_{C_2})$ denote the contribution from edges of P given configurations τ_{C_1} and τ_{C_2}, and $Z_K^\pi = \sum_{\tau_{T(K)}} Z_K^\pi(\tau_{T(K)})$. Moreover by (8.2), for $i = 1, 2$ and any subset $S \subseteq T(G_i)$, we have

$$(1 - \varepsilon')Q^\pi(\tau_S)Z_{G_i}^\pi \leqslant Z_{G_i}^\pi(\tau_S) \leqslant (1 + \varepsilon')Q^\pi(\tau_S)Z_{G_i}^\pi,$$

where we have used $Q^\pi(\tau_S)$ to denote the probability that the configuration on terminals in S

is τ_S in the distribution Q^{π}. Therefore by (8.4),

$$Z_K^+(\tau_{T(K)}) \leqslant (1+\varepsilon')^2 Q^+(\tau_{T(K)}) Z_{G_1}^+ Z_{G_2}^- \sum_{\tau_{C_1},\tau_{C_2}} Q^+(\tau_{C_1}) w(\tau_{C_1},\tau_{C_2}) \left(\frac{Z_{G_2}^+}{Z_{G_2}^-} Q^+(\tau_{C_2}) + Q^-(\tau_{C_2}) \right). \tag{8.5}$$

We need to calculate the quantity

$$\mu(\pi_1,\pi_2) := \sum_{\tau_{C_1},\tau_{C_2}} Q^{\pi_1}(\tau_{C_1}) Q^{\pi_2}(\tau_{C_2}) w(\tau_{C_1},\tau_{C_2}). \tag{8.6}$$

Recall our definitions of M and M^+. Let $N = M^+ M (M^+)^T$ where the superscript T means transposition. Then $\det(N) = (q^+ - q^-)^2(\beta\gamma - 1) < 0$. We write

$$N = \begin{bmatrix} N_{++} & N_{-+} \\ N_{+-} & N_{--} \end{bmatrix}$$

and let $c = \frac{N_{++}N_{--}}{N_{+-}N_{-+}} < 1$. Here $N_{\pi_1\pi_2}$ is the edge contribution when one end point is chosen with probability q^{π_1} and the other q^{π_2}. Also notice that each edge is independent under $Q^{\pm}$ so we can count them separately. Then the quantity in (8.6) is

$$\begin{aligned} \mu(+,+) &= \mu(-,-) = (N_{++}N_{--})^{t'}; \\ \mu(+,-) &= \mu(-,+) = (N_{+-}N_{-+})^{t'}. \end{aligned} \tag{8.7}$$

Plug (8.7) in (8.5),

$$\begin{aligned} Z_K^+(\tau_{T(K)}) &\leqslant (1+\varepsilon')^2 Z_{G_1}^+ Z_{G_2}^- \left(\frac{Z_{G_2}^+}{Z_{G_2}^-} (N_{++}N_{--})^{t'} + (N_{+-}N_{-+})^{t'} \right) \cdot Q^+(\tau_{T(K)}) \\ &= (1+\varepsilon')^2 Z_{G_1}^+ Z_{G_2}^- (N_{+-}N_{-+})^{t'} \left(\frac{Z_{G_2}^+}{Z_{G_2}^-} c^{t'} + 1 \right) \cdot Q^+(\tau_{T(K)}). \end{aligned} \tag{8.8}$$

Summing over $\tau_{T(K)}$ in (8.8) we get

$$Z_K^+ \leqslant (1+\varepsilon')^2 Z_{G_1}^+ Z_{G_2}^- (N_{+-}N_{-+})^{t'} \left(\frac{Z_{G_2}^+}{Z_{G_2}^-} c^{t'} + 1 \right). \tag{8.9}$$

Similarly we get an estimate for Z_K^-:

$$Z_K^- \geqslant (1-\varepsilon')^2 Z_{G_1}^- Z_{G_2}^+ (N_{+-}N_{-+})^{t'} \left(\frac{Z_{G_2}^-}{Z_{G_2}^+} c^{t'} + 1 \right). \tag{8.10}$$

Let $r = \frac{Z_{G_2}^-}{Z_{G_2}^+}$. Notice that $Z_{G_1}^\pi = Z_{G_2}^\pi$ as G_1 and G_2 are identical copies. Combine (8.9) and (8.10),

$$\frac{Z_K^-}{Z_K^+} \geqslant \left(\frac{1-\varepsilon'}{1+\varepsilon'} \right)^2 \cdot \frac{1 + c^{t'} r}{1 + c^{t'}/r}. \tag{8.11}$$

By (8.1) there is $f(t+t', \varepsilon')$ such that

$$\frac{1}{f(t+t', \varepsilon')} \leqslant r \leqslant f(t+t', \varepsilon'),$$

and $f(t+t', \varepsilon')$ is bounded above by a polynomial in $t+t'$ and $1/\varepsilon'$. To show (8.3) it suffices to show $\frac{Z_K^-}{Z_K^+} \geqslant \frac{1-\varepsilon}{1+\varepsilon}$ and $\frac{Z_K^+}{Z_K^-} \geqslant \frac{1-\varepsilon}{1+\varepsilon}$. Clearly $\frac{Z_K^-}{Z_K^+} \geqslant \frac{1-\varepsilon}{1+\varepsilon}$ can be achieved by picking $\varepsilon' = \frac{\varepsilon}{3}$ and $t' = O(\log(t + \varepsilon^{-1}))$ in (8.11). To show $\frac{Z_K^+}{Z_K^-} \geqslant \frac{1-\varepsilon}{1+\varepsilon}$ is similar and therefore omitted.

Establishing (8.2) is easy. We will show

$$\frac{Z_K^+(\tau_{T(K)})}{Z_K^+} \geqslant (1-\varepsilon) Q^+(\tau_{T(K)}),$$

and the other bounds are similar. By an argument similar to (8.5), we have

$$Z_K^+(\tau_{T(K)}) \geqslant (1-\varepsilon')^2 Q^+(\tau_{T(K)}) Z_{G_1}^+ Z_{G_2}^- \sum_{\tau_{C_1}, \tau_{C_2}} Q^+(\tau_{C_1}) w(\tau_{C_1}, \tau_{C_2}) \left(\frac{Z_{G_2}^+}{Z_{G_2}^-} Q^+(\tau_{C_2}) + Q^-(\tau_{C_2}) \right).$$

Moreover, summing over $\tau_{T(K)}$ in (8.5) we get:

$$Z_K^+ \leqslant (1+\varepsilon')^2 Z_{G_1}^+ Z_{G_2}^- \sum_{\tau_{C_1}, \tau_{C_2}} Q^+(\tau_{C_1}) w(\tau_{C_1}, \tau_{C_2}) \left(\frac{Z_{G_2}^+}{Z_{G_2}^-} Q^+(\tau_{C_2}) + Q^-(\tau_{C_2}) \right).$$

The desired bound follows as $\varepsilon' = \frac{\varepsilon}{3}$.

The other case is ferromagnetic, that is, $\beta\gamma > 1$. We construct K in the same way as in the previous case, but with the following change. To form the perfect matching P, we match $+/-$ terminals of G_1 to $-/+$ terminals of G_2. The proof goes similarly but $\det(N) = (q^+ - q^-)^2(\beta\gamma -$

$1) > 0$. However since we made a twist in connecting G_1 and G_2, it follows that

$$\mu(+,+) = \mu(-,-) = (N_{+-}N_{-+})^{t'};$$
$$\mu(+,-) = \mu(-,+) = (N_{--}N_{++})^{t'}.$$

Therefore we continue with the new constant $c' = \frac{N_{+-}N_{-+}}{N_{++}N_{--}} < 1$ and the rest of the proof is the same. □

8.4 Symmetry Breaking

In this section we prove Lemma 8.6 that almost all 2-spin models support symmetry breaking.

Proof of Lemma 8.6. Consider the sequence of gadgets $(H_k : k \in \mathbb{N})$, defined as follows. The vertex set of H_k is $V(H_k) = \{u, u', u'', v_1, v_2, \dots, v_k\}$, and u is considered the attachment vertex. The edge set of H_k is

$$E(H_k) = \big\{\{u', v_i\} : 1 \leqslant i \leqslant k\big\} \cup \big\{\{v_i, u''\} : 1 \leqslant i \leqslant k\big\} \cup \big\{\{u'', u\}\big\}.$$

We shall argue that if the first three graphs in the sequence, namely H_0, H_1 and H_2, all fail to be symmetry breakers then one of conditions (i) or (ii) holds. The graph H_0 has an isolated vertex that could clearly be removed; we leave it in to make the calculations uniform. Note that the maximum degree of any vertex in these graphs is 3.

Let $a = \beta^2 + \lambda$, $b = \beta + \lambda\gamma$ and $c = 1 + \lambda\gamma^2$. Then the effective weight of vertex u is given by the column vector,

$$\rho^k = \begin{bmatrix} \lambda & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \beta & 1 \\ 1 & \gamma \end{bmatrix} \begin{bmatrix} \lambda & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} a^k & b^k \\ b^k & c^k \end{bmatrix} \begin{bmatrix} \lambda \\ 1 \end{bmatrix} = T \begin{bmatrix} \lambda a^k + b^k \\ \lambda b^k + c^k \end{bmatrix}$$

where

$$T = \begin{bmatrix} \lambda^2\beta & \lambda \\ \lambda & \gamma \end{bmatrix}.$$

For h_k to be a symmetry breaker we require the vector ρ^k not to be a multiple of $\begin{bmatrix} \lambda \\ 1 \end{bmatrix}$.

Suppose ρ^0, ρ^1 and ρ^2 all fail to be symmetry breakers. Then they must all lie in a one-dimensional subspace of $\mathbb{R}^2$. One way this can happen is if the matrix T is rank 1, i.e., if $\beta\gamma = 1$. This is case (i). Otherwise, $\begin{bmatrix} 1+\lambda \\ 1+\lambda \end{bmatrix}$ and $\begin{bmatrix} \lambda a + b \\ \lambda b + c \end{bmatrix}$ and $\begin{bmatrix} \lambda a^2 + b^2 \\ \lambda b^2 + c^2 \end{bmatrix}$ lie in a one-dimensional subspace, namely the one generated by $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$. This implies $a+\lambda b = b+\lambda c$ and $a^2+\lambda b^2 = b^2+\lambda c^2$, or recasting,

$$a - b = \lambda(c - b)$$

$$(a - b)(a + b) = \lambda(c - b)(c + b).$$

So either $a = b = c$, or (dividing the second equation by the first) $a = c$ and $\lambda = 1$. In either case, substituting for a, b and c in terms of β, γ, λ, we obtain either $\beta = \gamma = 1$ (which belongs to case (i)) or $\beta = \gamma$ and $\lambda = 1$ (which is case(ii)). □

There are two exceptional cases. The first is $\beta\gamma = 1$. It is well-known that in this case the 2-spin system can be decomposed and hence tractable. The other case of $\beta = \gamma$ and $\lambda = 1$ is perfectly symmetric and this symmetry cannot be broken. This system is the Ising model without external fields. For this system, the marginal probability of any vertex in any graph is exactly $1/2$.

Regarding the ferromagnetic case, Jerrum and Sinclair [JS93] presented an FPRAS for the ferromagnetic Ising model with consistent external fields. On the other hand, anti-ferromagnetic Ising models without external field on bipartite graphs are actually equivalent to ferromagnetic Ising models. The trick is to flip the assignment on only one part of the bipartition, which has been used before by Goldberg and Jerrum [GJ07].

Lemma 8.9. *For* $0 < \alpha < 1$, $\#\text{BI-2SPIN}(\alpha, \alpha, 1) \equiv_T \#\text{BI-2SPIN}(1/\alpha, 1/\alpha, 1)$.

Proof. Let $B = (V_1, V_2, E)$ be a bipartite graph where V_1 and V_2 are two partitions of vertices.

Let $|E| = m$. Then we have

$$\begin{aligned}
Z_B(\alpha,\alpha,1) &= \sum_{\sigma_{V_1}:V_1\to\{0,1\}^{|V_1|}} \sum_{\sigma_{V_2}:V_2\to\{0,1\}^{|V_2|}} \prod_{(v,w)\in E} \alpha^{(1-\sigma_{V_1}(v))(1-\sigma_{V_2}(w))}\alpha^{\sigma_{V_1}(v)\sigma_{V_2}(w)} \\
&= \sum_{\sigma_{V_1}:V_1\to\{0,1\}^{|V_1|}} \sum_{\sigma_{V_2}:V_2\to\{0,1\}^{|V_2|}} \prod_{(v,w)\in E} \alpha^{(1-\sigma_{V_1}(v))\sigma_{V_2}(w)}\alpha^{\sigma_{V_1}(v)(1-\sigma_{V_2}(w))} \\
&= \sum_{\sigma_{V_1}:V_1\to\{0,1\}^{|V_1|}} \sum_{\sigma_{V_2}:V_2\to\{0,1\}^{|V_2|}} \prod_{(v,w)\in E} \alpha^{\sigma_{V_2}(w)-\sigma_{V_1}(v)\sigma_{V_2}(w)+\sigma_{V_1}(v)-\sigma_{V_1}(v)\sigma_{V_2}(w))} \\
&= \alpha^m \sum_{\sigma_{V_1}:V_1\to\{0,1\}^{|V_1|}} \sum_{\sigma_{V_2}:V_2\to\{0,1\}^{|V_2|}} \prod_{(v,w)\in E} \alpha^{-(1-\sigma_{V_1}(v))(1-\sigma_{V_2}(w))-\sigma_{V_1}(v)\sigma_{V_2}(w)} \\
&= \alpha^m Z_B\left(\alpha^{-1},\alpha^{-1},1\right),
\end{aligned}$$

where in the second line we flip the assignment of σ_{V_2}. □

Combining Lemma 8.9 with the FPRAS by Jerrum and Sinclair [JS93] for the ferromagnetic Ising model yields the following corollary.

Corollary 8.10. *For any* $\alpha > 0$, $\#\textsc{Bi-2Spin}(\alpha,\alpha,1)$ *has an FPRAS.*

This corollary explains why the notion of symmetry breaking is necessary to achieve #BIS-hardness.

8.5 The Reduction from #BIS

In this section we show our main reduction, Lemma 8.7, namely that the two properties of "nearly-independent phase-correlated spins" and "symmetry-breaking" lead to #BIS-hardness. Lemma 8.7 follows directly from Lemma 8.11 and Lemma 8.12, which will be proved in the following two subsections.

An Intermediate Problem

The goal of this section is to show that it is #BIS-hard to approximate the partition function of anti-ferromagnetic Ising models with non-uniform non-trivial external fields on bipartite graphs.

Lemma 8.11. *For any* $0 < \alpha < 1$, $\lambda > 0$ *and* $\lambda \neq 1$,

$$\#\text{BIS} \leqslant_{AP} \#\text{Bi-M-Nonuniform-2Spin}(\alpha, \alpha, \lambda).$$

Proof. By flipping 0 to 1 and 1 to 0 for each configuration σ, we see that #Bi-M-Nonuniform-2Spin$(\alpha, \alpha, \lambda)$ is in fact the same as #Bi-M-Nonuniform-2Spin$(\alpha, \alpha, 1/\lambda)$. Hence we may assume that $\lambda > 1$.

Let $M := \left[\begin{smallmatrix} \alpha & 1 \\ 1 & \alpha \end{smallmatrix}\right]$, and $\left[\begin{smallmatrix} \rho_0 \\ \rho_1 \end{smallmatrix}\right] := M \left[\begin{smallmatrix} \lambda \\ 1 \end{smallmatrix}\right] = \left[\begin{smallmatrix} \alpha\lambda+1 \\ \alpha+\lambda \end{smallmatrix}\right]$. Since $\alpha < 1$ and $\lambda > 1$, $\rho_1 > \rho_0$.

Let $B = (V, E)$ be an input to #BIS with $n = |V|$ and $m = |E|$. Let I_B be the number of independent sets of B. Let ε be the desired accuracy of the reduction. We will construct an instance $B' = (V', E')$ with a specified vertex subset $U \subset V'$ for #Bi-M-Nonuniform-2Spin$(\alpha, \alpha, \lambda)$ such that

$$\exp\left(-\frac{\varepsilon}{2}\right) I_B \leqslant \frac{Z_{\alpha,\alpha,\lambda}(B'; U)}{C} \leqslant \exp\left(\frac{\varepsilon}{2}\right) I_B,$$

where C is a quantity that is easy to approximate. Therefore it suffices to call oracle #Bi-M-Nonuniform-2Spin$(\alpha, \alpha, \lambda)$ on B' with the specified subset U with accuracy $\frac{\varepsilon}{4}$ and approximate C within $\frac{\varepsilon}{4}$.

The construction of B' involves two positive integers t_1 and t_2. Let t_1 be the least positive integer such that

$$\alpha^{2t_1} \leqslant \frac{\varepsilon}{6 \cdot 2^n}. \tag{8.12}$$

Note that t_1 depends on n and ε and there is a polynomial p in n and ε^{-1} such that $t_1 \leqslant p(n, \varepsilon^{-1})$. Let t_2 be the least positive integer depending on n, ε and t_1 such that

$$\left(\frac{\rho_0}{\rho_1}\right)^{t_2} \leqslant \frac{\alpha^{t_1 m} \cdot \varepsilon}{6 \cdot 2^{2t_1 m+n}}. \tag{8.13}$$

Once again, t_2 is bounded from above by a polynomial in n and ε^{-1}.

Given the integers t_1 and t_2, the graph B' is constructed as follows. Let $W_v = \{w_{v,j} \mid 1 \leqslant j \leqslant t_1 \deg(v)\}$ for each $v \in V$ where $\deg(v)$ is the degree of v in B. Let $U_{v,j} = \{u_{v,j,k} \mid 1 \leqslant k \leqslant t_2\}$

for any $v \in V$ and $1 \leqslant j \leqslant t_1 \deg(v)$. Let

$$W = \bigcup_{v \in V} W_v \qquad \text{and} \qquad U = \bigcup_{v \in V} \bigcup_{1 \leqslant j \leqslant t_1 \deg(v)} U_{v,j}.$$

The vertex set of B' is $V' = V \cup U \cup W$. Note that $|W| = 2t_1 m$ and $|U| = 2t_1 t_2 m$.

We add t_1 parallel edges in B' between u and v for each $(u,v) \in E$ and add edges between v and every vertex in W_v, and between $w_{v,j}$ and every vertex in $U_{v,j}$ for each $v \in V$ and $1 \leqslant j \leqslant t_1 \deg(v)$. Formally the edge set of B' is

$$E' = \left(\biguplus_{1 \leqslant i \leqslant t_1} E \right) \cup \bigcup_{v \in V} E_v \cup \bigcup_{\substack{v \in V \\ 1 \leqslant j \leqslant t_1 \deg(v)}} E_{v,j},$$

where $\biguplus$ denotes a disjoint union as a multiset of t_1 copies of E, $E_v = \{(v,w) | w \in W_v\}$ and $E_{v,j} = \{(w_{v,j}, u) | u \in U_{v,j}\}$ for each v and j.

Let $C = \rho_1^{2t_1 t_2 m} \alpha^{t_1 m}$ and $N = \begin{bmatrix} 1 & 1 \\ 1 & \alpha^{2t_1} \end{bmatrix}$.

For each $\sigma : V \cup W \to \{0,1\}$, let $w(\sigma)$ be the contribution to $Z_{\alpha,\alpha,\lambda}(B'; U)$ of configurations that are consistent with σ. First consider configurations σ such that $\sigma(w) = 1$ for all $w \in W$. Denote by Σ the set of all such configurations on $V \cup W$. Then for $\sigma \in \Sigma$,

$$\begin{aligned} w(\sigma) &= \rho_1^{t_2 |W|} \prod_{(u,v) \in E} (M_{1,\sigma(u)} M_{\sigma(u),\sigma(v)} M_{\sigma(v),1})^{t_1} \\ &= C \prod_{(u,v) \in E} N_{\sigma(u),\sigma(v)}. \end{aligned}$$

Let $\Sigma^{ind} \subset \Sigma$ be the subset of configurations which induce an independent set on the vertices V and Z^{ind} be its contribution to $Z_{\alpha,\alpha,\lambda}(B'; U)$. Let $\Sigma^{bad} = \Sigma \backslash \Sigma^{ind}$ and Z^{bad} be its contribution. If $\sigma \in \Sigma^{ind}$ then $w(\sigma) = C$. Otherwise, $w(\sigma) \leqslant \alpha^{2t_1} C$. It implies

$$Z^{ind} = I_B \cdot C \qquad \text{and} \qquad Z^{bad} \leqslant 2^n \alpha^{2t_1} C \leqslant \frac{\varepsilon}{6} \cdot C, \tag{8.14}$$

since t_1 satisfies (8.12).

Next consider configurations σ on $V \cup W$ such that $\sigma(w) = 0$ for at least one $w \in W$. Denote this set by Σ' and its contribution by Z^{small}. Then for $\sigma \in \Sigma'$,

$$w(\sigma) \leqslant \left(\rho_0 \rho_1^{|W|-1}\right)^{t_2} \leqslant \left(\frac{\rho_0}{\rho_1}\right)^{t_2} \rho_1^{t_2|W|} = \left(\frac{\rho_0}{\rho_1}\right)^{t_2} \frac{C}{\alpha^{t_1 m}}.$$

It implies

$$Z^{small} \leqslant 2^{2t_1 m+n} \left(\frac{\rho_0}{\rho_1}\right)^{t_2} \frac{C}{\alpha^{t_1 m}} \leqslant \frac{\varepsilon}{6} \cdot C, \tag{8.15}$$

since $|\Sigma'| \leqslant 2^{2t_1 m+n}$ and t_2 satisfies (8.13).

By (8.14) and (8.15) we have

$$\begin{aligned} Z_{\alpha,\alpha,\lambda}(B';U) &= Z^{ind} + Z^{bad} + Z^{small} \\ &\leqslant I_B \cdot C + \frac{\varepsilon}{6} \cdot C + \frac{\varepsilon}{6} \cdot C \\ &\leqslant \exp\left(\frac{\varepsilon}{3}\right) I_B \cdot C, \end{aligned}$$

and clearly $Z_{\alpha,\alpha,\lambda}(B';U) \geqslant I_B \cdot C$. It is also clear that C can be approximated accurate enough given FPRASes for λ and α. This finishes our proof. □

Simulating the Anti-Ferromagnetic Ising Model

In this section we show Lemma 8.12.

Lemma 8.12. *Suppose β, γ and λ are efficiently approximable reals satisfying $\beta, \gamma \geqslant 0$, $\lambda > 0$ and $\beta\gamma \neq 1$. Suppose that Δ is either an integer that is at least 3 or $\Delta = \infty$ (indicating that we do not have a degree bound). If $(\beta, \gamma, \lambda, \Delta)$ supports balanced nearly-independent phase-correlated spins and symmetry breaking, then there exist efficiently approximable $0 < \alpha < 1$ and $\lambda' > 0$ such that $\lambda' \neq 1$ and*

$$\#\text{Bi-M-Nonuniform-2Spin}(\alpha, \alpha, \lambda') \leqslant_{AP} \#\Delta\text{-Bi-2Spin}(\beta, \gamma, \lambda).$$

Proof. We prove the anti-ferromagnetic case first, that is, $\beta\gamma < 1$. α and λ' are chosen as follows. Let $M := \begin{bmatrix} \beta & 1 \\ 1 & \gamma \end{bmatrix}$ and $M^+ := \begin{bmatrix} q^+ & 1-q^+ \\ q^- & 1-q^- \end{bmatrix}$. Let $N = M^+ M (M^+)^\intercal = \begin{bmatrix} N_{++} & N_{-+} \\ N_{+-} & N_{--} \end{bmatrix}$. Then $\det(N) = (\beta\gamma - 1)(q^+ - q^-)^2 < 0$. Therefore $N_{--}N_{++} < N_{-+}N_{+-}$ and let $\alpha = \frac{N_{--}N_{++}}{N_{-+}N_{+-}} < 1$. Moreover, suppose H is the symmetry breaking gadget with distinguished vertex v_H. Let $\rho = \begin{bmatrix} \rho_0 \\ \rho_1 \end{bmatrix}$ where ρ_i denote $\Pr_{H;\beta,\gamma,\lambda}(\sigma_{v_H} = i)$ for spin $i \in \{0,1\}$ and $\rho_0 + \rho_1 = 1$. Let $\rho' = \begin{bmatrix} \rho'_0 \\ \rho'_1 \end{bmatrix} = M^+ \begin{bmatrix} \rho_0 \\ \rho_1/\lambda \end{bmatrix}$, and $\lambda' = \frac{\rho'_0}{\rho'_1}$. It is easy to verify that $\lambda' \neq 1$ as $\rho_0 \neq \lambda/(1+\lambda)$ by the symmetry breaking assumption.

Given $0 < \varepsilon < 1$ and a bipartite multigraph $B = (V, E)$ with a subset $U \subseteq V$ where $|V| = n$, $|E| = m$, and $|U| = n'$, our reduction first constructs a bipartite graph B' with degree at most Δ. The construction of B' involves a gadget G. Since $(\beta, \gamma, \lambda, \Delta)$ supports nearly-independent phase-correlated spins, by Lemma 8.3 $(\beta, \gamma, \lambda, \Delta)$ also supports balanced nearly-independent phase-correlated spins. Therefore we draw $G \sim \mathcal{G}(t, n(t, \varepsilon'), \Delta)$ such that (8.3) and (8.2) hold with probability at least 3/4, where $t = m + 1$ and $\varepsilon' = \frac{\varepsilon}{8n}$. Assume G satisfies them and otherwise the reduction fails. We will construct B' such that

$$\exp\left(-\frac{\varepsilon}{2}\right) Z_{\alpha,\alpha,\lambda'}(B; U) \leqslant \frac{Z_{B'}}{(N_{+-}N_{-+})^m \left(\rho'_1 Z_H\right)^{n'} \left(\frac{Z_G}{2}\right)^n} \leqslant \exp\left(\frac{\varepsilon}{2}\right) Z_{\alpha,\alpha,\lambda'}(B; U),$$

where we use the abbreviated expressions $Z_{B'} = Z_{\beta,\gamma,\lambda}(B')$, $Z_H = Z_{\beta,\gamma,\lambda}(H)$, and $Z_G = Z_{\beta,\gamma,\lambda}(G)$. The lemma follows by one oracle call for $Z_{B'}$ with accuracy $\frac{\varepsilon}{6}$, one oracle call for Z_G with accuracy $\frac{\varepsilon}{6n}$, and an approximation of other terms in the denominator with accuracy $\frac{\varepsilon}{6}$ using FPRASes for q^-, q^+, β, γ and λ.

The graph B' is constructed as follows. For each vertex $v \in V$ we introduce a copy of G, denoted by G_v with vertex set $V(G_v)$. Moreover, for each vertex $u \in U$ we introduce a copy of H, denoted by H_u. Whenever a terminal vertex is used in the construction once, we say it is occupied. For each $(u, v) \in E$, we connect one currently unoccupied positive (and respectively negative) terminal of G_u to one currently unoccupied positive (and respectively negative) terminal of G_v. Denote by E' all these edges between terminals. For each $u \in U$, we identify an unoccupied positive terminal of G_u with the distinguished vertex v_{H_u} of H_u. We denote this terminal by t_u. The resulting graph is B'. It is clear that B' is bipartite and has bounded degree Δ.

Let $\tilde{\sigma}\colon V \to \{-,+\}$ be a configuration of the phases of the G_v's. Let $Z_{B'}(\tilde{\sigma})$ be the contribution to $Z_{B'}$ from the configurations σ that are consistent with $\tilde{\sigma}$ in the sense that, for each $v \in V$, $Y(\sigma_{V(G_v)}) = \tilde{\sigma}(v)$. Then $Z_{B'} = \sum_{\tilde{\sigma}} Z_{B'}(\tilde{\sigma})$. Let T be the set of all terminals $T = \cup_{v \in V} T(G_v)$ and $\tau\colon T \to \{0,1\}$ be a configuration on T. Let $\tau_{T(G_v)}$ be the configuration τ restricted to $T(G_v)$. Recall that for $\pi \in \{-,+\}$, $Z^{\pi}_{G_v}(\tau_{T(G_v)})$ is the contribution to Z_{G_v} from configurations that have phase π and are consistent with $\tau_{T(G_v)}$. Also,

$$\Pr_{G_v;\beta,\gamma,\lambda}(\tau_{T(G_v)} \mid Y(\sigma_{V(G_v)}) = \pi) = \frac{Z^{\pi}_{G_v}(\tau_{T(G_v)})}{Z^{\pi}_{G_v}}.$$

Moreover, for each $u \in U$ and each spin $i \in \{0,1\}$, let $Z_{H_u}(i)$ be the contribution to Z_{H_u} from configurations σ with $\sigma(t_u) = i$. Hence,

$$\rho_i = \Pr_{H_u;\beta,\gamma,\lambda}(\sigma(t_u) = i) = \frac{Z_{H_u}(i)}{Z_{H_u}}.$$

We express $Z_{B'}(\tilde{\sigma})$ as

$$Z_{B'}(\tilde{\sigma}) = \sum_{\tau\colon T \to \{0,1\}} w_{E'}(\tau) \prod_{v \in V} Z^{\tilde{\sigma}(v)}_{G_v}(\tau_{T(G_v)}) \prod_{u \in U} \frac{Z_{H_u}(\tau(t_u))}{\lambda^{1-\tau(t_u)}},$$

where $w_{E'}(\tau)$ is the contribution of edges in E' given configuration τ. Notice that we divide the last factor by λ when $\tau(t_u) = 0$ because we counted the vertex weight twice in that case. Define $\widetilde{Z_{B'}}(\tilde{\sigma})$ to be an approximation version of the partition function where on each $T(G_v)$ the spins are chosen exactly according to $Q^{\tilde{\sigma}(v)}$. That is,

$$\begin{aligned}
\widetilde{Z_{B'}}(\tilde{\sigma}) &= \sum_{\tau\colon T \to \{0,1\}} w_{E'}(\tau) \prod_{v \in V} Z^{\tilde{\sigma}(v)}_{G_v} Q^{\tilde{\sigma}(v)}(\tau_{T(G_v)}) \prod_{u \in U} \frac{Z_{H_u}(\tau(t_u))}{\lambda^{1-\tau(t_u)}} \\
&= \left(\prod_{v \in V} Z^{\tilde{\sigma}(v)}_{G_v}\right) \cdot \left(\sum_{\tau\colon T \to \{0,1\}} w_{E'}(\tau) \prod_{v \in V} Q^{\tilde{\sigma}(v)}(\tau_{T(G_v)}) \prod_{u \in U} \frac{Z_{H_u}(\tau(t_u))}{\lambda^{1-\tau(t_u)}}\right). \qquad (8.16)
\end{aligned}$$

Let $\widetilde{Z_{B'}} = \sum_{\tilde{\sigma}} \widetilde{Z_{B'}}(\tilde{\sigma})$. Then (8.2) implies that $Z_{B'}(\tilde{\sigma})$ and $\widetilde{Z_{B'}}(\tilde{\sigma})$ are close, that is,

$$(1-\varepsilon')^n \leqslant \frac{Z_{B'}(\tilde{\sigma})}{\widetilde{Z_{B'}}(\tilde{\sigma})} \leqslant (1+\varepsilon')^n. \tag{8.17}$$

Moreover, (8.3) implies that

$$\left(\frac{1-\varepsilon'}{2}\right)^n \leqslant \frac{\prod_{v\in V} Z_{G_v}^{\tilde{\sigma}(v)}}{(Z_G)^n} \leqslant \left(\frac{1+\varepsilon'}{2}\right)^n. \tag{8.18}$$

Notice that here Z_{G_v} is the same for any $v \in V$ as the G_v's are identical copies of G.

Then we calculate the following quantity given $\tilde{\sigma}$

$$\sum_{\tau\colon T\to\{0,1\}} w_{E'}(\tau) \prod_{v\in V} Q^{\tilde{\sigma}(v)}(\tau_{T(G_v)}) \prod_{u\in U} \frac{Z_{H_u}(\tau(t_u))}{\lambda^{1-\tau(t_u)}}.$$

As the measure $Q^{\tilde{\sigma}(v)}$ is i.i.d., we may count the weight of each edge in E' independently. Notice that $N_{\pi_1\pi_2}$ is the edge contribution when one end point is chosen with probability q^{π_1} and the other q^{π_2}. For an edge $(u,v) \in V$, if u and v are assigned the same phase $+$, then an edge in E' connecting one $+$ terminal of G_u and one $+$ terminal of G_v gives a weight of N_{++} and an edge connecting two $-$ terminals gives N_{--}. The total weight is $\mu_1 = N_{++}N_{--}$. Similarly if u and v are assigned the same phase $-$, the total weight is μ_1 as well. On the other hand if u and v are assigned distinct phases $+$ and $-$, the total weight is $\mu_2 = N_{+-}N_{-+}$. Recall that $\alpha = \frac{\mu_1}{\mu_2}$. Moreover, for each $u \in U$, if $\tilde{\sigma}(u) = +$, then the contribution of H_u is $\rho_0' Z_{H_u}$ and otherwise $\rho_1' Z_{H_u}$. Notice that here Z_{H_u} is the same for any $u \in U$ as the H_u's are identical copies of H. Recall that $\lambda' = \frac{\rho_0'}{\rho_1'}$.

Plug these calculation into (8.16), we have

$$\begin{aligned}\widetilde{Z_{B'}}(\tilde{\sigma}) &= \left(\prod_{v\in V} Z_{G_v}^{\tilde{\sigma}(v)}\right) \cdot \left(\mu_1^{m_+(\tilde{\sigma})} \mu_2^{m-m_+(\tilde{\sigma})} \left(\rho_0' Z_H\right)^{n_+(\tilde{\sigma})} \left(\rho_1' Z_H\right)^{n'-n_+(\tilde{\sigma})}\right) \\ &= \mu_2^m \left(\rho_1' Z_H\right)^{n'} \left(\prod_{v\in V} Z_{G_v}^{\tilde{\sigma}(v)}\right) \cdot \left(\alpha^{m_+(\tilde{\sigma})} \left(\lambda'\right)^{n_+(\tilde{\sigma})}\right),\end{aligned} \tag{8.19}$$

where $m_+(\tilde{\sigma})$ denotes the number of edges of which the two endpoints are of the same phase given $\tilde{\sigma}$, and $n_+(\tilde{\sigma})$ denotes the number of vertices in U that are assigned $+$ given $\tilde{\sigma}$. Apply

(8.18) to (8.19),

$$(1-\varepsilon')^n \left(\alpha^{m_+(\tilde{\sigma})} \left(\lambda'\right)^{n_+(\tilde{\sigma})}\right) \leqslant \frac{\widetilde{Z_{B'}}(\tilde{\sigma})}{\mu_2^m \left(\rho_1' Z_H\right)^{n'} \left(\frac{Z_G}{2}\right)^n} \leqslant (1+\varepsilon')^n \left(\alpha^{m_+(\tilde{\sigma})} \left(\lambda'\right)^{n_+(\tilde{\sigma})}\right). \quad (8.20)$$

Then we sum over $\tilde{\sigma}$ in (8.20),

$$(1-\varepsilon')^n \left(\sum_{\tilde{\sigma}} \alpha^{m_+(\tilde{\sigma})} \left(\lambda'\right)^{n_+(\tilde{\sigma})}\right) \leqslant \frac{\widetilde{Z_{B'}}}{\mu_2^m \left(\rho_1' Z_H\right)^{n'} \left(\frac{Z_G}{2}\right)^n} \leqslant (1+\varepsilon')^n \left(\sum_{\tilde{\sigma}} \alpha^{m_+(\tilde{\sigma})} \left(\lambda'\right)^{n_+(\tilde{\sigma})}\right)$$
(8.21)

However notice that $Z_{\alpha,\alpha,\lambda'}(B;U) = \sum_{\tilde{\sigma}} \alpha^{m_+(\tilde{\sigma})} (\lambda')^{n_+(\tilde{\sigma})}$ by just mapping $+$ to 0 and $-$ to 1 in each configuration $\tilde{\sigma}$. Combine (8.17) and (8.21),

$$(1-\varepsilon')^{2n} Z_{\alpha,\alpha,\lambda'}(B;U) \leqslant \frac{Z_{B'}}{\mu_2^m \left(\rho_1' Z_H\right)^{n'} \left(\frac{Z_G}{2}\right)^n} \leqslant (1+\varepsilon')^{2n} Z_{\alpha,\alpha,\lambda'}(B;U).$$

Recall that $\varepsilon' = \frac{\varepsilon}{8n}$ and we get the desired bounds.

The other case is ferromagnetic, that is, $\beta\gamma > 1$. Notice that in this case $\det(N) = (\beta\gamma - 1)(q^+ - q^-)^2 > 0$, So we choose $\alpha = \frac{N_{+-}N_{-+}}{N_{++}N_{--}} < 1$ and λ' to be the same as the antiferromagnetic case. The construction of B' is similar to the previous case, with the following change. For each $(u,v) \in E$, we connect one unoccupied positive terminal of G_u to one unoccupied negative terminal of G_v, and vice versa. The rest of the construction is the same. With this change, given a configuration $\tilde{\sigma}\colon V \to \{-,+\}$, if two endpoints are assigned the same spin, the contribution is $N_{+-}N_{-+}$ and otherwise $N_{++}N_{--}$. Therefore the effective edge weight is $\alpha < 1$ when the spins are the same, after normalizing the weight to 1 when the spins are distinct. The rest of the proof is the same. □

Chapter 9

Ferromagnetic 2-Spin Systems

In this chapter we study ferromagnetic 2-spin systems. We will consider a slightly more general problem, where we allow non-uniform fields, specified by a mapping $\pi : V \to \mathbb{R}^+$. When a vertex is assigned "0", we give it a weight $\pi(v)$. Similar as before, we assume the edge interaction function is $\begin{bmatrix} \beta & 1 \\ 1 & \gamma \end{bmatrix}$. For a particular configuration σ, its weight $w(\sigma)$ now becomes,

$$w(\sigma) = \beta^{m_0(\sigma)} \gamma^{m_1(\sigma)} \prod_{v \mid \sigma(v)=1} \pi(v), \tag{9.1}$$

where $m_0(\sigma)$ is the number of $(0,0)$ edges given by the configuration σ and $m_1(\sigma)$ is the number of $(1,1)$ edges.

We also write $\lambda_v := \pi(v)$. If π is a constant function such that $\lambda_v = \lambda > 0$ for all $v \in V$, we also denote it by λ and this is consistent with our previous notations. We say π has a lower bound $\lambda > 0$ (or an upper bound $\lambda > 0$), if π satisfies the guarantee that $\lambda_v \geqslant \lambda$ (or $\lambda_v \leqslant \lambda$).

The partition function $Z_{\beta,\gamma,\pi}(G) = \sum_{\sigma : V \to \{0,1\}} w(\sigma)$ is defined in the same way as before. We consider the following computation problem, where fields are taken from an interval.

Name #2Spin$(\beta, \gamma, [\lambda_1, \lambda_2])$

Instance A graph $G = (V, E)$ and a mapping $\pi : V \to \mathbb{R}^+$, such that $\pi(v) \in [\lambda_1, \lambda_2]$ for any $v \in V$.

Output $Z_{\beta,\gamma,\pi}(G)$.

When the field is uniform, that is, $\lambda_1 = \lambda_2 = \lambda$, we simply write #2Spin$(\beta, \gamma, \lambda)$, which is consistent with our notations.

9.1 The Uniqueness Condition on Regular Trees

Recall that $\mathbb{T}_d$ denote the infinite d-regular tree, also known as the *Bethe lattice* or the *Cayley tree*. If we pick an arbitrary vertex as the root of $\mathbb{T}_d$, then the root has d children and every other vertex has $d-1$ children. The difference between $\mathbb{T}_d$ and an infinite $(d-1)$-ary tree is only the degree of the root. We consider the uniqueness of Gibbs measures on $\mathbb{T}_d$, where the field is uniformly $\lambda > 0$. Due to the symmetric structure of $\mathbb{T}_d$, the standard recursion (7.3) thus becomes $R_v = f_{d-1}(R_{v_i})$ for any child v_i of v, where $f_d(x) = \lambda \left(\frac{\beta x+1}{x+\gamma}\right)^d$ is the symmetrized version of $F_d(\mathbf{x})$.

If $\beta\gamma > 1$, then $f'_d(x) > 0$ for any $x > 0$. There may be 1 or 3 positive fixed points such that $x = f_d(x)$. It is known [Kel85, Geo11] that the Gibbs measure of two-state spin systems in $\mathbb{T}_d$ is unique if and only if there is only one fixed point for $x = f_{d-1}(x)$.

We do some calculation here. Take the derivative of $f_d(x)$:

$$f'_d(x) = \frac{d(\beta\gamma - 1)f_d(x)}{(\beta x + 1)(x + \gamma)}. \tag{9.2}$$

Then take the second derivative:

$$\begin{aligned} f''_d(x) &= f'_d(x) \cdot \frac{1}{f_d(x)} - \frac{\beta}{\beta x + 1} - \frac{1}{x+\gamma} \\ &= \frac{d(\beta\gamma - 1) - \beta\gamma - 1 - 2\beta x}{(\beta x + 1)(x+\gamma)}. \end{aligned}$$

Therefore, at $x^* := \frac{d(\beta\gamma-1)-(\beta\gamma+1)}{2\beta}$, $f''_d(x^*) = 0$. It's easy to see when $d < \frac{\beta\gamma+1}{\beta\gamma-1}$, $f''_d(x) < 0$ for all $x > 0$. So $f_d(x)$ is concave and therefore has only one fixed point.

Since $f_d(x)$ has only one inflection point, there are at most three fixed points. Moreover, the uniqueness condition is equivalent to say that for all fixed points $\widehat{x}_d$ of $f_d(x)$, $f'_d(\widehat{x}_d) < 1$. For a fixed point $\widehat{x}_d$, we plug it in (9.2):

$$f'_d(\widehat{x}_d) = \frac{d(\beta\gamma - 1)\widehat{x}_d}{(\beta\widehat{x}_d + 1)(\widehat{x}_d + \gamma)}.$$

Let $\Delta_c := \frac{\sqrt{\beta\gamma}+1}{\sqrt{\beta\gamma}-1}$. If $d < \Delta_c$, we have that for any x,

$$\begin{aligned}(\beta x+1)(x+\gamma)-d(\beta\gamma-1)x &= \beta x^2+((\beta\gamma+1)-d(\beta\gamma-1))x+\gamma\\ &> \beta x^2+(\beta\gamma+1-(\sqrt{\beta\gamma}+1)^2)x+\gamma\\ &= (\sqrt{\beta}x-\sqrt{\gamma})^2 \geqslant 0.\end{aligned}$$

Hence $(\beta x+1)(x+\gamma) > d(\beta\gamma-1)x$. In particular, $f_d'(\widehat{x}_d) < 1$ for any fixed point $\widehat{x}_d$ and the uniqueness condition holds.

Proposition 9.1. *If $d < \Delta_c = \frac{\sqrt{\beta\gamma}+1}{\sqrt{\beta\gamma}-1}$, then the uniqueness condition holds regardless the field.*

The condition $d < \Delta_c$ matches the exact threshold of fast mixing for Gibbs samplers in the Ising model [MS13].

Next we assume $d \geqslant \Delta_c$. We may also assume that $\gamma \geqslant \beta$. The equation $(\beta x+1)(\gamma+x) = d(\beta\gamma-1)x$ has two solutions, which are

$$x_0 = x^* - \frac{\sqrt{((\beta\gamma+1)-d(\beta\gamma-1))^2-4\beta\gamma}}{2\beta} \quad \text{and} \quad x_1 = x^* + \frac{\sqrt{((\beta\gamma+1)-d(\beta\gamma-1))^2-4\beta\gamma}}{2\beta}.$$

Notice that both of them are positive since $x_0+x_1 = 2x^* > 0$ and $x_0x_1 = \beta/\gamma$.

We show that $f_d(x_0) > x_0$ or $f_d(x_1) < x_1$ is equivalent to the uniqueness condition. First we assume this condition doesn't hold, that is $f_d(x_0) \leqslant x_0$ and $f_d(x_1) \geqslant x_1$. If any of the equation holds, then x_0 or x_1 is a fixed point and the derivative is 1. So we have non-uniqueness. Otherwise, we have $f_d(x_0) < x_0$ and $f_d(x_1) > x_1$. Since $x_0 < x_1$, there is some fixed point $\widetilde{x}$ satisfying $f_d(\widetilde{x}) = \widetilde{x}$ and $x_0 < \widetilde{x} < x_1$. The second inequality implies that $(\beta\widetilde{x}+1)(\widetilde{x}+\gamma) < d(\beta\gamma-1)\widetilde{x}$. Therefore $f_d'(\widetilde{x}) > 1$ and non-uniqueness holds.

To show the other direction, if $f_d(x_0) > x_0$, then

$$f_d'(x_0) = \frac{d(\beta\gamma-1)f(x_0)}{(\beta x_0+1)(x_0+\gamma)} > \frac{d(\beta\gamma-1)x_0}{(\beta x_0+1)(x_0+\gamma)} = 1.$$

Assume for contradiction that $f_d(x)$ has three fixed points, denoted by $\widetilde{x}_0 < \widetilde{x}_1 < \widetilde{x}_2$. Then the middle fixed point $\widetilde{x}_1$ satisfies $f_d'(\widetilde{x}_1) > 1$. Therefore $\widetilde{x}_1 > x_0$ and there are two fixed points larger than x_0. However, for $x_0 < x \leqslant x^*$, $f_d'(x) > 1$ and $f_d(x_0) > x_0$. Hence there is no fixed point in this interval. For $x > x^*$, the function is concave and has exactly one fixed point. So there is only 1 fixed point larger than x_0. Contradiction. The case that $f_d(x_1) < x_1$ is similar.

These two conditions could be rewritten as

$$\lambda > \frac{x_0(x_0+\gamma)^d}{(\beta x_0+1)^d} \tag{9.3}$$

and

$$\lambda < \frac{x_1(x_1+\gamma)^d}{(\beta x_1+1)^d}. \tag{9.4}$$

Notice that the right hand side has nothing to do with λ in both (9.3) and (9.4).

We want to study conditions (9.3) and (9.4) as d changes. In particular, we are interested in the case when $\beta \leqslant 1$. Treat d as a continuous variable. Define

$$g_i(d) = \frac{x_i(x_i+\gamma)^d}{(\beta x_i+1)^d}.$$

where $i = 0, 1$ and x_i is defined above depending on β, γ and d. Take the derivative:

$$\begin{aligned}
\frac{g_i'(d)}{g_i(d)} &= \frac{\partial x_i}{\partial d}\left(\frac{1}{x_i} + \frac{d}{x_i+\gamma} - \frac{d\beta}{\beta x_i+1}\right) + \log(x_i+\gamma) - \log(\beta x_i+1) \\
&= \frac{\partial x_i}{\partial d}\left(\frac{1}{x_i} + \frac{d(1-\beta\gamma)}{(x_i+\gamma)(\beta x_i+1)}\right) + \log\frac{x_i+\gamma}{\beta x_i+1} \\
&= \frac{\partial x_i}{\partial d}\left(\frac{1}{x_i} - \frac{1}{x_i}\right) + \log\frac{x_i+\gamma}{\beta x_i+1} = \log\frac{x_i+\gamma}{\beta x_i+1} > 0.
\end{aligned}$$

Therefore, these two functions are increasing in d.

Recall that $\Delta_c = \frac{\sqrt{\beta\gamma}+1}{\sqrt{\beta\gamma}-1}$. Let $\lambda_c^{int} := g_1(\lceil\Delta_c\rceil) = (\gamma/\beta)^{\frac{\lceil\Delta_c+1\rceil}{2}}$. Thus if $\lambda < \lambda_c^{int}$, (9.4) holds for all integers d. On the other hand,

$$\begin{aligned}
g_0(d) = \frac{x_0(x_0+\gamma)^d}{(\beta x_0+1)^d} &> x_0\beta^{-d} = \frac{\beta}{\gamma x_1}\beta^{-d} > \frac{\beta}{\gamma 2x^*}\beta^{-d} \\
&= \frac{\beta^2}{\gamma(d(\beta\gamma-1)-(\beta\gamma+1))}\cdot\beta^{-d} \\
&\to \infty \text{ as } d \text{ goes to } \infty.
\end{aligned}$$

Hence there is no λ such that (9.3) holds for all integers d.

Proposition 9.2. *Let* (β, γ) *be two parameters such that* $\beta\gamma > 1$ *and* $\beta \leqslant 1 < \gamma$*. The uniqueness condition holds in* $\mathbb{T}_d$ *for all degrees* $d \geqslant 2$ *if and only if* $\lambda < \lambda_c^{int}$*.*

In Section 9.3, we will show that, there exist an FPTAS for the partition function, given the same assumption as in Proposition 9.1. This is Theorem 9.8. Moreover, let $\lambda_c := g_1(\Delta_c) = (\gamma/\beta)^{\frac{\Delta_c+1}{2}}$. In Section 9.4 we show that if $\beta \leqslant 1 < \gamma$ and $\lambda < \lambda_c$, there also exists an FPTAS. This is Theorem 9.9. Note that this is the same condition as in Proposition 9.2, with λ_c^{int} replaced by $\lambda_c \leqslant \lambda_c^{int}$.

9.2 The Potential Method for Ferromagnetic Systems

To show strong spatial mixing in arbitrary trees, we will use the same potential analysis as in Section 7.3. Here we just outline some necessary tweaks we made for ferromagnetic systems

We use the same notations as in Section 7.3. The definition of a good potential function is exactly the same as that of Definition 7.8. Note that now the range of our variables is $[\lambda\gamma^{-d}, \lambda\beta^{d}]$ instead of $[\lambda\beta^{d}, \lambda\gamma^{-d}]$. We say $\Phi(x)$ is a good potential function for d and a field π, if $\Phi(x)$ is a good potential function for d and any λ in the codomain of π, The next lemma is a counterpart of Lemma 7.11. The proof is slightly different.

Lemma 9.3. *Let* $\Phi(x)$ *be a good potential function for the field* λ *at* v*. Then there exists a constant* C *such that* $\delta_v \leqslant C\varepsilon_v$ *for any* $\operatorname{dist}(v, S) \geqslant 2$*.*

Proof. By (7.6) and the Mean Value Theorem, there exists an $\widetilde{R} \in [R_v, R^v]$ such that

$$\varepsilon_v = \Phi(R^v) - \Phi(R_v) = \Phi'(\widetilde{R}) \cdot \delta_v = \varphi(\widetilde{R}) \cdot \delta_v. \tag{9.5}$$

Since $\operatorname{dist}(v, S) \geqslant 2$, we have that $R_v \geqslant \lambda\gamma^{-d}$ and $R^v \leqslant \lambda\beta^{d}$. Hence $\widetilde{R} \in [\lambda\gamma^{-d}, \lambda\beta^{d}]$, and by Condition 1 of Definition 7.8, there exists a constant C_1 such that $\varphi(\widetilde{R}) \geqslant C_1$. Therefore $\delta_v \leqslant 1/C_1\varepsilon_v$. □

Lemma 7.12 holds without modification. The next one is the analogue of Lemma 7.13.

Lemma 9.4. *Let* (β,γ) *be two parameters such that* $\beta\gamma > 1$ *and* $\beta < \gamma$. *Let* v *be a vertex and* v_i *be its children for* $1 \leqslant i \leqslant d$. *Suppose* $\delta_{v_i} \leqslant C$ *for some* $C > 0$ *and all* $1 \leqslant i \leqslant d$. *Then,*

$$\delta_v \leqslant d\lambda_v(\beta\gamma - 1)\gamma^{-1}\beta^d C.$$

Proof. It is easy to see that $\gamma \geqslant 1$. By the same argument as in Lemma 7.12 and (7.3), there exists x_i's such that

$$\delta_v = \left|\nabla F_d(x_1,\ldots,x_d)\cdot(\delta_{v_1},\ldots,\delta_{v_d})\right| \leqslant C\sum_{i=1}^{d}\left|\frac{\partial F_d(\mathbf{x})}{\partial x_i}\right|,$$

where $\mathbf{x}$ is the vector composed by x_i's. Then, we have that

$$\left|\frac{\partial F_d(\mathbf{x})}{\partial x_i}\right| = \frac{d(\beta\gamma-1)F_d(\mathbf{x})}{(x_i+\gamma)(\beta x_i+1)} \leqslant d\lambda_v(\beta\gamma-1)\gamma^{-1}\beta^d,$$

where we use the fact that $F_d(\mathbf{x}) \leqslant \lambda_v\beta^d$ for any $x_i \in [0,\infty)$ and $\beta\gamma > 1$. The lemma follows. □

The algorithmic implication is also the same as Lemma 7.15. Note that we make the dependence of λ explicit here, as we are considering non-uniform fields.

Lemma 9.5. *Let* (β,γ) *be two parameters such that* $\beta\gamma > 1$. *Let* $G = (V, E)$ *be a graph with a maximum degree* Δ *and* n *many vertices and* π *be a field on* G. *Let* $\lambda = \max_{v\in V}\{\pi(v)\}$. *If there exists a good potential function for* π *and all* $d \in [1, \Delta-1]$ *with contraction ratio* $\alpha < 1$, *then* $Z_{\beta,\gamma,\pi}(G)$ *can be approximated deterministically within a relative error* ε *in time* $O\left(n\left(\frac{n\lambda}{\varepsilon}\right)^{\frac{\log(\Delta-1)}{-\log\alpha}}\right)$.

When the degree is unbounded, we also need to use the M-based depth from Definition 7.16. We make some tweaks in the definition of a universal potential function. We assume that $\beta \leqslant 1 < \gamma$ in the definition 9.6, as it will be the range where our potential function works.

Definition 9.6. *Let* $\Phi : \mathbb{R}^+ \to \mathbb{R}^+$ *be a differentiable and monotonically increasing function. Let* $\varphi(x)$ *and* $C_{\varphi,d}(\mathbf{x})$ *defined in the same way as in Definition 7.8. Let* $\beta,\gamma > 0$ *be two parameters such that* $\beta\gamma > 1$ *and* $\beta \leqslant 1 < \gamma$. *Then* $\Phi(x)$ *is a* universal potential function *for* (β,γ,λ) *if it satisfies the following conditions:*

1. *there exists two constants* $C_1, C_2 > 0$ *such that* $C_1 \leqslant \varphi(x) \leqslant C_2$ *for any* $x \in (0,\lambda)$;

2. *there exists a constant* $\alpha < 1$ *such that for all* d, $C_{\varphi,d}(\mathbf{x}) \leqslant \alpha^{\lceil \log_M(d+1) \rceil}$ *for all* $x_i \in (0, \lambda)$*;*

Basically Conditions 1 and 2 in Definition 7.17 are replaced by the stronger Condition 1 in Definition 9.6.

We say $\Phi(x)$ is a universal potential function for a field π, if $\Phi(x)$ is a universal potential function for any λ in the codomain of π, We also call α the contraction ratio and call M the base.

Lemma 9.7. *Let* (β, γ) *be two parameters such that* $\beta\gamma > 1$ *and* $\beta \leqslant 1 < \gamma$*. Let* $G = (V, E)$ *be a graph with* n *many vertices and* π *be a field on* G*. Let* $\lambda = \max_{v \in V}\{\pi(v)\}$*. If there exists a universal potential function* $\Phi(x)$ *for* π *with contraction ratio* $\alpha < 1$ *and base* M*, then* $Z_{\beta,\gamma,\pi}(G)$ *can be approximated deterministically within a relative error* ε *in time* $O\left(n^3 \left(\frac{n\lambda}{\varepsilon}\right)^{\frac{\log M}{-\log \alpha}}\right)$*.*

Proof. By the same proof of Lemma 7.15, we only need to approximate the marginal probability at the root v of a tree T. By Condition 2 of Definition 9.6, $C_{\varphi,d}(x_1, \cdots, x_d) < \alpha^{\lceil \log_M(d+1) \rceil}$. Denote by $B(\ell)$ the set of all vertices whose M-based depths of v is at most ℓ in T. Hence $|B(\ell)| \leqslant M^{\ell}$. Let $S = \{u \mid \mathrm{dist}(u, B(\ell)) > 1\}$, which is essentially the same S as in Lemma 9.5, but under a different metric. We can recursively compute upper and lower bounds R^v and R_v of $R_T^{\sigma_\Lambda}$ such that $R_v \leqslant R_T^{\sigma_\Lambda} \leqslant R^v$, with the base case that for any vertex $u \in S$ trivial bounds $R_u = 0$ and $R^u = \infty$ are used.

We proceed as in the proof of Lemma 9.5. Without loss of generality, we construct a path $u_0 u_1 \cdots u_k$ in T from the root $u_0 = v$ to a u_k with $\ell_M(u_{k-1}) \leqslant \ell$ and $\ell_M(u_k) > \ell$. As in the proof of Lemma 7.12, $\varepsilon_{u_j} \leqslant C^{\varphi}_{d_j}(x_{j,1}, \ldots, x_{j,d_j}) \cdot \varepsilon_{u_{j+1}}$for all $0 \leqslant j \leqslant k-1$, where d_j is the number of children of u_j and $x_{j,i} \in [0, \infty)$, $1 \leqslant i \leqslant d_j$. Hence we have that

$$\begin{aligned}
\varepsilon_v &\leqslant \varepsilon_{u_k} \cdot \prod_{j=0}^{k-1} \alpha^{\lceil \log_M(d_j+1) \rceil} \leqslant \varepsilon_{u_k} \cdot \alpha^{\sum_{j=0}^{k-1} \lceil \log_M(d_j+1) \rceil} \\
&= \varepsilon_{u_k} \cdot \alpha^{\ell_M(u_k)} \leqslant \varepsilon_{u_k} \cdot \alpha^{\ell}.
\end{aligned}$$

Note that $\mathrm{dist}(u_k, B(\ell)) = 1$ and hence $u_k \notin S$. So $\delta_{u_k} < \lambda_{u_k} \leqslant \lambda$. By (9.5), we have that $\varepsilon_{u_k} \leqslant \varphi(\widetilde{R})\delta_{u_k}$, for some $\widetilde{R} \in [\lambda_{u_k}\gamma^{-d_k}, \lambda_{u_k}\beta^{d_k}]$. Hence $\varepsilon_{u_k} < C_2\lambda$ by Condition 1 of Definition 9.6, and $\varepsilon_v < \lambda\alpha^{\ell}C_2$. By (9.5) and Condition 1 of Definition 9.6 again, we have that $\delta_v \leqslant \lambda\alpha^{\ell}C_2/C_1$.

The rest of the proof goes the same as that of Lemma 7.15. The running time has an extra n^2 factor since we need to go down two more levels (in the worst case) outside of $B(\ell)$. □

9.3 Bounded Degree Graphs

Our first result is an FPTAS for any graphs with degree bound $\Delta < \Delta_c + 1 = \frac{2\sqrt{\beta\gamma}}{\sqrt{\beta\gamma}-1}$.

Theorem 9.8. *Let (β, γ) be two parameters such that $\beta\gamma > 1$. Let $G = (V, E)$ be a graph with a maximum degree $\Delta < \Delta_c + 1$ and n many vertices, and let π be a field on G. Let $\lambda = \max_{v \in V}\{\pi(v)\}$. Then $Z_{\beta,\gamma,\pi}(G)$ can be approximated deterministically within a relative error ε in time $O\left(n\left(\frac{n\lambda}{\varepsilon}\right)^{\frac{\log(\Delta-1)}{-\log\alpha}}\right)$, where $\alpha = \frac{\Delta-1}{\Delta_c}$.*

Proof. We choose our potential function to be $\Phi_1(x) = \log x$ such that $\varphi_1(x) := \Phi_1'(x) = \frac{1}{x}$. We verify the conditions of Definition 7.8. Condition 1 is trivial. Then we verify Condition 2, for any integer $1 \leqslant d \leqslant \Delta - 1$,

$$\begin{aligned}
C_{\varphi_1,d}(\mathbf{x}) &= \varphi_1(F_d(\mathbf{x})) \sum_{i=1}^{d} \frac{\partial F_d}{\partial x_i} \cdot \frac{1}{\varphi_1(x)} \\
&= \frac{1}{F_d(\mathbf{x})} \sum_{i=1}^{d} F_d(\mathbf{x}) \cdot \frac{\beta\gamma - 1}{(x_i + \beta)(\gamma x_i + 1)} \cdot x_i \\
&= \sum_{i=1}^{d} \frac{(\beta\gamma - 1)x_i}{(\gamma x_i + 1)(x_i + \beta)} \leqslant \sum_{i=1}^{d} \frac{1}{\Delta_c} = \frac{d}{\Delta_c} \leqslant \frac{\Delta - 1}{\Delta_c} = \alpha,
\end{aligned}$$

where we used the fact that for any $x > 0$,

$$\frac{(\beta\gamma - 1)x}{(\gamma x + 1)(x + \beta)} \leqslant \frac{1}{\Delta_c}.$$

Hence $\Phi_1(x)$ is a good potential function for all degrees $d \in [1, \Delta - 1]$ with contraction ratio α. The theorem follows by Lemma 9.5. □

Note that Theorem 9.8 matches the fast mixing bound of Gibbs samplers for the Ising model in [MS13].

9.4 General Graphs

Now we assume that $\beta \leqslant 1 < \gamma$ and $\beta\gamma > 1$. Recall that $\lambda_c = \left(\frac{\gamma}{\beta}\right)^{\frac{\Delta_c+1}{2}} = \left(\frac{\gamma}{\beta}\right)^{\frac{\sqrt{\beta\gamma}}{\sqrt{\beta\gamma}-1}}$.

Theorem 9.9. *Let* (β, γ) *be two parameters such that* $\beta\gamma > 1$ *and* $\beta \leqslant 1 < \gamma$. *Let* $G = (V, E)$ *be a graph with* n *many vertices, and let* π *be a field on* G. *Let* $\lambda = \max_{v \in V}\{\pi(v)\}$. *If* $\lambda < \lambda_c$, *then* $Z_{\beta,\gamma,\pi}(G)$ *can be approximated deterministically within a relative error* ε *in time* $O\left(n\left(\frac{n\lambda}{\varepsilon}\right)^{\frac{\log M}{-\log \alpha}}\right)$, *where* $M > 1$ *and* $\alpha < 1$ *are two constants depending on* (β, γ, λ).

We will apply Lemma 9.7. We first prove a technical lemma:

Lemma 9.10. *Let* β, γ *be two parameters such that* $\beta\gamma > 1$ *and* $\beta \leqslant 1 < \gamma$. *For any* $0 < x < \lambda_c$, *it holds that*

$$(\beta\gamma - 1)x \log \frac{\lambda_c}{x} \leqslant (\beta x + 1)(x + \gamma) \log \frac{x + \gamma}{\beta x + 1}. \tag{9.6}$$

Proof. Let $g(x) := (\beta\gamma - 1)x \log \frac{\lambda_c}{x} - (\beta x + 1)(x + \gamma) \log \frac{x+\gamma}{\beta x+1}$. Hence it is equivalent to show that $g(x) \leqslant 0$ for all $0 < x < \lambda_c$. Take the derivative of $g(x)$ and we have that

$$\begin{aligned} g'(x) &= (\beta\gamma - 1)(\log \frac{\lambda_c}{x} - 1) - (2\beta x + \beta\gamma + 1) \log \frac{x + \gamma}{\beta x + 1} - (\beta x + 1)(x + \gamma)\left(\frac{1}{x + \gamma} - \frac{\beta}{\beta x + 1}\right) \\ &= (\beta\gamma - 1) \log \frac{\lambda_c}{x} - (2\beta x + \beta\gamma + 1) \log \frac{x + \gamma}{\beta x + 1}. \end{aligned}$$

By direct calculation, $g\left(\sqrt{\frac{\gamma}{\beta}}\right) = 0$ and $g'\left(\sqrt{\frac{\gamma}{\beta}}\right) = 0$. Then we prove (9.6) for the case of $0 < x < \sqrt{\frac{\gamma}{\beta}}$ and $\sqrt{\frac{\gamma}{\beta}} < x < \lambda_c$ separately.

If $0 < x < \sqrt{\frac{\gamma}{\beta}}$, it is sufficient to verify that $g'(x) > 0$. We only need to show that $g'(x)$ is decreasing since $g'\left(\sqrt{\frac{\gamma}{\beta}}\right) = 0$. It is easily verified by taking the derivative again:

$$\begin{aligned} g''(x) &= -\frac{\beta\gamma - 1}{x} - 2\beta \log \frac{x + \gamma}{\beta x + 1} - (2\beta x + \beta\gamma + 1)\left(\frac{1}{x + \gamma} - \frac{\beta}{\beta x + 1}\right) \\ &= -2\beta \log \frac{x + \gamma}{\beta x + 1} - (\beta\gamma - 1)\left(\frac{1}{x} - \frac{2\beta x + \beta\gamma + 1}{(x + \gamma)(\beta x + 1)}\right) \\ &= -2\beta \log \frac{x + \gamma}{\beta x + 1} - (\beta\gamma - 1)\frac{r - \beta x^2}{x(x + \gamma)(\beta x + 1)} < 0, \end{aligned}$$

where the last inequality uses the fact that $\frac{x+\gamma}{\beta x+1} > 1$ (since $\beta \leqslant 1 < \gamma$) and $x < \sqrt{\frac{\gamma}{\beta}}$.

If $\sqrt{\frac{\gamma}{\beta}} < x < \lambda_c$, then we show (9.6) directly. First notice that as $x \neq \sqrt{\frac{\gamma}{\beta}}$,

$$\frac{x}{(\beta x+1)(x+\gamma)} = \frac{1}{\beta x + \frac{\gamma}{x} + \beta\gamma + 1} < (\sqrt{\beta\gamma}+1)^{-2},$$

Given this, in order to get (9.6), it is sufficient to show that $h(x) < 0$ where

$$h(x) := \frac{\sqrt{\beta\gamma}-1}{\sqrt{\beta\gamma}+1} \log\frac{\lambda_c}{x} - \log\frac{x+\gamma}{\beta x+1}.$$

In fact, $h(x)$ is a decreasing function as

$$\begin{aligned} h'(x) &= -\frac{\sqrt{\beta\gamma}-1}{x(\sqrt{\beta\gamma}+1)} - \frac{1}{x+\gamma} + \frac{\beta}{\beta x+1} \\ &= -\frac{(\sqrt{\beta\gamma}-1)\left(\sqrt{\beta}x-\sqrt{\gamma}\right)^2}{x(\sqrt{\beta\gamma}+1)(x+\gamma)(\beta x+1)} \leqslant 0. \end{aligned}$$

Notice that $h\left(\sqrt{\frac{\gamma}{\beta}}\right) = 0$. It implies that $h(x) < 0$ for all $x > \sqrt{\frac{\gamma}{\beta}}$. This completes the proof. □

We then want to show that if $\lambda < \lambda_c$, $g_\lambda(x) \leqslant \alpha$ for some $\alpha < 1$, where

$$g_\lambda(x) := \frac{(\beta\gamma-1)x\log\frac{\lambda}{x}}{(\beta x+1)(x+\gamma)\log\frac{x+\gamma}{\beta x+1}}.$$

By Lemma 9.10, $g_{\lambda_c}(x) \leqslant 1$. Note that $\lim_{x\to 0} g_\lambda(x) = 0$. Hence there exists $0 < \varepsilon < \lambda$ and $0 < \delta < 1$ such that if $0 < x < \varepsilon$, $g_\lambda(x) < \delta$. Moreover, if $\varepsilon \leqslant x < \lambda$, then $\frac{g_\lambda(x)}{g_{\lambda_c}(x)} = \frac{\log\lambda - \log x}{\log\lambda_c - \log x} \leqslant \frac{\log\lambda-\log\varepsilon}{\log\lambda_c-\log\varepsilon}$. Let

$$\alpha_\lambda := \max\left\{\delta, \frac{\log\lambda - \log\varepsilon}{\log\lambda_c - \log\varepsilon}\right\} < 1.$$

Then we have the following lemma.

Lemma 9.11. *Let β,γ be two parameters such that $\beta\gamma > 1$ and $\beta \leqslant 1 < \gamma$. If $\lambda < \lambda_c$, then $g_\lambda(x) \leqslant \alpha_\lambda$ for any $0 < x < \lambda$, where $\alpha_\lambda < 1$ is defined above.*

Let $t := \frac{\alpha_\lambda \gamma}{\beta\gamma - 1} \log \frac{\lambda+\gamma}{\beta\lambda+1}$ so that for any $0 < x < \lambda$,

$$t \leqslant \frac{\alpha_\lambda(\beta x + 1)(x+\gamma)}{\beta\gamma - 1} \log \frac{x+\gamma}{\beta x + 1}.$$

Note that $x \log \frac{\lambda}{x} \leqslant \frac{\lambda}{e}$ for any $0 < x < \lambda$. If $t \geqslant \frac{\lambda}{e}$, then $\frac{1}{t} \cdot x \log \frac{\lambda}{x} \leqslant 1$ for any $0 < x < \lambda$. In this case, we let

$$\varphi_2(x) := \frac{1}{t}. \tag{9.7}$$

Otherwise $t < \frac{\lambda}{e}$, and there are two roots to $x \log \frac{\lambda}{x} = t$ in $(0, \lambda)$. Denote them by x_0 and x_1. We define

$$\varphi_2(x) := \begin{cases} \frac{1}{t} & 0 \leqslant x < x_0; \\ \frac{1}{x \log \frac{\lambda}{x}} & x_0 \leqslant x < x_1; \\ \frac{1}{t} & x_1 \leqslant x < \lambda. \end{cases} \tag{9.8}$$

By our choice of $\varphi_2(x)$, it always holds that for any $0 < x < \lambda$,

$$\varphi_2(x) x \log \frac{\lambda}{x} \leqslant 1, \tag{9.9}$$

and by Lemma 9.11,

$$\frac{(\beta\gamma - 1)}{(\beta x + 1)(x + \gamma)} \cdot \frac{1}{\varphi_2(x)} \leqslant \alpha_\lambda \log \frac{x+\gamma}{\beta x + 1}. \tag{9.10}$$

Now, we are ready to prove Theorem 9.9.

Proof of Theorem 9.9. We claim that Φ_2 is a universal potential function for any field π with an upper bound λ, with contraction ratio α_λ and base M, which will be determined shortly. We verify the two conditions in Definition 9.6

For Condition 1, it is easy to see that in case (9.7), $\varphi_2(x) = \frac{1}{t}$ for any $x \in (0, \lambda)$, and in case (9.8), $\frac{e}{\lambda} \leqslant \varphi_2(x) \leqslant \frac{1}{t}$ for any $x \in (0, \lambda)$.

For Condition 2, we have that

$$
\begin{aligned}
C_{\varphi_2,d}(\mathbf{x}) &= \varphi_2(F_d(\mathbf{x}))\sum_{i=1}^{d}\frac{\partial F_d}{\partial x_i}\cdot\frac{1}{\varphi_2(x_i)} \\
&= \varphi_2(F_d(\mathbf{x}))F_d(\mathbf{x})\sum_{i=1}^{d}\frac{\beta\gamma-1}{(\beta x_i+1)(x_i+\gamma)}\cdot\frac{1}{\varphi_2(x_i)} \\
&\leqslant \varphi_2(F_d(\mathbf{x}))F_d(\mathbf{x})\sum_{i=1}^{d}\alpha_\lambda\log\frac{x_i+\gamma}{\beta x_i+1} && \text{(by (9.10))} \\
&= \alpha_\lambda\varphi_2(F_d(\mathbf{x}))F_d(\mathbf{x})\log\frac{\lambda}{F_d(\mathbf{x})} \\
&\leqslant \alpha_\lambda. && \text{(by (9.9))}
\end{aligned}
$$

Moreover, $F_d(\mathbf{x}) < \lambda\left(\frac{\beta\lambda+1}{\lambda+\gamma}\right)^d$ for any $x_i \in (0,\lambda)$, and $\frac{\beta\lambda+1}{\lambda+\gamma} < 1$. Then there exists $d_0 \geqslant 1$ such that $\left(\frac{\beta\lambda+1}{\lambda+\gamma}\right)^{d_0} < e^{-1}$. Hence, for any $d > d_0$,

$$
\begin{aligned}
C_{\varphi_2,d}(\mathbf{x}) &\leqslant \frac{\alpha_\lambda}{t}F_d(\mathbf{x})\log\frac{\lambda}{F_d(\mathbf{x})} \\
&\leqslant \frac{\alpha_\lambda\lambda}{t}\left(\frac{\beta\lambda+1}{\lambda+\gamma}\right)^d d\log\frac{\beta\lambda+1}{\lambda+\gamma}.
\end{aligned}
$$

Therefore, there exists an integer $M \geqslant d_0$ such that for any $1 \leqslant d < M$, $C_{\varphi_2,d}(\mathbf{x}) \leqslant \alpha_\lambda \leqslant \alpha_\lambda^{\lceil \log_M(d+1)\rceil}$ and for any $d \geqslant M$, $C_{\varphi_2,d}(\mathbf{x}) \leqslant \frac{\alpha_\lambda\lambda}{t}\left(\frac{\beta\lambda+1}{\lambda+\gamma}\right)^d d\log\left(\frac{\beta\lambda+1}{\lambda+\gamma}\right) \leqslant \alpha_\lambda^{\lceil \log_M(d+1)\rceil}$. Condition 2 holds. □

9.5 Correlation Decay Beyond λ_c

Let β,γ be two parameters such that $\beta \leqslant 1 < \gamma$ and $\beta\gamma > 1$. In this section we give an example to show that if Δ_c is not an integer, then correlation decay still holds for a small interval beyond λ_c. To simplify the presentation, we assume that π is a uniform field such that $\pi(v) = \lambda$.

Let $\beta = 0.6$ and $\gamma = 2$. Then $\Delta_c = \frac{\sqrt{\beta\gamma}+1}{\sqrt{\beta\gamma}-1} \approx 21.95$ and $\lambda_c = (\gamma/\beta)^{\frac{\Delta_c+1}{2}} < 1002761$. Let $\lambda = 1002762 > \lambda_c$. We will show that #2Spin(β,γ,λ) is still approximable.

Define a constant t as

$$t := \frac{\sqrt{\beta\gamma}+1}{\sqrt{\beta\gamma}-1} \cdot \frac{\log \sqrt{\gamma/\beta}}{\sqrt{\gamma/\beta}+1} - \log\left(1+\sqrt{\beta/\gamma}\right) \approx 4.24032. \tag{9.11}$$

We consider the potential function $\Phi_3(x)$ so that $\varphi_3(x) := \frac{1}{x(\log(1+1/x)+t)}$. With this choice,

$$\begin{aligned} C_{\varphi_3,d}(\mathbf{x}) &= \varphi_3(F_d(\mathbf{x})) \sum_{i=1}^{d} \frac{\partial F_d}{\partial x_i} \cdot \frac{1}{\varphi_3(x)} \\ &= \frac{\beta\gamma - 1}{\log\left(1 + 1/F_d(\mathbf{x})\right) + t} \sum_{i=1}^{d} \frac{x_i\left(\log(1+1/x_i)+t\right)}{(\beta x_i+1)(x_i+\gamma)}. \end{aligned}$$

We do a change of variables. Let $r_i = \frac{\beta x_i + 1}{x_i+\gamma}$. Then $x_i = \frac{\gamma r_i - 1}{\beta - r_i}$, $\beta x_i + 1 = \frac{r_i(\beta\gamma-1)}{\beta - r_i}$, and $x_i + \gamma = \frac{\beta\gamma-1}{\beta-r_i}$. Hence,

$$\begin{aligned} \sum_{i=1}^{d} \frac{x_i(\log(1+1/x_i)+t)}{(\beta x_i+1)(x_i+\gamma)} &= \sum_{i=1}^{d} \frac{(\gamma r_i - 1)(\beta - r_i)}{r_i(\beta\gamma-1)^2} \cdot \left(\log\left(1+\frac{\beta - r_i}{\gamma r_i - 1}\right)+t\right) \\ &= \frac{1}{(\beta\gamma-1)^2} \sum_{i=1}^{d} \left(1+\beta\gamma - \frac{\beta}{r_i} - \gamma r_i\right)\left(\log\left(1+\frac{\beta-r_i}{\gamma r_i - 1}\right)+t\right). \end{aligned}$$

Furthermore, let $s_i = \log r_i$. As $r_i \in \left(\frac{1}{\gamma}, \beta\right)$, $s_i \in (-\log\gamma, \log\beta)$. Let

$$\rho(x) := \left(1+\beta\gamma - \beta e^{-x} - \gamma e^{x}\right)\left(\log\left(1+\frac{\beta - e^x}{\gamma e^x - 1}\right)+t\right).$$

Then $\rho(x)$ is concave for any $x \in (-\log\gamma, \log\beta)$. It can be easily verified, as the second derivative is

$$\begin{aligned} \rho''(x) = &\frac{(\beta+1)(\beta\gamma-1)}{\beta - 1 + e^x(\gamma-1)} + \frac{\gamma(\beta\gamma-1)}{\gamma-1} - \frac{\beta(\beta\gamma-1)}{\beta - e^x} \\ &- \frac{(\beta-1)(\beta\gamma-1)^2}{(\gamma-1)(\beta-1+e^x(\gamma-1))^2} - \beta t e^{-x} - \gamma t e^{x} - e^{-x}\left(\beta + e^{2x}\gamma\right) \operatorname{Log}\left(1+\frac{\gamma e^x - 1}{\beta - e^x}\right). \\ \leqslant &\ \gamma(\beta+1) + \frac{\gamma(\beta\gamma-1)}{\gamma-1} - \beta\gamma - \frac{\beta-1}{\gamma-1} - 2t < -5.68 < 0, \end{aligned} \tag{9.12}$$

where in the last line we used (9.11) and the fact that $1/\gamma \leqslant e^x \leqslant \beta$. Hence, by concavity, we have that for any $x_i \in (0, \lambda)$,

$$C_{\varphi_3,d}(\mathbf{x}) = \frac{\beta\gamma - 1}{\log\left(1 + 1/F_d(\mathbf{x})\right) + t} \sum_{i=1}^{d} \frac{x_i \left(\log(1 + 1/x_i) + t\right)}{(\beta x_i + 1)(x_i + \gamma)},$$
$$\leqslant \frac{\beta\gamma - 1}{\log\left(1 + 1/f_d(\widetilde{x})\right) + t} \cdot \frac{d\widetilde{x}\left(\log(1 + \widetilde{x}^{-1}) + t\right)}{(\beta\widetilde{x} + 1)(\widetilde{x} + \gamma)} = c_{\varphi_3,d}(\widetilde{x}), \tag{9.13}$$

where $\widetilde{x} > 0$ is the unique solution such that $f_d(\widetilde{x}) = F_d(\mathbf{x})$.

Next we show that there exists an $\alpha < 1$ such that for any integer d and $x > 0$, $c_{\varphi_3,d}(x) < \alpha$. In fact, by (9.11), our choice of t, it is not hard to show that the maximum of $c_{\varphi_3,d}(x)$ is achieved at $x = \sqrt{\gamma/\beta}$ and $d = \Delta_c$, which is 1 if $\lambda = \lambda_c$ and is larger than 1 if $\lambda > \lambda_c$. However, since the degree d has to be an integer, we can verify that for any integer $1 \leqslant d \leqslant 100$, the maximum of $c_{\varphi_3,d}(x)$ is $c_{\varphi_3,22}(x_{22}) = 0.999983$ where $x_{22} \approx 1.83066$. If $d > 100$, then

$$c_{\varphi_3,d}(x) = \frac{d(\beta\gamma - 1)}{\log\left(1 + 1/f_d(x)\right) + t} \cdot \frac{x\left(\log(1 + x^{-1}) + t\right)}{(\beta x + 1)(x + \gamma)}$$
$$\leqslant C_0 \cdot C_1 < 1,$$

where $C_0 < 1.07191$ is the maximum of $\frac{x\left(\log(1+x^{-1})+t\right)}{(\beta x+1)(x+\gamma)}$ for any $x > 0$, and $C_1 < 0.481875$ is the maximum of $\frac{d(\beta\gamma-1)}{\log\left(1+\lambda^{-1}\beta^{-d}\right)+t}$ for any $d > 100$. Then, due to (9.13), we have that for any $x_i \in (0, \lambda)$, $C_{\varphi_3,d}(\mathbf{x}) < \alpha = 0.999983 < 1$. This is the counterpart of $C_{\varphi_2,d}(\mathbf{x}) < \alpha_\lambda$ in the proof of Theorem 9.9. To make $\varphi_3(x)$ satisfy Condition 1 and Condition 2 in Definition 9.6, it is sufficient to do a simple "chop-off" trick to $\varphi_3(x)$ as in (9.8).

Proposition 9.12. *For* $\beta = 0.6$, $\gamma = 2$, *and* $\lambda = 1002762 > \lambda_c$, #2SPIN$(\beta, \gamma, \lambda)$ *has an FPTAS.*

It is easy to see that the above proof works for any $\beta \leqslant 1 < \gamma$ and $\beta\gamma > 1$, except (9.12), the concavity of $\rho(x)$. Indeed, the concavity does not hold if, say, $\beta = 1$ and $\gamma = 2$. Nevertheless, the key point here is that λ_c is not the tight bound for FPTAS. Short of an conjectured optimal bound, we did not try to optimize the potential function nor the applicable range of the proof above.

9.6 Complementary Hardness Results

In this final section, we discuss limitations of approximation algorithms for ferromagnetic 2-spin models based on correlation decay analysis.

The problem of counting independent sets in bipartite graphs (#BIS) plays an important role in classifying approximate counting complexity. #BIS is not known to have any efficient approximation algorithm, despite many attempts. However there is no known approximation preserving reduction (AP-reduction) to reduce #BIS to #SAT either. It is conjectured to have intermediate approximation complexity, and in particular, to have no FPRAS [DGGJ03].

Goldberg and Jerrum [GJ07] showed that for any $\beta\gamma > 1$, approximating #2SPIN$(\beta, \gamma, (0, \infty))$ can be reduced to approximating #BIS. This is the (approximation) complexity upper bound of all ferromagnetic 2-spin models. In contrast, by Theorem 9.8, #Δ-2SPIN$(\beta, \gamma, (0, \infty))$ has an FPTAS if $\Delta < \Delta_c + 1$ and the field is at most polynomial in size of n, the number of vertices.

We then consider fields with some constant bound. Recall that $\lambda_c^{int} = (\gamma/\beta)^{\frac{\lceil \Delta_c+1 \rceil}{2}}$. Let $\lambda_c^{int\prime} = (\gamma/\beta)^{\frac{\lfloor \Delta_c+1 \rfloor+1}{2}}$. Then $\lambda_c^{int\prime} = \lambda_c^{int}$ unless Δ_c is an integer. By reducing to anti-ferromagnetic 2-spin models in bipartite graphs, one can show the following, which is first observed in [LLZ14a, Theorem 3].

Proposition 9.13. *Let (β, γ, λ) be a set of parameters such that $\beta < \gamma$, $\beta\gamma > 1$, and $\lambda > \lambda_c^{int\prime}$. Then #2SPIN$(\beta, \gamma, (0, \lambda])$ is #BIS-hard.*

Proof. We apply Lemma 7.42 and Theorem 8.8. Let $\alpha = \sqrt{\beta\gamma} > 1$. Let $\Delta = \lfloor \Delta_c + 1 \rfloor + 1 = \left\lfloor \frac{2\alpha}{\alpha-1} \right\rfloor + 1 = \left\lfloor \frac{2}{1-\alpha^{-1}} \right\rfloor + 1$. Then $\lambda > (\gamma/\beta)^{\Delta/2}$. By Lemma 7.42, there exists λ_c such that $\lambda_c > 1$, $\lambda_c < \lambda\,(\gamma/\beta)^{-\Delta/2}$ and $\mathrm{Ctr}(\alpha^{-1}, \alpha^{-1}, \lambda_c, \Delta - 1) > 1$. By Theorem 8.8, #$\Delta$-BI-2SPIN$(\alpha^{-1}, \alpha^{-1}, \lambda_c)$ is #BIS-hard.

Recall that the instance of #Δ-BI-2SPIN$(\alpha^{-1}, \alpha^{-1}, \lambda_c)$ is a bipartite graph. We first do a flip of the truth table on one side, say the left, of the instance, that is, to renaming "0" to "1" and "1" to "0". Effectively, after the flip, the edge interaction becomes $\left[\begin{smallmatrix} \alpha & 1 \\ 1 & \alpha \end{smallmatrix}\right]$ (after normalization). Moreover, the vertices on the left have external fields $\lambda_c^{-1} < \lambda_c$ and those on the right still have λ_c. This is a ferromagnetic Ising model with inconsistent fields on two sides.

The next step is a standard diagonal transformation from the Ising model to a general 2-

spin model. More precisely, we do $T = \begin{bmatrix} 1 & 0 \\ 0 & \sqrt{\gamma/\beta} \end{bmatrix}$. The edge interaction becomes $T \begin{bmatrix} \alpha & 1 \\ 1 & \alpha \end{bmatrix} T = \begin{bmatrix} \sqrt{\beta\gamma} & \sqrt{\gamma/\beta} \\ \sqrt{\gamma/\beta} & \sqrt{\beta\gamma}\cdot\gamma/\beta \end{bmatrix} = \sqrt{\gamma/\beta} \begin{bmatrix} \beta & 1 \\ 1 & \gamma \end{bmatrix}$, whereas the external field becomes $\lambda_v' = \lambda_v (\gamma/\beta)^{d_v/2}$ for a vertex v of degree $d_v \leqslant \Delta$. Since $\lambda_v \leqslant \lambda_c$, we have that $\lambda_v' \in (0, \lambda_c (\gamma/\beta)^{\Delta/2}] \subseteq (0, \lambda]$. Hence we obtain a reduction from #Δ-Bi-2Spin($\alpha^{-1}, \alpha^{-1}, \lambda_c$) to #2Spin($\beta, \gamma, (0, \lambda]$). The proposition follows. □

The hardness bound in Proposition 9.13 matches the failure of uniqueness due to Proposition 9.2, unless Δ_c is an integer. In contrast to Proposition 9.13, Theorem 9.9 implies that if $\beta \leqslant 1 < \gamma$ and $\lambda < \lambda_c = (\gamma/\beta)^{\frac{\Delta_c+1}{2}}$, then #2Spin($\beta, \gamma, (0, \lambda]$) has an FPTAS. Hence Theorem 9.9 is almost optimal, up to an integrality gap.

We note that λ_c is not the tight bound for FPTAS, as observed in Proposition 9.12. Since the degree d has to be an integer, with an appropriate choice of the potential function, there is a small interval beyond λ_c such that strong spatial mixing still holds. Interestingly, it seems that λ_c^{int} is not the right bound either. Let us make a concrete example. Let $\beta = 1$ and $\gamma = 2$. Then $\Delta_c = \frac{\sqrt{\beta\gamma}+1}{\sqrt{\beta\gamma}-1} = \frac{\sqrt{2}+1}{\sqrt{2}-1} \approx 5.8$. Hence $\lambda_c \approx 10.6606$ and $\lambda_c^{int} = (2)^{\frac{6+1}{2}} \approx 11.3137$. However, even if $\lambda < \lambda_c^{int}$, the system may not exhibit strong spatial mixing. To see that, we take any $\lambda \in [10.9759, 10.9965]$ so that $\lambda_c < \lambda < \lambda_c^{int}$. Consider an infinite tree where at even layers, each vertex has 5 children, and at odd layers, each vertex has 7 children. The uniqueness condition fails in this tree. This can be easily verified from the fact that the two layer recursion function $f_5(f_7(x))$ has three fixed points such that $x = f_5(f_7(x))$. This example shows that one cannot expect correlation decay algorithms to work all the way up to λ_c^{int}.

At last, if we consider the uniform field case #2Spin(β, γ, λ), then our tractability results still holds. However, to extend the hardness results as in Proposition 9.13 from an interval of fields to a uniform one, there seems to be some technical difficulty. Suppose we want to construct a combinatorial gadget to effectively realize another field. There is a gap between λ and the next largest possible field to realize. This is why in [LLZ14a], there are some extra conditions transiting from an interval of fields to the uniform case. The observation above about the failure of SSM in irregular trees may suggest a random bipartite construction of uneven degrees. However, to analyze such a gadget is beyond the scope of the current work.

Chapter 10

Complex Weighted Ising Models

In this last chapter, we extend our classification of 2-spin systems to complex weights. As we shall see, it has connections to the classical simulation of quantum computation.

10.1 Approximating Complex Numbers

With complex weights, the first issue is what do we mean by approximating a complex number. We are usually interested in the norm or the argument of a complex number. It makes sense that we approximate the norm of a complex number relatively, whereas we approximate the argument additively. This is natural because multiplying complex numbers multiplies norms and adds arguments, so it preserves the usual property that if you can approximate two numbers, you can approximate the product.

Other notions of approximation have been proposed. Most notably, Ziv [Ziv82] has proposed that the distance between two complex numbers y and y' should be measured as

$$d(y', y) := \frac{|y' - y|}{\max(|y'|, |y|)},$$

where $d(0, 0) := 0$.

We will use the following technical lemma concerning Ziv's distance measure.

Lemma 10.1. *If z and z' are two non-zero complex numbers and if $d(z', z) \leqslant \varepsilon$, then $|z'|/|z| \leqslant 1/(1 - \varepsilon)$ and $|\arg z - \arg z'| \leqslant \sqrt{36\varepsilon/11}$.*

Proof. Suppose $d(z', z) \leqslant \varepsilon$ and $|z'| \geqslant |z|$.

First, by the triangle inequality, $|z| + |z' - z| \geqslant |z'|$ so

$$\frac{|z'|}{|z|} = 1 + \frac{|z'| - |z|}{|z|} \leqslant 1 + \frac{|z' - z|}{|z|} = 1 + \frac{|z' - z|}{|z'|}\frac{|z'|}{|z|} \leqslant 1 + \varepsilon\frac{|z'|}{|z|},$$

as required.

Second, letting $z = r\exp(i\theta)$ and $z' = r'\exp(i\theta')$ we have

$$((r'\cos(\theta') - r\cos(\theta))^2 + ((r'\sin(\theta') - r\sin(\theta))^2 \leqslant \varepsilon^2 r'^2.$$

The left-hand-side is equal to $r^2 + r'^2 - 2rr'\cos(\theta - \theta')$. But we already proved

$$1 \leqslant \frac{r'}{r} \leqslant \frac{1}{1-\varepsilon},$$

so

$$r'^2(1-\varepsilon)^2 + r'^2 - 2r'^2\cos(\theta - \theta') \leqslant \varepsilon^2 r'^2,$$

and

$$\cos(\theta - \theta') \geqslant 1 - \frac{3\varepsilon}{2} + \frac{\varepsilon^2}{2}.$$

But $\cos(x) = 1 - x^2/2! + x^4/4! - x^6/6! + \cdots$, so

$$\frac{(\theta-\theta')^2}{2!} - \frac{(\theta-\theta')^4}{4!} + \frac{(\theta-\theta')^6}{6!} - \cdots \leqslant \frac{3\varepsilon}{2} - \frac{\varepsilon^2}{2}.$$

Provided that ε is sufficiently small (so $\theta - \theta' \leqslant 1$) the left-hand-side is at least $11(\theta-\theta')^2/24$, so $|\theta - \theta'| \leqslant \sqrt{36\varepsilon/11}$. □

10.2 Problem Definitions and Results

The main subject that we study in this chapter is the partition function of the Ising model, that is, $Z_{\beta,\beta,\lambda}(G)$. However, in this chapter we will require use of the Tutte polynomial as well. To uniform the notation, and to differentiate Ising partition functions from Tutte ones, we will use the following definition. Given an edge interaction y and an external field λ, the Ising partition function is defined for a (multi)graph $G = (V, E)$ as

$$Z_{\text{Ising}}(G; y, \lambda) = \sum_{\sigma: V \to \{0,1\}} y^{m(\sigma)} \lambda^{n_1(\sigma)}, \tag{10.1}$$

where $m(\sigma)$ is the number of monochromatic edges under σ (that is, the number of edges (u, v) with $\sigma(u) = \sigma(v)$) and $n_1(\sigma)$ is the number of vertices v with $\sigma(v) = 1$. We write $Z_{\text{Ising}}(G; y)$ to denote $Z_{\text{Ising}}(G; y, 1)$. It is easy to see that the definition (10.1) is consistent with (7.2).

To avoid unnecessary technical issues with complex numbers, we will restrict complex parameters y and λ to the set $\overline{\mathbb{Q}}$ of algebraic numbers. Thus, the real and imaginary parts of y and λ will be algebraic. For fixed y and λ, we study several computational problems. We will define them in terms of constant approximation. It is well known that for partition functions, a constant approximation is as good as an inverse polynomial approximation, by a standard amplification trick. The first of them is approximating the norm of $Z_{\text{Ising}}(G; y, \lambda)$ within a factor $K > 1$.

Name FACTOR-K-NORMISING(y, λ).

Instance A (multi)graph G.

Output A rational number $\widehat{N}$ such that $\widehat{N}/K \leqslant \left|Z_{\text{Ising}}(G; y, \lambda)\right| \leqslant K\widehat{N}$.

We also consider the problem of approximating the argument of the partition function within an additive distance of $\rho \in (0, 2\pi)$. Here we have to treat the zero case exceptionally since the argument is undefined.

Name DISTANCE-ρ-ARGISING(y, λ).

Instance A (multi)graph G.

Output If $Z_{\text{Ising}}(G; y, \lambda) = 0$, then the algorithm should output 0. Otherwise, it should output a rational number $\widehat{A}$ such that

$$\left|\widehat{A} - \arg(Z_{\text{Ising}}(G; y, \lambda))\right| \leqslant \rho.$$

We drop the argument λ when it is equal to 1, so FACTOR-K-NORMISING(y) denotes the problem FACTOR-K-NORMISING($y, 1$) and DISTANCE-ρ-ARGISING(y) is DISTANCE-ρ-ARGISING($y, 1$).

As discussed in Section 10.1, it makes sense to approximate complex numbers under Ziv's measure. We also study the following approximation problem.

Name COMPLEXAPX-ISING(y, λ)

Instance A (multi)graph G and a positive integer R, in unary.

Output If $|Z_{\text{Ising}}(G; y, \lambda)| = 0$ then the algorithm should output 0. Otherwise, it should output a complex number y such that $d(y, Z_{\text{Ising}}(G; y, \lambda)) \leqslant \frac{1}{R}$.

As with the other problems, we use the notation COMPLEXAPX-ISING(y) for COMPLEXAPX-ISING($y, 1$). We have specified the error R as an input of the problem, rather than as a parameter, in order to emphasise the suitability of COMPLEXAPX-ISING(y, λ) as an appropriate notion of approximation for the Ising partition function when y is complex. The number R is expressed in unary, so a polynomial time algorithm for COMPLEXAPX-ISING(y, λ) would give an FPTAS or FPRAS for the norm of the partition function. Again, for partition functions, it is well-known that approximating the norm within a factor that is an inverse polynomial in a unary input R is equivalent in difficultly to approximating the norm with any specific factor $K > 1$. We will return to this point later in Lemma 10.8.

Main results for the Ising model

The following theorem gives our main complexity results about the Ising model. These results are illustrated in Figure 10.1.

Theorem 10.2. *Let* $y = re^{i\theta}$ *be an algebraic complex number with* $\theta \in [0, 2\pi)$*. Suppose* $K > 1$*.*

1. *If* $y = 0$ *or if* $r = 1$ *and* $\theta \in \{0, \frac{\pi}{2}, \pi, \frac{3\pi}{2}\}$*, then* FACTOR-K-NORMISING(y)*,* DISTANCE-($\pi/3$)-ARGISING(y) *and* COMPLEXAPX-ISING(y) *are in* **FP***.*

2. *If* $y > 1$ *is a real number then* FACTOR-K-NORMISING(y) *is in* **RP** *and* DISTANCE-($\pi/3$)-ARGISING(y) *is in* **FP***.*

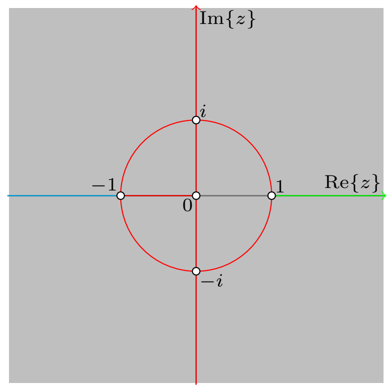


Figure 10.1: An illustration of Theorem 10.2. The five white points correspond to the easy evaluations described in Item 1. The green line segment corresponds to a region where approximation is in **RP**— See Item 2. The blue line segment corresponds to a region where approximation is equivalent to approximately counting perfect matchings. See Item 4. The red points on the axes and on the unit circle correspond to regions where approximation is #**P**-hard. See Items 5, 6, and 7. Elsewhere the points are coloured grey, and approximation is known to be NP-hard (Items 3, 9 and 10) and sometimes to be #**P**-hard (Item 8, not pictured).

3. *If* y *is a real number in* $(0, 1)$ *then* FACTOR-K-NORMISING(y) *is* **NP**-*hard and* DISTANCE-($\pi/3$)-ARGISING(y) *is in* **FP**.

4. *If* $y < -1$ *is a real number then* FACTOR-K-NORMISING(y) *is equivalent in complexity to the problem of approximately counting perfect matchings in a graph and* DISTANCE-($\pi/3$)-ARGISING(y) *is in* **FP**.

5. *If* y *is a real number in* $(-1, 0)$ *then* FACTOR-K-NORMISING(y) *is* #**P**-*hard, and so is* DISTANCE-($\pi/3$)-ARGISING(y).

6. *If* $r = 1$ *and* $\theta \notin \{0, \frac{\pi}{2}, \pi, \frac{3\pi}{2}\}$*, then* FACTOR-K-NORMISING(y), DISTANCE-($\pi/3$)-ARGISING(y) *and* COMPLEXAPX-ISING(y) *are* #**P**-*hard.*

7. *If* $r \notin \{-1, 0, 1\}$*, and* $\theta = \frac{\pi}{2}$ *or* $\theta = \frac{3\pi}{2}$*, then* FACTOR-K-NORMISING(y), DISTANCE-($\pi/3$)-ARGISING(y) *and* COMPLEXAPX-ISING(y) *are* #**P**-*hard.*

8. *If* $r > 0$ *and* $\theta = \frac{a\pi}{2b}$*, where* a *and* b *are two co-prime positive integers and* a *is odd, then* FACTOR-K-NORMISING(y), DISTANCE-($\pi/3$)-ARGISING(y) *and* COMPLEXAPX-ISING(y) *are* #**P**-*hard.*

9. *If* $r < 1$ *and* $y \neq 0$*, then* FACTOR-K-NORMISING(y) *and* COMPLEXAPX-ISING(y) *are* **NP**-*hard.*

10. *If* $r > 1$ *and* $\theta \notin \{0, \pi\}$ *then* FACTOR-K-NORMISING(y) *and* COMPLEXAPX-ISING(y) *are* **NP**-*hard.*

Relaxed versions of the problems

A polynomial-time algorithm for any of the problems that we have defined is required to output 0 if it is given an input G such that $Z_{\text{Ising}}(G; y, \lambda) = 0$. Theorem 10.2 gives hardness results for these problems. The hardness is not due to special difficulties which arise when the value of the partition function is zero. In order to demonstrate this point, (and in order to make certain reductions easier later on), we also consider the following, more relaxed versions of the problems, where the output is unconstrained if the value of the partition function is zero. As before, there parameter K is greater than 1 and the parameter ρ is in $(0, 2\pi)$.

Name FACTOR-K-NONZERO-NORMISING(y, λ)

Instance A (multi)graph G.

Output If $|Z_{\text{Ising}}(G; y, \lambda)| = 0$ then the algorithm may output any rational number. Otherwise, it should output a rational number $\widehat{N}$ such that $\widehat{N}/K \leqslant |Z_{\text{Ising}}(G; y, \lambda)| \leqslant K\widehat{N}$.

Name DISTANCE-ρ-NONZERO-ARGISING(y, λ)

Instance A (multi)graph G.

Output If $Z_{\text{Ising}}(G; y, \lambda) = 0$, then the algorithm may output any rational number. Otherwise, it should output a rational number $\widehat{A}$ such that $\left|\widehat{A} - \arg(Z_{\text{Ising}}(G; y, \lambda))\right| \leqslant \rho$.

Name COMPLEXAPX-NONZERO-ISING(y, λ)

Instance A (multi)graph G and a positive integer R, in unary.

Output If $|Z_{\text{Ising}}(G; y, \lambda)| = 0$ then the algorithm may output any complex number. Otherwise, it should output a complex number z such that $d(z, Z_{\text{Ising}}(G; y, \lambda)) \leqslant \frac{1}{R}$.

As in the un-relaxed versions of the problems, we drop the parameter "λ" from the problem name when it is 1. We give the following generalisation of Theorem 10.2.

Theorem 10.3. *All of the results in Theorem 10.2 extend to the relaxed case. That is, the results still hold even if* FACTOR-K-NORMISING(y), DISTANCE-$(\pi/3)$-ARGISING(y) *and* COMPLEXAPX-ISING(y) *are replaced by* FACTOR-K-NONZERO-NORMISING(y), DISTANCE-$(\pi/3)$-NONZERO-ARG-ISING(y), *and* COMPLEXAPX-NONZERO-ISING(y), *respectively.*

We note that due to Lemma 10.1, the problem of COMPLEXAPX-ISING(y, λ) is always harder than FACTOR-K-NORMISING(y, λ) or DISTANCE-ρ-ARGISING(y, λ), even if they are both relaxed.

Lemma 10.4. *Suppose* $K > 1$ *and* $0 < \rho < 2\pi$*. Then the following polynomial-time Turing reductions exist.*

$$\text{FACTOR-K-NORMISING}(y, \lambda) \leqslant_T \text{COMPLEXAPX-ISING}(y, \lambda),$$

$$\text{FACTOR-K-NONZERO-NORMISING}(y, \lambda) \leqslant_T \text{COMPLEXAPX-NONZERO-ISING}(y, \lambda),$$

$$\text{DISTANCE-}\rho\text{-ARGISING}(y, \lambda) \leqslant_T \text{COMPLEXAPX-ISING}(y, \lambda),$$

$$\text{DISTANCE-}\rho\text{-NONZERO-ARGISING}(y, \lambda) \leqslant_T \text{COMPLEXAPX-NONZERO-ISING}(y, \lambda).$$

Proof. Let R be any (sufficiently large) integer so that $1 - 1/R > 1/K$ and $\sqrt{36/11R} \leqslant \rho$.

Consider a multigraph G where $|Z_{\text{Ising}}(G; y, \lambda)| \neq 0$. Given input G and R, an oracle for COMPLEXAPX-ISING(y, λ)or COMPLEXAPX-NONZERO-ISING(y, λ) returns a complex number z such that $d(z, Z_{\text{Ising}}(G; y, \lambda)) \leqslant \frac{1}{R}$. On the other hand, if $|Z_{\text{Ising}}(G; y, \lambda)| = 0$, then the oracle for COMPLEXAPX-ISING(y, λ) returns the complex number $z = 0$ and the oracle for COMPLEXAPX-NONZERO-ISING(y, λ) returns any complex number z.

For the first two reductions, suppose first that $|Z_{\text{Ising}}(G; y, \lambda)| \neq 0$. Then $d(z, Z_{\text{Ising}}(G; y, \lambda)) \leqslant \frac{1}{R}$ and Lemma 10.1 imply that

$$\frac{|z|}{K} \leqslant \left(1 - \frac{1}{R}\right)|z| \leqslant |Z_{\text{Ising}}(G; y, \lambda)| \leqslant \frac{|z|}{1 - \frac{1}{R}} \leqslant K|z|,$$

so $|z|$ is a suitable output to FACTOR-K-NORMISING(y, λ) or FACTOR-K-NONZERO-NORMISING(y, λ) with input G. On the other hand, if $|Z_{\text{Ising}}(G; y, \lambda)| = 0$, then $|z|$ is still suitable in both cases.

For the last two reductions, suppose first that $|Z_{\text{Ising}}(G; y, \lambda)| \neq 0$. Then $d(z, Z_{\text{Ising}}(G; y, \lambda)) \leqslant$

$\frac{1}{R}$ and Lemma 10.1 imply that

$$\left|\arg z - \arg Z_{\text{Ising}}(G; y, \lambda)\right| \leqslant \sqrt{36\varepsilon/11} \leqslant \rho,$$

so $\arg z$ is a suitable output to either DISTANCE-ρ-ARGISING(y, λ) or DISTANCE-ρ-NONZERO-ARGISING(y, λ) with input G. On the other hand, if $|Z_{\text{Ising}}(G; y, \lambda)| = 0$ and $z = 0$, then 0 is a suitable output in both cases. If $|Z_{\text{Ising}}(G; y, \lambda)| = 0$ and $z \neq 0$, then $\arg z$ is suitable (as an output for DISTANCE-ρ-NONZERO-ARGISING(y, λ)). □

Ising models with fields

Our Theorems 10.2 and 10.3 are about the complexity of evaluating the Ising partition function in the absence of an external field (that is, $\lambda = 1$). This is appropriate for the application to **IQP**, a complexity subclass of **BQP**. However, Ising models with external fields are important for their own sake. Moreover, De las Cuevas et al. [DDVM11, Result 2] showed that with edge interaction i and external field $e^{i\pi/4}$ an additive approximation of the partition function is **BQP**-hard. Motivated by the Ising model itself and such quantum connections, we focus on the problem of (multiplicatively) approximating the norm of the partition function when both the interaction parameter and the external field are roots of unity. We extend our hardness results to show that, for most such parameters, including the one studied by De las Cuevas et al., the approximation problem is #**P**-hard. For the remaining parameters, the partition function can be evaluated exactly in polynomial time, and thus we get a complete dichotomy. This extension relies on some lower bounds from transcendental number theory, which allow us to convert additive distances into multiplicative ones. Details are in Section 10.7.

Theorem 10.5. *Let $K > 1$. Let y and z be two roots of unity. Then the following holds:*

1. *If $y = \pm i$ and $z \in \{1, -1, i, -i\}$, or $y = \pm 1$, $Z_{Ising}(-; y, z)$ can be computed exactly in polynomial time.*
2. *Otherwise* FACTOR-K-NONZERO-NORMISING(y, z) *is #**P**-hard.*

10.3 The Tutte polynomial

The partition function $Z_{\text{Ising}}(G;y)$ is equivalent to a specialisation of the *Tutte polynomial*, which is a graph polynomial with two parameters, x and y, defined as follows,

$$T(G;x,y) = \sum_{A \subseteq E(G)} (x-1)^{\kappa(A)-\kappa(E(G))}(y-1)^{|A|-n+\kappa(A)}, \tag{10.2}$$

where $n = |V(G)|$ and $\kappa(A)$ is the number of connected components in the subgraph $(V(G),A)$. If the quantity $q = (x-1)(y-1)$ is a positive integer, then the Tutte polynomial with parameters x and y is closely-related to the partition function of the Potts model, which includes the Ising model as the special case $q = 2$. In particular, when $q = 2$,

$$T(G;x,y) = (y-1)^{-n}(x-1)^{-\kappa(E(G))} Z_{\text{Ising}}(G;y). \tag{10.3}$$

We will encounter the following two problems, where x, y are two real numbers.

Name SIGN-REALTUTTE(x,y)

Instance A (multi)graph G.

Output Determine whether the sign of the real part of $T(G;x,y)$ is positive, negative, or 0.

Name SIGN-REAL-NONZEROTUTTE(x,y)

Instance A (multi)graph G.

Output A correct statement of the form "$T(G;x,y) \geqslant 0$" or "$T(G;x,y) \leqslant 0$".

We will require the random cluster formulation of the multivariate Tutte polynomial. Given a (multi) graph G with edge weights $\boldsymbol{\gamma} : E(G) \to \overline{\mathbb{Q}}$ and $q \in \overline{\mathbb{Q}}$, this is defined as

$$Z_{\text{Tutte}}(q,\boldsymbol{\gamma}) := \sum_{A \subseteq E} q^{\kappa(A)} \prod_{e \in A} \gamma_e. \tag{10.4}$$

Suppose x and y satisfy $q = (x-1)(y-1)$. For a graph $G = (V,E)$, let $\boldsymbol{\gamma} : E \to \overline{\mathbb{Q}}$ be the constant function which maps every edge to the value $y-1$. Then (see, for example [Sok05, (2.26)])

$$T(G;x,y) = (y-1)^{-n}(x-1)^{-\kappa(E(G))} Z_{\mathrm{Tutte}}(q,\gamma). \tag{10.5}$$

Obviously from (10.3), this implies that if $q = 2$, then $Z_{\mathrm{Ising}}(G;y) = Z_{\mathrm{Tutte}}(q,\boldsymbol{\gamma})$.

To apply a technique from [GJ14] we will require a multivariate version of the problem FACTOR-K-NONZERO-NORMISING(y,λ). We could do this for general q, but we will only use the following version, which is restricted to $q = 2$ and has two complex parameters, γ_1 and γ_2.

Name FACTOR-K-NONZERO-NORM2TUTTE(γ_1,γ_2)

Instance A (multi)graph $G = (V,E)$ and edge weights $\boldsymbol{\gamma} : E \to \{\gamma_1,\gamma_2\}$.

Output If $|Z_{\mathrm{Tutte}}(2,\boldsymbol{\gamma})| = 0$ then the algorithm may output any rational number. Otherwise, it should output a rational number $\widehat{N}$ such that $\widehat{N}/K \leqslant |Z_{\mathrm{Tutte}}(2,\boldsymbol{\gamma})| \leqslant K\widehat{N}$.

Suppose that s and t are two distinguished vertices of G. Let $Z_{st}(G;q,\boldsymbol{\gamma})$ be the contribution to $Z_{\mathrm{Tutte}}(q,\boldsymbol{\gamma})$ from subgraphs where s and t are in the same component, that is,

$$Z_{st}(G;q,\boldsymbol{\gamma}) := \sum_{\substack{A\subseteq E:\\ s \text{ and } t \text{ are}\\ \text{in the same component}}} q^{\kappa(A)} \prod_{e\in A} \gamma_e.$$

Similarly, let $Z_{s|t}$ denote the contribution to $Z_{\mathrm{Tutte}}(q,\boldsymbol{\gamma})$ from configurations A in which s and t are in different components.

10.4 Series and Parallel Compositions

In later reductions, we will want to implement new edge weights. The construction that will be used repeatedly is series and parallel compositions. Our treatment is completely standard and is taken from [GJM15, Section 2.1]. We include it here for completeness.

Fix $W \subseteq \overline{\mathbb{Q}}$ and $q \in \overline{\mathbb{Q}}$. Let $w^* \in \overline{\mathbb{Q}}$ be a weight (which may not be in W) which we want to "implement". Suppose that there is a graph Υ, with distinguished vertices s and t and a weight function $\widehat{\boldsymbol{\gamma}} : E(\Upsilon) \to W$ such that

$$w^* = qZ_{st}(\Upsilon; q, \widehat{\boldsymbol{\gamma}})/Z_{s|t}(\Upsilon; q, \widehat{\boldsymbol{\gamma}}). \tag{10.6}$$

In this case, we say that Υ and $\widehat{\boldsymbol{\gamma}}$ implement w^* (or even that W implements w^*).

The purpose of "implementing" edge weights is this. Let G be a graph with weight function $\boldsymbol{\gamma}$. Let f be some edge of G with weight $\gamma_f = w^*$. Suppose that W implements w^*. Let Υ be a graph with distinguished vertices s and t with a weight function $\widehat{\boldsymbol{\gamma}} : E(\Upsilon) \to W$ satisfying (10.6). Construct the weighted graph G' by replacing edge f with a copy of Υ (identify s with either endpoint of f (it doesn't matter which one) and identify t with the other endpoint of f and remove edge f). Let the weight function $\boldsymbol{\gamma}'$ of G' inherit weights from $\boldsymbol{\gamma}$ and $\widehat{\boldsymbol{\gamma}}$ (so $\gamma'_e = \hat{\gamma}_e$ if $e \in E(\Upsilon)$ and $\gamma'_e = \gamma_e$ otherwise). Then the definition of the multivariate Tutte polynomial gives

$$Z_{\text{Tutte}}(q, \boldsymbol{\gamma}') = \frac{Z_{s|t}(\Upsilon; q, \widehat{\boldsymbol{\gamma}})}{q^2} Z_{\text{Tutte}}(q, \boldsymbol{\gamma}). \tag{10.7}$$

So, as long as $q \neq 0$ and $Z_{s|t}(\Upsilon; q, \widehat{\boldsymbol{\gamma}})$ is easy to evaluate, evaluating the multivariate Tutte polynomial of G' with weight function $\boldsymbol{\gamma}'$ is essentially the same as evaluating the multivariate Tutte polynomial of G with weight function $\boldsymbol{\gamma}$.

Since the norm of the product of two complex numbers is the product of the norms, this reduces computing (or relatively approximating) the norm with weight function $\boldsymbol{\gamma}$ to the problem of computing (or relatively approximating) the norm with weight function $\boldsymbol{\gamma}'$. Also, since the argument of the product of two complex numbers is the sum of the arguments of the numbers, this reduces computing (or additively approximating) the argument with weight function $\boldsymbol{\gamma}$ to the problem of computing (or additively approximating) the argument with weight function γ'.

Two especially useful implementations are series and parallel compositions. Parallel composition is the case in which Υ consists of two parallel edges e_1 and e_2 with endpoints s and t and $\hat{\gamma}_{e_1} = w_1$ and $\hat{\gamma}_{e_2} = w_2$. It is easily checked from (10.6) that $w^* = (1 + w_1)(1 + w_2) - 1$. Also, the extra factor in (10.7) cancels, so in this case $Z_{\text{Tutte}}(q, \boldsymbol{\gamma}') = Z_{\text{Tutte}}(q, \boldsymbol{\gamma})$.

Series composition is the case in which Υ is a length-2 path from s to t consisting of edges

e_1 and e_2 with $\hat{\gamma}_{e_1} = w_1$ and $\hat{\gamma}_{e_2} = w_2$. It is easily checked from (10.6) that $w^* = w_1 w_2/(q + w_1 + w_2)$. Also, the extra factor in (10.7) is $q + w_1 + w_2$, so in this case $Z_{\text{Tutte}}(q, \boldsymbol{\gamma}') = (q + w_1 + w_2) Z_{\text{Tutte}}(q, \boldsymbol{\gamma})$. It is helpful to note that w^* satisfies

$$\left(1 + \frac{q}{w^*}\right) = \left(1 + \frac{q}{w_1}\right)\left(1 + \frac{q}{w_2}\right).$$

We say that there is a "shift" from (q, α) to (q, α') if there is an implementation of α' consisting of some Υ and $\widehat{w} : E(\Upsilon) \to W$ where W is the singleton set $W = \{\alpha\}$. Taking $y = \alpha + 1$ and $y' = \alpha' + 1$ and defining x and x' by $q = (x-1)(y-1) = (x'-1)(y'-1)$, we equivalently refer to this as a shift from (x, y) to (x', y'). It is an easy, but important observation that shifts may be composed to obtain new shifts. So, if we have shifts from (x, y) to (x', y') and from (x', y') to (x'', y''), then we also have a shift from (x, y) to (x'', y'').

The k-thickening of [JVW90] is the parallel composition of k edges of weight α. It implements $\alpha' = (1 + \alpha)^k - 1$ and is a shift from (x, y) to (x', y') where $y' = y^k$ (and x' is given by $(x'-1)(y'-1) = q$). Similarly, the k-stretch is the series composition of k edges of weight α. It implements an α' satisfying

$$1 + \frac{q}{\alpha'} = \left(1 + \frac{q}{\alpha}\right)^k.$$

It is a shift from (x, y) to (x', y') where $x' = x^k$. (In the classical bivariate (x, y) parameterisation, there is effectively one edge weight, so the stretching or thickening is applied uniformly to every edge of the graph.)

Thus, we have the following observation.

Proposition 10.6. *The k-thickening operation gives the following polynomial-time reductions.*

- Factor-K-NormIsing$(y^k) \leqslant$ Factor-K-NormIsing(y),
- Distance-ρ-ArgIsing$(y^k) \leqslant$ Distance-ρ-ArgIsing(y),
- Sign-RealTutte$(1 + (x-1)(y-1)/(y^k - 1), y^k) \leqslant$ Sign-RealTutte(x, y), *where* $y^k \neq 1$, *and*
- ComplexApx-Ising$(y^k) \leqslant$ ComplexApx-Ising(y).

Similarly, k*-stretching gives the following polynomial-time reductions for* $y \neq 1$*.*

- FACTOR-K-NORMISING$(1+2/((1+2/(y-1))^k-1)) \leqslant$ FACTOR-K-NORMISING(y),
- DISTANCE-ρ-ARGISING$(1+2/((1+2/(y-1))^k-1)) \leqslant$ DISTANCE-ρ-ARGISING(y),
- SIGN-REALTUTTE$(x^k, 1+(x-1)(y-1)/(x^k-1)) \leqslant$ SIGN-REALTUTTE(x,y), *where* $x^k \neq 1$, *and*
- COMPLEXAPX-ISING$(1+2/((1+2/(y-1))^k-1)) \leqslant$ COMPLEXAPX-ISING(y).

Similar statements hold for the relaxed versions of the problems.

10.5 Hardness Results for the Ising Model

In this section we prove Theorems 10.2 and 10.3. We will start with real weights and then extend the results to the whole complex plane.

Real Weights

First we gather some known results regarding approximating the partition function $Z_{\text{Ising}}(G; y)$ of the Ising model when y is an algebraic real number.

If $y \in \{-1, 0, 1\}$, then computing $Z_{\text{Ising}}(G; y)$ is trivial from the definition (10.1). A classical result by Jerrum and Sinclair [JS93] settles the complexity of approximating $Z_{\text{Ising}}(G; y)$ when $y > 0$. They show that there is an FPRAS when $y > 1$ and that it is **NP**-hard to approximate the partition function when $0 < y < 1$. The negative case appears to be more complicated. Goldberg and Jerrum [GJ08] showed that if $-1 < y < 0$, it is also **NP**-hard to approximate $Z_{\text{Ising}}(G; y)$, but if $y < -1$, the problem is equivalent to approximating the number of perfect matchings in a graph and it is not known whether there is an FPRAS. Technically, neither Jerrum and Sinclair nor Goldberg and Jerrum worked over the algebraic numbers. In order to avoid issues of real arithmetic, Jerrum and Sinclair used a computational model in which real arithmetic is performed with perfect accuracy, and Goldberg and Jerrum restricted attention to rationals. However, the operations in those papers are easily implemented over the algebraic real numbers. Using our notation, these results are summarised as follows.

Lemma 10.7 ([JS93, GJ08]). *Suppose* $y \in \overline{\mathbb{Q}}$ *and* $K > 1$. *Then* FACTOR-K-NORMISING(y)

- *is in* **FP** *if* $y \in \{-1, 0, 1\}$*;*
- *is in* **RP** *if* $y > 1$*;*
- *is* **NP***-hard if* $0 < y < 1$ *or* $-1 < y < 0$*; and*
- *is equivalent in difficulty to approximately counting perfect matchings if* $y < -1$.

Technically, the results in [JS93, GJ08] were not about the problem FACTOR-K-NORMISING(y) with fixed K. Instead, the accuracy parameter was viewed as part of the input as in the following problem.

Name FPRAS-NORMISING(y, λ)

Instance A (multi)graph G and a positive integer R, in unary.

Output A rational number $\widehat{N}$ such that

$$\left(1 - \tfrac{1}{R}\right) \widehat{N} \leqslant |Z_{\text{Ising}}(G; y, \lambda)| \leqslant \left(1 + \tfrac{1}{R}\right) \widehat{N}.$$

Nevertheless, the hardness results in Lemma 10.7 follow easily from those papers using the following standard powering lemma.

Lemma 10.8. *For any* $K > 1$*, there are polynomial-time Turing reductions between* FACTOR-K-NORMISING(y, λ) *and* FPRAS-NORMISING(y, λ).

Proof. The reduction from FACTOR-K-NORMISING(y, λ) to FPRAS-NORMISING(y, λ) is straightforward. Given an input G to FACTOR-K-NORMISING(y, λ), choose R so that $K \geqslant R/(R-1)$ and run an algorithm for FPRAS-NORMISING(y, λ) with inputs G and R, returning the result.

The other direction is almost as easy. Given an input (G, R) to FPRAS-NORMISING(y, λ), choose an integer k sufficiently large so that $(1 - 1/R)^k \leqslant 1/K$ and $(1 + 1/R)^k \geqslant K$. Then form G_k by taking k disjoint copies of G. Run an algorithm for FACTOR-K-NORMISING(y, λ) with input G_k, obtaining a number $\widehat{N}$ such that $\widehat{N}/K \leqslant |Z_{\text{Ising}}(G_k; y, \lambda)| \leqslant K\widehat{N}$. Then note that

$Z_{\text{Ising}}(G_k; y, \lambda) = Z_{\text{Ising}}(G; y, \lambda)^k$, so

$$\left(1 - \tfrac{1}{R}\right) \widehat{N}^{1/k} \leqslant \widehat{N}^{1/k}/K^{1/k} \leqslant |Z_{\text{Ising}}(G; y, \lambda)| \leqslant K^{1/k}\widehat{N}^{1/k} \leqslant \widehat{N}^{1/k} \left(1 + \tfrac{1}{R}\right),$$

so $\widehat{N}^{1/k}$ is a suitable output. □

Note that the **NP**-hardness result for $0 < y < 1$ in Lemma 10.7 is essentially best possible in the sense that the problem is not much harder than **NP**. As [GJ08] observed, the problem can be solved in randomised polynomial time using an oracle for an **NP** predicate by using the bisection technique of Valiant and Vazirani [VV86]. The situation is different for $y < 0$. Goldberg and Jerrum [GJ14, Theorem 1, Region G] showed that it is #**P**-hard to determine the sign of $Z_{\text{Ising}}(G; y)$ if $-1 < y < 0$. Again, they stated their theorem for the case in which y is rational, but the proof applies equally well when y is an algebraic real number. In terms of our notation, they proved the following lemma.

Lemma 10.9 ([GJ14])**.** *For any algebraic real number* $y \in (-1, 0)$, SIGN-REALTUTTE(x, y) *is #**P**-hard, where* $x = 1 + 2/(y - 1)$.

If y is real then $Z_{\text{Ising}}(G; y)$ is real so, either $Z_{\text{Ising}}(G; y) = 0$, or $\arg(Z_{\text{Ising}}(G; y)) \in \{0, \pi\}$. Hence, approximating the argument within $\pm\pi/3$ enables one to determine the sign of the real part. Using the connection (10.3) between the Tutte polynomial and the partition function of the Ising model and Lemma 10.4, Lemma 10.9 implies the following corollary.

Corollary 10.10. *Suppose* y *is an algebraic real number in the range* $y \in (-1, 0)$*. Then the problem* DISTANCE-$(\pi/3)$-ARGISING(y) *is #**P**-hard and so is* COMPLEXAPX-ISING(y)*.*

In fact, we can extend Goldberg and Jerrum's #**P**-hardness interval-shrinking technique from [GJ14] to also obtain #**P**-hardness for the relaxed version of the problems. We start with a general discussion of interval shrinking. Suppose that we have a linear function $f(\varepsilon) = -\varepsilon A + B$ for positive A and B and that we wish to find a value $\hat{\varepsilon}$ that is very close to the root $\varepsilon^* = B/A$. Suppose that we also have an interval $[\varepsilon', \varepsilon'']$ such that $f(\varepsilon') > 0$ and $f(\varepsilon'') < 0$. Suppose that $\varepsilon'' - \varepsilon' = \ell$ (so the interval has length ℓ). Roughly, Goldberg and Jerrum had to hand an oracle for computing the sign of $f(\varepsilon)$ (using an oracle for SIGN-REALTUTTE(x, y)) and, using this, it is easy to bisect the interval, getting very close to ε^* by binary search.

Using an oracle for the relaxed problem SIGN-REAL-NONZEROTUTTE(x, y) we can compute the sign whenever it is positive or negative, but we receive an unreliable answer for the sign of $f(\varepsilon)$ if $f(\varepsilon) = 0$. Nevertheless, we observe that having a reliable answer in this case is not important for the progress of the binary search. If the binary search queries the value of $f(\varepsilon)$ and $f(\varepsilon) \neq 0$, then the reply from the oracle is correct. Otherwise, it is still possible to recurse into a sub-interval that contains a zero of the function, as required. Thus, we have the following lemma.

Lemma 10.11. *For any algebraic real number* $y \in (-1, 0)$, SIGN-REAL-NONZEROTUTTE(x, y) *is* #**P***-hard, where* $x = 1 + 2/(y-1)$. *Also, the problems* DISTANCE-$(\pi/3)$-NONZERO-ARGISING(y) *and* COMPLEXAPX-NONZERO-ISING(y) *are* #**P***-hard.*

We next show how to further extend the #**P**-hardness interval-shrinking technique to obtain #**P**-hardness for the problem FACTOR-K-NONZERO-NORMISING(y). This requires new ideas, so we will provide more details. Let us return to the discussion of interval shrinking. Let $\eta = 1/21$ and $\rho = 22/21$. Instead of having an oracle for the sign of $f(\varepsilon) = -\varepsilon A + B$, we only will be able to assume that we have an oracle that, on input ε, returns a value $\hat{f}(\varepsilon)$ satisfying

$$(1-\eta)|f(\varepsilon)| \leqslant |f(\varepsilon)|/\rho \leqslant \hat{f}(\varepsilon) \leqslant \rho|f(\varepsilon)| \leqslant (1+\eta)|f(\varepsilon)|,$$

except that again the value $\hat{f}(\varepsilon)$ is completely unreliable if $f(\varepsilon) = 0$. Our strategy will be to divide the interval into 10 equal-length sub-intervals $[\varepsilon_i, \varepsilon_{i+1}]$ for $i \in \{0, \dots, 9\}$ with $\varepsilon_0 = \varepsilon'$ and $\varepsilon_{10} = \varepsilon''$. (The number 10 is not chosen to be optimal — however, it is easy to see that it suffices.) We then let s_i be the sign (positive, negative, or zero) of $\hat{f}(\varepsilon_i) - \hat{f}(\varepsilon_{i+1})$, for each $i \in \{0, \dots, 9\}$. The s_i values can be computed by the oracle. Now consider what happens if $\varepsilon_i < \varepsilon_{i+1} < \varepsilon^*$ (so $f(\varepsilon_i) > f(\varepsilon_{i+1}) > 0$). In this case,

$$\begin{aligned} \hat{f}(\varepsilon_i) - \hat{f}(\varepsilon_{i+1}) &\geqslant (1-\eta)f(\varepsilon_i) - (1+\eta)f(\varepsilon_{i+1}) \\ &= A(\varepsilon_{i+1} - \varepsilon_i - \eta(2\varepsilon^* - \varepsilon_i - \varepsilon_{i+1})). \end{aligned}$$

Now $\varepsilon_{i+1} - \varepsilon_i \geqslant \ell/10$. Also $\varepsilon^* - \varepsilon_i$ and $\varepsilon^* - \varepsilon_{i+1}$ are both at most ℓ. So since $\eta < 1/20$, s_i is positive. Similarly, if $\varepsilon^* < \varepsilon_i < \varepsilon_{i+1}$ (so $f(\varepsilon_{i+1}) < f(\varepsilon_i) < 0$), then

$$\begin{aligned}\hat{f}(\varepsilon_i) - \hat{f}(\varepsilon_{i+1}) &\geqslant (1-\eta)(-f(\varepsilon_i)) - (1+\eta)(-f(\varepsilon_{i+1})) \\ &= -A(\varepsilon_{i+1} - \varepsilon_i - \eta(2\varepsilon^* - \varepsilon_i - \varepsilon_{i+1})),\end{aligned}$$

so s_i is negative. If $\varepsilon_i \leqslant \varepsilon^*$ and $\varepsilon_{i+1} \geqslant \varepsilon^*$, then we don't know what the value of s_i will be. However, this is true for at most two values of i. So either s_0, s_1, s_2 and s_3 are all positive (in which case $\varepsilon_2 < \varepsilon^*$ and we can recurse on the interval $[\varepsilon_2, \varepsilon_{10}]$) or s_6, s_7, s_8 and s_9 are all negative (in which case $\varepsilon_8 > \varepsilon^*$ and we can recurse on the interval $[\varepsilon_0, \varepsilon_8]$). Either way, the interval shrinks to 4/5 of its original length.

Applying this idea in the proof of [GJ14, Lemma 1] yields the following.

Lemma 10.12. *Suppose that* γ_1 *and* γ_2 *are algebraic reals with* $\gamma_1 \in (-2,-1)$ *and* $\gamma_2 \notin [-2,0]$. *Then* FACTOR-($\frac{22}{21}$)-NONZERO-NORM2TUTTE(γ_1, γ_2) *is* #**P**-*hard.*

Proof. Apart from the interval shrinking idea discussed above, the proof is similar in structure to the proof of [GJ14, Lemma 1]. We defer some calculations (which are unchanged) to [GJ14] but we provide the rest of the proof to show how to get the stronger result. We use the fact that the following problem is #**P**-complete. This was shown by Provan and Ball [PB83].

Name #MINIMUM CARDINALITY (s,t)-CUT.

Instance A graph $G = (V, E)$ and distinguished vertices $s, t \in V$.

Output $|\{S \subseteq E : S \text{ is a minimum cardinality } (s,t)\text{-cut in } G\}|$.

We will give a Turing reduction from #MINIMUM CARDINALITY (s,t)-CUT to the problem FACTOR-($\frac{22}{21}$)-NONZERO-NORM2TUTTE(γ_1, γ_2).

Let G, s, t be an instance of #MINIMUM CARDINALITY (s,t)-CUT. Assume without loss of generality that G has no edge from s to t. Let $n = |V(G)|$ and $m = |E(G)|$. Assume without loss of generality that G is connected and that $m \geqslant n$ is sufficiently large. Let k be the size of a minimum cardinality (s,t)-cut in G and let C be the number of size-k (s,t)-cuts.

Let $q = 2$ and $M^* = 2^{4m}$. Let h be the smallest integer such that $(\gamma_2 + 1)^h - 1 > M^*$ and let $M = (\gamma_2 + 1)^h - 1$. Note that we can implement M from γ_2 via an h-thickening, and h is at most a polynomial in m.

Let $\delta = 4^m/M$. Let $\mathbf{M}$ be the constant weight function which gives every edge weight M. We will use the following facts:

$$qM^m(1-\delta) \leqslant Z_{st}(G; q, \mathbf{M}) \leqslant qM^m(1+\delta) \tag{10.8}$$

and

$$CM^{m-k}q^2(1-\delta) \leqslant Z_{s|t}(G; q, \mathbf{M}) \leqslant CM^{m-k}q^2(1+\delta). \tag{10.9}$$

Fact (10.8) follows from the fact that each of the (at most 2^m) terms in $Z_{st}(G; q, \mathbf{M})$, other than the term with all edges in A, has size at most $M^{m-1}q^n$ and $2^m M^{m-1} q^n \leqslant \delta M^m q$. Fact (10.9) follows from the fact that all terms in $Z_{s|t}(G; q, \mathbf{M})$ are complements of (s, t)-cuts. If more than k edges are cut, then the term is at most $M^{m-k-1}q^n$ and

$$2^m M^{m-k-1} q^n \leqslant \delta C M^{m-k} q^2.$$

For a parameter ε in the open interval $(0, 1)$ which we will tune later, let $\gamma' = -1 - \varepsilon \in (-2, -1)$. We will discuss the implementation of γ' later. Let G' be the graph formed from G by adding an edge from s to t. Let $\boldsymbol{\gamma}$ be the edge-weight function for G' that assigns weight M to every edge of G and assigns weight γ' to the new edge. Using the definition of the (random cluster) Tutte polynomial, Goldberg and Jerrum noted that

$$\begin{aligned} Z_{\mathrm{Tutte}}(2, \boldsymbol{\gamma}) &= Z_{st}(G; 2, \mathbf{M})(1 + \gamma') + Z_{s|t}(G; 2, \mathbf{M}) \left(1 + \frac{\gamma'}{2}\right) \\ &= -\varepsilon Z_{st}(G; 2, \mathbf{M}) + Z_{s|t}(G; 2, \mathbf{M}) \left(1 - \frac{1+\varepsilon}{2}\right). \end{aligned} \tag{10.10}$$

It is easily checked that $Z_{\mathrm{Tutte}}(2, \boldsymbol{\gamma})$ is positive if ε is sufficiently small ($\varepsilon = M^{-2m}$ will do) and it is negative at $\varepsilon = 1$. Thus, viewing $Z_{\mathrm{Tutte}}(2, \boldsymbol{\gamma})$ as a function of ε, we can perform interval shrinking (as discussed before the statement of the lemma) to find a value of ε for which

$Z_{\text{Tutte}}(2, \gamma)$ is very close to 0. The interval shrinking uses an oracle for FACTOR-$(\frac{22}{21})$-NONZERO-NORM2TUTTE(γ_1, γ_2).

If we find an ε where $Z_{\text{Tutte}}(q, \gamma) = 0$, then for this value of ε, we have $\varepsilon Z_{st}(G; q, \mathbf{M}) = Z_{s|t}(G; q, \mathbf{M})\left(1 - \frac{1+\varepsilon}{2}\right)$. Thus, using ε, we can calculate the fraction $Z_{s|t}(G; q, \mathbf{M})/Z_{st}(G; q, \mathbf{M})$. Plugging this (known) value into (10.8) and (10.9), we obtain

$$\frac{Cq(1-\delta)}{M^k(1+\delta)} \leqslant \frac{Z_{s|t}(G; q, \mathbf{M})}{Z_{st}(G; q, \mathbf{M})} \leqslant \frac{Cq(1+\delta)}{M^k(1-\delta)}.$$

Now, we don't know k, but C is an integer between 1 and 2^m, whereas $M > 2^{4m}$, so there is only one value of k that gives a solution C in the right range. Using the value of k, we can calculate C exactly.

Technical issues arise both because we are somewhat constrained in what values ε we can implement and because we won't be able to discover the exact value of ε that we need (but we will be able to approximate it closely). These technical issues provide no more difficulty than they did in [GJ14]. Suppose first that we are able, for any given $\varepsilon \in (M^{-2m}, 1)$ to implement $\gamma' = -1 - \varepsilon$. Then our basic strategy is to do the interval shrinking, repeatedly sub-dividing the current interval $\Theta(\log(M^{m^2}))$ times, so eventually we'll get an interval of width at most M^{-m^2} which contains an ε where $Z_{\text{Tutte}}(2, \gamma) = 0$. Goldberg and Jerrum [GJ14] have already shown that knowing such an interval enables the exact calculation of C (so having a small interval is OK — it is not necessary to know ε exactly).

The only issue, then, is implementing the weights $\gamma' = -1 - \varepsilon$ during the interval shrinking. As in [GJ14] we cannot expect to implement any particular desired γ' precisely. However, using stretching and thickening, we can implement a value that is within an additive error of $M^{-m^2}/20$ of any desired ε, and this suffices. The fact that we have algebraic, rather than rational, numbers is irrelevant since stretchings and thickenings can be computed on algebraic numbers. $\square$

Using stretching and thickening, we get the following corollary.

Corollary 10.13. *Suppose* $K > 1$ *and that* $y \in (-1, 0)$ *is an algebraic real number. Then* FACTOR-K-NONZERO-NORMISING(y) *is #**P**-hard.*

Proof. We first show that FACTOR-(22/21)-NONZERO-NORMISING(y) is #**P**-hard. Consider the edge interaction $y \in (-1, 0)$. Using the correspondence from (10.3) and (10.5), this corresponds directly to the quantity $\gamma_1 \in (-2, -1)$ in Lemma 10.12. We now consider how to use y to implement the quantity γ_2. A 2-thickening from (x, y) gives an effective weight (x', y') with $y' = y^2 \in (0, 1)$ and $x' = 2/(y' - 1) + 1 < -1$. Then a 2-stretch from (x', y') gives an effective weight (x'', y'') with $x'' = (x')^2 > 1$ and $y'' = 2/(x'' - 1) + 1 > 1$, corresponding $\gamma_2 > 0$, as required.

We apply Lemma 10.8 to reduce from FACTOR-(22/21)-NONZERO-NORMISING(y) to FACTOR-K-NONZERO-NORMISING(y). □

Using Lemma 10.4 and the trivial reduction from FACTOR-K-NONZERO-NORMISING(y) to FACTOR-K-NORMISING(y) and from COMPLEXAPX-NONZERO-ISING(y) to COMPLEXAPX-ISING(y) we get the following.

Corollary 10.14. *Let* $y \in (-1, 0)$ *be an algebraic real number. Then for any* $K > 1$*,* FACTOR-K-NORMISING(y) *and* COMPLEXAPX-NONZERO-ISING(y) *and* COMPLEXAPX-ISING(y) *are #**P**-hard.*

Complex Weights

Next we study complex weights.

Lemma 10.15. *Let* $\theta \in [0, 2\pi)$ *and* $\theta \notin \{0, \frac{\pi}{2}, \pi, \frac{3\pi}{2}\}$*. There exists a positive integer* k *such that* $k\theta \in (\frac{\pi}{2}, \pi) \cup (\pi, \frac{3\pi}{2})$ *modulo* 2π*.*

Proof. Clearly if $\theta \in (\frac{\pi}{2}, \pi) \cup (\pi, \frac{3\pi}{2})$ then we are done by letting $k = 1$. Otherwise $\theta \in (0, \frac{\pi}{2}) \cup (\frac{3\pi}{2}, 2\pi)$. If θ is a irrational fraction of 2π then we can go through the whole unit circle by taking multiple of θ. So assume $\theta = \frac{2\pi a}{b}$ where a and b are co-prime and $b = 3$ or $b \geqslant 5$ as $\theta \notin \{0, \frac{\pi}{2}, \pi, \frac{3\pi}{2}\}$. Moreover $b = 3$ contradicts to $\theta \in (0, \frac{\pi}{2}) \cup (\frac{3\pi}{2}, 2\pi)$. Hence $b \geqslant 5$ and there exists an integer $t \neq b/2$ such that $b < 4t < 3b$. As a and b are relatively prime, there exist integers l_1, l_2 such that $l_1 a + l_2 b = 1$ and $l_1 > 0$. It is easy to see that $tl_1\theta = \frac{2\pi t l_1 a}{b} = -2\pi t l_2 + \frac{2\pi t}{b}$. As $t/b \in (1/4, 1/2) \cup (1/2, 3/4)$ we have that $\frac{2\pi t}{b} \in (\frac{\pi}{2}, \pi) \cup (\pi, \frac{3\pi}{2})$. □

The following lemma enables us to determine the complexity of evaluating the Ising partition function when the complex edge interaction $y \in \overline{\mathbb{Q}}$ is on the unit circle.

Lemma 10.16. *Let* $y = e^{i\theta} \in \mathbb{C}$ *be an algebraic complex number such that* $\theta \in [0, 2\pi)$ *and* $\theta \notin \{0, \frac{\pi}{2}, \pi, \frac{3\pi}{2}\}$. *There exists an algebraic real number* $y' \in (-1, 0)$ *that can be implemented by a sequence of stretchings and thickenings from* y.

Proof. By Lemma 10.15, there exists a positive integer k that $k\theta \in (\frac{\pi}{2}, \pi) \cup (\pi, \frac{3\pi}{2})$. As a k-thickening realizes $y^k = e^{ik\theta}$, we may assume $\theta \in (\frac{\pi}{2}, \pi) \cup (\pi, \frac{3\pi}{2})$.

Since $\theta \notin \{0, \frac{\pi}{2}, \pi, \frac{3\pi}{2}\}$, we have $\cos\theta \neq 1$ and $\sin\theta\cos\theta \neq 0$. The latter implies that $\sin\theta + \cos\theta \neq 1$. Let $x = \frac{y+1}{y-1}$. Note that $x = \frac{\sin\theta}{\cos\theta - 1}i$. Moreover $\theta \in (\frac{\pi}{2}, \pi) \cup (\pi, \frac{3\pi}{2})$, implies that $\cos\theta < 0$ and hence $|x| < 1$. We do a 2-stretch and the effective weight is $y' = 1 - \frac{2}{|x|^2+1} \in (-1, 0)$. □

Combining Lemma 10.16 with Proposition 10.6, Corollary 10.13, Lemma 10.11, and Corollary 10.14 we get the following corollary.

Corollary 10.17. *Let* $y = e^{i\theta} \in \mathbb{C}$ *be an algebraic complex number such that* $\theta \in [0, 2\pi)$ *and* $\theta \notin \{0, \frac{\pi}{2}, \pi, \frac{3\pi}{2}\}$. *Then for any* $K > 1$, FACTOR-K-NONZERO-NORMISING(y), DISTANCE-($\pi/3$)-NONZERO-ARGISING(y) *and* COMPLEXAPX-NONZERO-ISING(y) *are* #**P**-*hard. Hence, so are the un-relaxed versions of all three problems.*

The hardness on the unit circle extends directly to the whole imaginary axis.

Lemma 10.18. *For any* $y = ri$ *and* $r \neq 0, \pm 1$ *where* r *is algebraic. There exists an algebraic real number* $y' \in (-1, 0)$ *that can be implemented by a sequence of stretchings and thickenings from* y.

Proof. If $0 < |y| < 1$, then a 2-thickening yields effective weight $y^2 = -r^2 \in (-1, 0)$. Let $y' = -r^2$ and the claim holds.

Otherwise suppose $|y| > 1$. We know that a k-stretch yields the weight $z_k = 1 + 2/(x^k - 1)$ where $x = 1 + 2/(y - 1) = (y + 1)/(y - 1)$. Re-arranging, we find that $z_k = \frac{(y+1)^k + (y-1)^k}{(y+1)^k - (y-1)^k}$. We will now argue that z_k is purely imaginary. To see this, note that monomials in the numerator all have degrees of the same parity as k, whereas those in the denominator have degrees of the same parity as $k - 1$. Therefore, it must be the case that the numerator is real and the denominator is purely imaginary, or vice versa. In either case, z_k is purely imaginary. Therefore, if we can find a positive integer k such that $0 < |z_k| < 1$ then we have reduced our problem to the previous case.

Since y is purely imaginary, we have that $|y+1| = |y-1|$. Since $x = (y+1)/(y-1)$, this implies that $|x| = 1$. It is easy to see that $0 < |z_k| < 1$ if and only if $|x^k + 1| < |x^k - 1|$ and $x^k \neq -1$. This in turn is equivalent to $\arg\left(x^k\right) \in \left(\frac{\pi}{2}, \pi\right) \cup \left(\pi, \frac{3\pi}{2}\right)$. By Lemma 10.15, such a k always exists unless $\arg(x) = \frac{t\pi}{2}$ where $t = 0, 1, 2, 3$, in which case $y = \pm 1, \pm i$ and contradicts our assumption. □

Combining Lemma 10.18 with Proposition 10.6, Corollary 10.13, Lemma 10.11, and Corollary 10.14, we get the following corollary.

Corollary 10.19. *Let $y = ri$ where $r \neq 0, \pm 1$ and r is algebraic. Let $K > 1$. Then* FACTOR-K-NONZERO-NORMISING(y)*,* DISTANCE-$(\pi/3)$-NONZERO-ARGISING(y) *and* COMPLEXAPX-NONZERO-ISING(y) *are #**P**-hard. Hence, so are the un-relaxed versions of all three problems.*

Finally, this hardness can be extended to some algebraic complex numbers off of the unit circle.

Lemma 10.20. *Let $y = re^{i\theta}$ be an algebraic complex number such that $r > 0$ and $\theta = \frac{a\pi}{2b}$, where a and b are two co-prime positive integers and a is odd. There exists an algebraic real number $y' \in (-1, 0)$ that can be implemented by a sequence of stretchings and thickenings from y.*

Proof. If $r = 1$ then we are done by Lemma 10.16. Otherwise $r \neq 1$ and by a b-thickening it reduces to the case of Lemma 10.18. □

Corollary 10.21. *Let $y = re^{i\theta}$ be an algebraic complex number such that $r > 0$ and $\theta = \frac{a\pi}{2b}$, where a and b are two co-prime positive integers and a is odd. Then for any $K > 1$,* FACTOR-K-NONZERO-NORMISING(y)*,* DISTANCE-$(\pi/3)$-NONZERO-ARGISING(y) *and* COMPLEXAPX-NONZERO-ISING(y) *are #**P**-hard. Hence, so are the un-relaxed versions of all three problems.*

To obtain obtain **NP**-hardness results for other values of y, we start with the well-known **NP**-hard problem MAX-CUT.

Name MAX-CUT

Instance A (multi)graph G and a positive integer b.

Output Is there a cut of size at least b.

Lemma 10.22. *Suppose* $K > 1$. *Let* y *be an algebraic complex number such that* $|y| < 1$ *and* $y \neq 0$. *Then* FACTOR-K-NONZERO-NORMISING(y) *is* **NP**-*hard and so is* COMPLEXAPX-NONZERO-ISING(y).

Proof. We will reduce MAX-CUT to FACTOR-K-NONZERO-NORMISING(y). Given a graph G and a constant b, we want to decide whether G has a cut of size at least b. We do a k-thickening on G, where k is the least positive integer such that $2^m|y|^k < 1/4$. Then the effective edge weight is $y_k = y^k$. Clearly $|y_k| = |y|^k < 1$.

Suppose the maximum cut of G has size c. Now rewrite (10.1) as

$$Z_{\text{Ising}}(G; y_k) = \sum_{i=0}^{c} C_i y_k^{m-i},$$

where m is the number of edges in G and C_i is the number of configurations under which there are exactly i bichromatic edges. Since the maximum cut of G has size c and G has m edges, $\sum_{i=0}^{m-c} C_i = 2^m$. Also, since $2^m|y_k| < 1$, the $i = c$ term dominates the sum, so $Z_{\text{Ising}}(G; y_k)$ is not equal to 0.

If $c \geqslant b$, then our choice of k together with the triangle inequality implies that

$$\begin{aligned} |Z_{\text{Ising}}(G; y_k)| \;&=\; |C_c y_k^{m-c} + \sum_{i=0}^{c-1} C_i y_k^{m-i}| \;>\; C_c|y_k|^{m-c} - 2^m|y_k|^{m-c+1} \\ &>\; |y_k|^{m-c}|1 - 2^m|y|^k| \;>\; \tfrac{3}{4}|y_k|^{m-b}. \end{aligned}$$

Otherwise we have $c \leqslant b - 1$ and

$$\begin{aligned} |Z_{\text{Ising}}(G; y_k)| \;&=\; |\sum_{i=0}^{c} C_i y_k^{m-i}| \;<\; \sum_{i=0}^{c} C_i |y_k|^{m-i} \\ &\leqslant\; 2^m|y_k|^{m-b+1} \;<\; \tfrac{1}{4}|y_k|^{m-b} \end{aligned}$$

again by the triangle inequality and $2^m|y_k| < 1/4$. Therefore we could solve MAX-CUT in polynomial time using an oracle for FACTOR-1.1-NONZERO-NORMISING(y_k). By Proposition 10.6

it suffices to use an oracle for FACTOR-1.1-NONZERO-NORMISING(y). By Lemma 10.8, an oracle for FACTOR-K-NONZERO-NORMISING(y) will do. Finally, Lemma 10.4 gives the result for COMPLEXAPX-NONZERO-ISING(y). □

The other case, when the norm of y is larger than 1, can be shown to be **NP**-hard by reduction from the previous case, unless the edge weight is real.

Lemma 10.23. *Suppose* $K > 1$. *Let* y *be an algebraic complex number such that* $|y| > 1$ *and* $y \notin \mathbb{R}$. *Then* FACTOR-K-NONZERO-NORMISING(y) *is* **NP**-*hard and so is* COMPLEXAPX-NONZERO-ISING(y).

Proof. We will prove that there exists a positive integer k such that the effective weight y_k of a k-stretch satisfies $|y_k| < 1$. Then we are done by Lemma 10.22.

Recall that $y_k = \frac{x^k+1}{x^k-1}$ where $x = \frac{y+1}{y-1}$. Clearly $|y_k| < 1$ if and only if $|x^k + 1| < |x^k - 1|$. The latter is equivalent to $\arg(x^k) = k\arg(x) \in (\pi/2, 3\pi/2)$ (plus some multiple of 2π). Let $\theta = \arg(x) \in [0, 2\pi)$. The fact that $|y| > 1$ implies that $\theta \in [0, \pi/2) \cup (3\pi/2, 2\pi)$. If $\theta = 0$, then $y \in \mathbb{R}$, which is a contradiction. Therefore $\theta \in (0, \pi/2) \cup (3\pi/2, 2\pi)$. By Lemma 10.15, there exists a positive integer k such that $k\theta \in (\pi/2, 3\pi/2)$ modulo 2π. This is exactly what we need. Moreover, k does not depend on the input G. This finishes our proof. □

Proof of Theorems 10.2 and 10.3

Theorems 10.2 and 10.3 follow from the following combined theorem. The hardness result in Item 3 of Theorem 10.2 (and its counterpart in Theorem 10.3) follows from Item 9 of the combined theorem.

Theorem 10.24. *Let* $y = re^{i\theta}$ *be an algebraic complex number with* $\theta \in [0, 2\pi)$. *Suppose* $K > 1$.

1. *If* $y = 0$ *or if* $r = 1$ *and* $\theta \in \{0, \frac{\pi}{2}, \pi, \frac{3\pi}{2}\}$, *then* FACTOR-K-NORMISING(y), DISTANCE-($\pi/3$)-ARGISING(y) *and* COMPLEXAPX-ISING(y) *are in* **FP**.

2. *If* $y > 1$ *is a real number then* FACTOR-K-NORMISING(y) *is in* **RP** *and* DISTANCE-($\pi/3$)-ARGISING(y) *is in* **FP**.

3. *If* y *is a real number in* $(0, 1)$, *then* DISTANCE-($\pi/3$)-ARGISING(y) *is in* **FP**.

4. *If* $y < -1$ *is a real number, then* Factor-K-Nonzero-NormIsing(y) *is equivalent in complexity to the problem of approximately counting perfect matchings in graphs and* Distance-($\pi/3$)-ArgIsing(y) *is in* **FP**.

5. *If* y *is a real number in* $(-1, 0)$, *then* Factor-K-Nonzero-NormIsing(y) *and* Distance-($\pi/3$)-Nonzero-ArgIsing(y) *are* #**P**-*hard.*

6. *If* $r = 1$ *and* $\theta \notin \{0, \frac{\pi}{2}, \pi, \frac{3\pi}{2}\}$, *then* Factor-K-Nonzero-NormIsing(y), Distance-($\pi/3$)-Nonzero-ArgIsing(y), *and* ComplexApx-Nonzero-Ising(y) *are* #**P**-*hard.*

7. *If* $\theta \in \{\frac{\pi}{2}, \frac{3\pi}{2}\}$ *and* $r \notin \{-1, 0, 1\}$, *then* Factor-K-Nonzero-NormIsing(y), Distance-($\pi/3$)-Nonzero-ArgIsing(y), *and* ComplexApx-Nonzero-Ising(y) *are* #**P**-*hard.*

8. *If* $r > 0$ *and* $\theta = \frac{a\pi}{2b}$, *where* a *and* b *are two positive integers that are co-prime and* a *is odd, then* Factor-K-Nonzero-NormIsing(y), Distance-($\pi/3$)-Nonzero-ArgIsing(y), *and* ComplexApx-Nonzero-Ising(y) *are* #**P**-*hard.*

9. *If* $r < 1$ *and* $y \neq 0$, *then* Factor-K-Nonzero-NormIsing(y) *and* ComplexApx-Nonzero-Ising(y) *are* **NP**-*hard.*

10. *If* $r > 1$ *and* $\theta \notin \{0, \pi\}$, *then* Factor-K-Nonzero-NormIsing(y) *is* **NP**-*hard, and so is* ComplexApx-Nonzero-Ising(y).

Proof. Item 1 is from [JVW90]. The randomised algorithm referred to in Item 2 is from [JS93]. See also Lemma 10.7 and the surrounding text for a discussion of algebraic numbers and accuracy parameters. The deterministic algorithm referred to in Items 2 and 3 is trivial because the argument of a positive real number is 0. The approximation equivalence in Item 4 is from [GJ08], since one can decide in polynomial time the existence of perfect matchings to lift the non-zero restriction. The deterministic sign algorithm in Item 4 is from [GJ14]. Item 5 is from Lemma 10.11 and Corollary 10.13. Item 6 is from Corollary 10.17. Item 7 is from Corollary 10.19. Item 8 is from Corollary 10.21. Item 9 is from Lemma 10.22. Finally,item 10 is from Lemma 10.23. □

10.6 Quantum Circuits and Counting Complexity

In this section we explain the connection between quantum computation and complex weighted Ising models. We begin with some basic notions about quantum circuits. We view qubits $|0\rangle$ and $|1\rangle$ as column vectors $\left[\begin{smallmatrix}1\\0\end{smallmatrix}\right]$ and $\left[\begin{smallmatrix}0\\1\end{smallmatrix}\right]$. Similarly $\langle 0|$ and $\langle 1|$ are row vectors $(1,0)$ and $(0,1)$. For $\mathbf{x} \in \{0,1\}^n$, let $|\mathbf{x}\rangle$ denote the tensor product $\otimes_{j=1}^n |x_j\rangle$ and similarly $\langle \mathbf{x}|$.

Suppose C is a quantum circuit on n qubits and consists of m quantum gate $U_1, \ldots, U_m$ sequentially. A quantum gate is a function taking k input and k output variables and returning a value in $\mathbb{C}$. Such a gate is called k-local and has a natural 2^k by 2^k square unitary matrix representation. In a circuit we also need to specify on which qubits the gate acts upon. To make the notation uniform we view unaffected qubits as simply copied and associate each quantum gate with the following 2^n by 2^n square unitary matrix. Let U be a quantum gate and $\mathbf{x}, \mathbf{y} \in \{0,1\}^n$ two vectors specifying the input and output on all n qubits. Define the 2^n by 2^n matrix M_U corresponding to gate U as $M_{U;\mathbf{x},\mathbf{y}} = U(\mathbf{x}, \mathbf{y})$.

For example, let H be the Hadamard gate $\frac{1}{\sqrt{2}}\left[\begin{smallmatrix}1 & 1\\1 & -1\end{smallmatrix}\right]$ acting on the first qubit and suppose there are two qubits in total, illustrated as in Figure 10.2. Then the matrix M_H is $\frac{1}{\sqrt{2}}\left[\begin{smallmatrix}1 & 1\\1 & -1\end{smallmatrix}\right] \otimes \left[\begin{smallmatrix}1 & 0\\0 & 1\end{smallmatrix}\right] = \frac{1}{\sqrt{2}}\left[\begin{smallmatrix}1 & 0 & 1 & 0\\0 & 1 & 0 & 1\\1 & 0 & -1 & 0\\0 & 1 & 0 & -1\end{smallmatrix}\right]$.

Using this notation, given an input $\mathbf{x} \in \{0,1\}^n$, the output of the quantum circuit C is a random variable $\mathbf{Y}$ subject to the distribution

$$\Pr_C(\mathbf{Y} = \mathbf{y}) = \left| \langle \mathbf{y}| \prod_{j=1}^{m} M_{U_{m+1-j}} |\mathbf{x}\rangle \right|^2, \tag{10.11}$$

where $\mathbf{y} \in \{0,1\}^n$. It is not necessary that we measure all qubits in the output. We may measure a subset I of all n qubits. Let $\mathbf{y}' \in \{0,1\}^s$ where $|I| = s$. Then the output is a random variable $\mathbf{Y}'$ subject to the distribution

$$\Pr_{C;I}(\mathbf{Y}' = \mathbf{y}') = \sum_{\mathbf{z} \in \{0,1\}^n \text{ such that } \mathbf{z}|_I = \mathbf{y}'} \Pr_C(\mathbf{Y} = \mathbf{z}). \tag{10.12}$$

Alternatively, we may treat such marginal probability in the counting perspective, as a par-

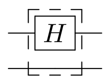


Figure 10.2: Gate H applying only on the first qubit.

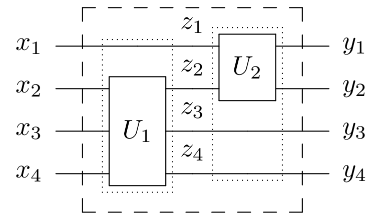


Figure 10.3: Two quantum gates U_1 and U_2 composed together.

tition function in the "sum of product" fashion. First let us consider composing two quantum gates, say U_1 and U_2. Let input variables of U_1 be $\{x_j\}$, intermediate variables between U_1 and U_2 be $\{z_j\}$, and outputs of U_2 be $\{y_j\}$. Then the composition U of U_1 followed by U_2 is given by

$$U(\mathbf{x}, \mathbf{y}) = \sum_{\sigma:\{z_j\}\to\{0,1\}} U_1(\mathbf{x}, \sigma(\mathbf{z}))U_2(\sigma(\mathbf{z}), \mathbf{y}), \tag{10.13}$$

where the summation is over all possible assignment of $\mathbf{z}$ to $\{0, 1\}$. Figure 10.3 illustrates the composition of gate U_1 acting upon qubits $2, 3, 4$ followed by U_2 acting upon $1, 2$. In the matrix notation, it is easy to see that $M_U = M_{U_1}M_{U_2}$.

We associate an intermediate variable $z_{j,k}$ to each edge on qubit k between gate U_j and U_{j+1} for all $2 \leqslant j \leqslant m-1$ and $1 \leqslant k \leqslant n$. Denote by $\mathbf{z_j}$ the vector $\{z_{j,k} \mid 1 \leqslant k \leqslant n\}$ and $\mathbf{z} = \cup_{j=2}^{m-1}\mathbf{z_j}$. As the initial input and output of a quantum circuit are column vectors and row vectors respectively, they may be treated as function/gates with no output variables or no input variables. In particular, on the product input state $|\mathbf{x}\rangle$ input variables are set to $\{x_k\}$ where $\mathbf{x} \in \{0, 1\}^n$. Using (10.13) recursively we can rewrite (10.11) as follows:

$$\Pr_C(\mathbf{Y} = \mathbf{y}) = \left| \sum_{\sigma:\mathbf{z}\to\{0,1\}} U_1(\mathbf{x}, \sigma(\mathbf{z_1}))U_m(\sigma(\mathbf{z_{m-1}}), \mathbf{y}) \prod_{j=2}^{m-1} U_j(\sigma(\mathbf{z_{j-1}}), \sigma(\mathbf{z_j}))\right|^2. \tag{10.14}$$

To simulate classically a quantum circuit, one can either (approximately) compute the probability $\Pr_C(\mathbf{Y} = \mathbf{y})$ — this is called "strong simulation" — or one can sample from a distribution that is sufficiently close to the one given by (10.11) or (10.14). This is called "weak simulation"

IQP and the Ising partition function

IQP, which stands for "instantaneous quantum polynomial time", is characterised by a restricted class of quantum circuits introduced by Shepherd and Bremner [SB09]. Bremner et al. [BJS11] showed that if **IQP** can be simulated classically in the sense of "weak simulation" with multiplicative error, then the polynomial hierarchy collapses to the third level. Fujii and Morimae [FM13] showed that the marginal probabilities of possible outcomes of **IQP** circuits correspond to partition functions of Ising models with complex edge weights.

The key property of **IQP** is that all gates are diagonal in the $|0\rangle \pm |1\rangle$ basis. Therefore all gates are commutable. In other words, there is no temporal structure and hence it is called "instantaneous". Let H be the Hadamard gate $\frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$. If a gate U is diagonal in the $|0\rangle \pm |1\rangle$ basis, there exists a diagonal matrix D such that $M_U = H^{\otimes n} D H^{\otimes n}$. Moreover H is its own inverse; That is, $HH = I_2$. Any two H's between each pair of gates cancel. This leads to an alternative view of **IQP** circuit in which each qubit line starts and ends with an H gate and all gates in between are diagonal.

Definition 10.25. *An* **IQP** *circuit on* n *qubit lines is a quantum circuit with the following structure: each qubit line starts and ends with an* H *gate, and all other gates are diagonal.*

We will focus particularly on $1,2$-local **IQP**, which means that every intermediate gate acts on 1 or 2 qubits. It was shown that a classical weak simulation of **IQP** with multiplicative error implies the polynomial hierarchy collapse to the third level [BJS11]. Let $Z = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$. The hardness of simulation holds even if we restrict gates to the phase gate $e^{i(\pi/8)Z} = \begin{bmatrix} e^{i\pi/8} & 0 \\ 0 & e^{-i\pi/8} \end{bmatrix}$ and controlled Z-gate $CZ = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix}$ other than H gates on two ends of each line. We will show that these circuits correspond to Ising models with complex edge interactions. Therefore the strong simulation of these circuits, which is to compute the marginal probabilities, is #**P**-hard, even allowing an error of any factor $K > 1$.

To show the relationship between these circuits and Ising partition functions, it is convenient to use another set of gates. Let $P_\theta = e^{i\theta Z} = \begin{bmatrix} e^{i\theta} & 0 \\ 0 & e^{-i\theta} \end{bmatrix}$ and $R_\theta = e^{i\theta Z \otimes Z} = \begin{bmatrix} e^{i\theta} & 0 & 0 & 0 \\ 0 & e^{-i\theta} & 0 & 0 \\ 0 & 0 & e^{-i\theta} & 0 \\ 0 & 0 & 0 & e^{i\theta} \end{bmatrix}$. Note from (10.11) that we may multiply a gate by any norm 1 constant without affecting the outcome of the gate. By multiplying by $e^{-i\pi/4}$, we may decompose CZ as:

$$e^{-i\pi/4}\begin{bmatrix}1&0&0&0\\0&1&0&0\\0&0&1&0\\0&0&0&-1\end{bmatrix} = \begin{bmatrix}e^{i\pi/8}&0&0&0\\0&e^{-i\pi/8}&0&0\\0&0&e^{-i\pi/8}&0\\0&0&0&e^{i\pi/8}\end{bmatrix}^2 \times \begin{bmatrix}e^{i\pi/8}&0&0&0\\0&e^{-i\pi/8}&0&0\\0&0&e^{i\pi/8}&0\\0&0&0&e^{-i\pi/8}\end{bmatrix}^{14}$$
$$\times \begin{bmatrix}e^{i\pi/8}&0&0&0\\0&e^{i\pi/8}&0&0\\0&0&e^{-i\pi/8}&0\\0&0&0&e^{-i\pi/8}\end{bmatrix}^{14}$$
$$= \left(R_{\pi/8}\right)^2 \left(P_{\pi/8}\otimes I_2\right)^{14} \left(I_2\otimes P_{\pi/8}\right)^{14}. \tag{10.15}$$

Hence we can replace every CZ gate on qubits j, k by 2 copies of $R_{\pi/8}$ on j, k, 14 copies of $P_{\pi/8}$ on qubit j, and 14 $P_{\pi/8}$ on qubit k. It is easy to see that $R_{\pi/8}$ can be replaced by CZ and $P_{\pi/8}$ as well. We may therefore assume every gate is either $P_{\pi/8}$ on 1 qubit or $R_{\pi/8}$ on 2 qubits without changing the power of the circuit. In general we give the following definition.

Definition 10.26. *An* $\mathbf{IQP}_{1,2}(\theta)$ *circuit on* n *qubit lines is a quantum circuit with the following structure: each qubit line starts and ends with an* H *gate, and every other gate is either* P_θ *on* 1 *qubit or* R_θ *on* 2 *qubits. We assume the input state is always* $|0^n\rangle$.

An example $\mathbf{IQP}_{1,2}(\theta)$ circuit is given in Figure 10.4.

We consider the following strong simulation problem where $K > 1$ is an error parameter.

Name FACTOR-K-STRONGSIMIQP$_{1,2}(\theta)$.

Instance An $\mathbf{IQP}_{1,2}(\theta)$ circuit C, a subset $I \subseteq [n]$ of lines, and a string $\mathbf{y} \in \{0,1\}^{|I|}$.

Output A rational number p such that $\Pr_{C;I}(\mathbf{Y}=\mathbf{y})/K \leqslant p \leqslant K \Pr_{C;I}(\mathbf{Y}=\mathbf{y})$.

We will show that $\mathbf{IQP}_{1,2}(\theta)$ circuits correspond to Ising models with complex edge interactions. Therefore their strong simulation is #**P**-hard, even allowing an error of any factor $K > 1$.

Theorem 10.27. *Suppose* $K > 1$ *and* $\theta \in (0, 2\pi)$. *If* $e^{i\theta}$ *is an algebraic complex number and* $e^{i8\theta} \neq 1$, *then* FACTOR-K-STRONGSIMIQP$_{1,2}(\theta)$ *is #**P**-hard.*

The relationship between $\mathbf{IQP}_{1,2}(\theta)$ circuits and Ising models was first observed by Fujii and Morimae [FM13]. These connections will be shown next. For completeness we include our own proofs, which have a more combinatorial flavour than the original ones by Fujii and

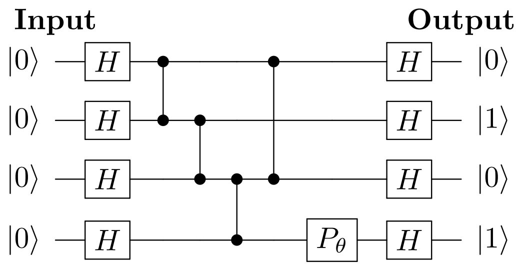


Figure 10.4: An $\mathbf{IQP}_{1,2}(\theta)$ circuit. We use two solid dots to denote R_θ gate as it is diagonal and symmetric.

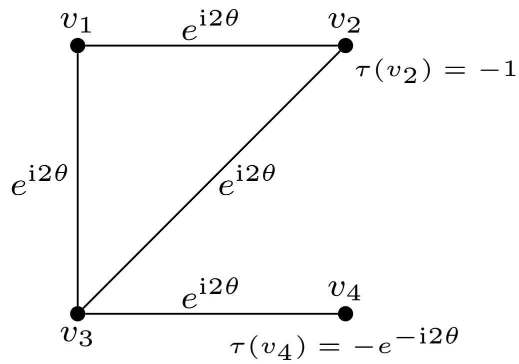


Figure 10.5: The equivalent Ising instance to the circuit in Figure 10.4.

Morimae [FM13]. We introduce the following non-uniform Ising model which has been studied previously. See, for example [Sok05]. Let $G = (V, E)$ be a (multi)graph. The edge interaction is specified by a function $\varphi : E \to \mathbb{C}$ and the external field is specified by a function $\tau : V \to \mathbb{C}$. The partition function is defined as

$$Z_{\text{Ising}}(G; \varphi, \tau) = \sum_{\sigma: V \to \{0,1\}} \prod_{e=(v_j, v_k) \in E} \varphi(e)^{\delta(\sigma(v_j), \sigma(v_k))} \prod_{v \in V} \tau(v)^{\sigma(v)}, \tag{10.16}$$

where $\delta(x, y) = 1$ if $x = y$ and $\delta(x, y) = 0$ if $x \neq y$. We write $Z_{\text{Ising}}(G; y, \tau)$ when $\varphi(e) = y$ is a constant function and similarly $Z_{\text{Ising}}(G; \varphi, \lambda)$ when $\tau(v) = \lambda$. Notice that this notation is consistent with (10.1).

We will show that the following problem is related to FACTOR-K-STRONGSIMIQP$_{1,2}(\theta)$ when $e^{i\theta}$ is a root of unity.

Name FACTOR-K-NORMIQPISING(θ)

Instance A (multi)graph G with an edge interaction function $\varphi(-)$ taking value $e^{i\theta}$ or $e^{-i\theta}$, and an external field function τ so that for each vertex v there are non-negative integers a_v and b_v so that $\tau(v) = (-1)^{a_v} \left(e^{i\theta}\right)^{b_v}$ or $\tau(v) = (-1)^{a_v} \left(e^{-i\theta}\right)^{b_v}$.

Output A rational number p such that $|Z_{\text{Ising}}(G; \varphi, \tau)|/K \leqslant p \leqslant K|Z_{\text{Ising}}(G; \varphi, \tau)|$.

We will first consider inputs to $\mathbf{IQP}_{1,2}(\theta)$ where $I = [n]$ so all qubits are measured. Given an $\mathbf{IQP}_{1,2}(\theta)$ circuit C on n qubits and a string $\mathbf{y} \in \{0,1\}^n$, we can construct a non-uniform Ising instance G_C with edge interaction $e^{i2\theta}$ and external field $\tau_{C;\mathbf{y}}$ such that

$$\Pr_C(\mathbf{Y} = \mathbf{y}) = 2^{-2n} \left| Z_{\text{Ising}}(G_C; e^{i2\theta}, \tau_{C;\mathbf{y}}) \right|^2. \tag{10.17}$$

The construction is as follows. The vertex set $\{v_j\}$ contains n vertices and each vertex corresponds to a qubit. For each gate R_θ on two qubits j, k, add an edge (j, k) in G_C. For qubit j, let p_j be the number of gates P_θ acting on qubit j in C. Let $\tau_{C;\mathbf{y}}(v_j) = e^{-i(2p_j\theta)}(-1)^{y_j}$. An example of the construction is given in Figure 10.5.

Lemma 10.28. *Let* C *be an* $\mathbf{IQP}_{1,2}(\theta)$ *circuit on* n *qubits and* $\mathbf{y} \in \{0,1\}^n$ *be the output. Let* G_C *and* $\tau_{C;\mathbf{y}}$ *be constructed as above. Then* (10.17) *holds.*

Proof. Suppose C is composed sequentially by $U_1 = H^{\otimes n}$, U_2, ..., U_{m-1}, $U_m = H^{\otimes n}$, where U_j is either P_θ on 1 qubit or R_θ on 2 qubits for $2 \leqslant j \leqslant m-1$. Notice that $U_1(\mathbf{x}, \mathbf{x}') = U_m(\mathbf{x}, \mathbf{x}') = 2^{-n/2} \prod_{k=1}^n (-1)^{x_k x'_k}$. As the input $|\mathbf{x}\rangle = |0^n\rangle$, we can rewrite (10.14):

$$\begin{aligned}
\Pr_C(\mathbf{Y} = \mathbf{y}) &= \left| \sum_{\sigma:\mathbf{z}\to\{0,1\}} U_1(\mathbf{0}, \sigma(\mathbf{z_1}))U_m(\sigma(\mathbf{z_{m-1}}), \mathbf{y}) \prod_{j=2}^{m-1} U_j(\sigma(\mathbf{z_{j-1}}), \sigma(\mathbf{z_j})) \right|^2 \\
&= \left| 2^{-n} \sum_{\sigma:\mathbf{z}\to\{0,1\}} \prod_{k=1}^{n} (-1)^{0\cdot\sigma(z_{1,k})} \prod_{k=1}^{n} (-1)^{y_k\sigma(z_{m-1,k})} \prod_{j=2}^{m-1} U_j(\sigma(\mathbf{z_{j-1}}), \sigma(\mathbf{z_j})) \right|^2 \\
&= 2^{-2n} \left| \sum_{\sigma:\mathbf{z}\to\{0,1\}} \prod_{k=1}^{n} (-1)^{y_k\sigma(z_{m-1,k})} \prod_{j=2}^{m-1} U_j(\sigma(\mathbf{z_{j-1}}), \sigma(\mathbf{z_j})) \right|^2
\end{aligned} \tag{10.18}$$

Let Q denote the quantity inside the norm, that is,

$$Q := \sum_{\sigma:\mathbf{z}\to\{0,1\}} \prod_{k=1}^{n} (-1)^{y_k\sigma(z_{m-1,k})} \prod_{j=2}^{m-1} U_j(\sigma(\mathbf{z_{j-1}}), \sigma(\mathbf{z_j})).$$

Since U_j's are diagonal for $2 \leqslant j \leqslant m-1$, any configuration σ with a non-zero contribution to Q must satisfy that for any k, $\sigma(z_{1,k}) = \sigma(z_{2,k}) = \cdots = \sigma(z_{m-1,k})$. Therefore we may replace $z_{j,k}$ by a single variable v_k for all $1 \leqslant j \leqslant m-1$ so that

$$Q = \sum_{\sigma:V\to\{0,1\}} \prod_{k=1}^{n} (-1)^{y_k\sigma(v_k)} \prod_{j=2}^{m-1} U_j(\sigma(V), \sigma(V)).$$

Moreover, if U_j is the gate P_θ on qubit k, then $U_j(\sigma(V), \sigma(V)) = e^{i\theta}\left(e^{-i2\theta}\right)^{\sigma(\nu_k)}$. If U_j is the gate R_θ on qubits k_1 and k_2, then $U_j(\sigma(V), \sigma(V)) = e^{-i\theta}\left(e^{i2\theta}\right)^{\delta(\sigma(\nu_{k_1}),\sigma(\nu_{k_2}))}$, where $\delta(x, y) = 1$ if $x = y$ and $\delta(x, y) = 0$ if $x \neq y$. Recall that p_k is the number of P_θ gates on qubit k and $\tau_{C;\mathbf{y}}(\nu_k) = e^{-i(2p_k\theta)}(-1)^{y_k}$. Collecting all the contributions, we have

$$\begin{aligned} Q &= e^{i(m_1-m_2)\theta} \sum_{\sigma:V\to\{0,1\}} \left(e^{i2\theta}\right)^{m(\sigma)} \prod_{k=1}^{n} (-1)^{y_k\sigma(\nu_k)} \left(e^{-i2\theta}\right)^{p_k\sigma(\nu_k)} \\ &= e^{i(m_1-m_2)\theta} \sum_{\sigma:V\to\{0,1\}} \left(e^{i2\theta}\right)^{m(\sigma)} \prod_{k=1}^{n} \tau_{C;\mathbf{y}}(\nu_k)^{\sigma(\nu_k)} \\ &= e^{i(m_1-m_2)\theta} Z_{\text{Ising}}(G_C; e^{i2\theta}, \tau_{C;\mathbf{y}}), \end{aligned} \tag{10.19}$$

where m_j is the number of j qubit(s) gates for $j = 1, 2$, and, from (10.1), $m(\sigma)$ is the number of monochromatic edges under σ. We get (10.17) by substituting (10.19) in (10.18). □

Similar results hold when some qubits are not measured. To show it, we need the following fact. It can be viewed as an application of Parsevals's identity on the length-2^n vector $\{C_\mathbf{z}\}$ indexed by $\mathbf{z} \in \{0, 1\}^n$ over an orthonormal basis $\{e_\mathbf{z}\}$ where basis element $e_\mathbf{z}$ has value $2^{-\frac{n}{2}}(-1)^{\mathbf{z}\cdot\mathbf{z}'}$ in position $\mathbf{z}'$. We include a proof for completeness.

Claim 10.29. *Let $\{C_\mathbf{z}\}$ be 2^n complex numbers where $\mathbf{z}$ runs over $\{0, 1\}^n$. Then we have*

$$\sum_{\mathbf{z}'\in\{0,1\}^n} \left| \sum_{\mathbf{z}\in\{0,1\}^n} C_\mathbf{z}(-1)^{\mathbf{z}\cdot\mathbf{z}'} \right|^2 = 2^n \sum_{\mathbf{z}\in\{0,1\}^n} |C_\mathbf{z}|^2 .$$

Proof. Notice that for two complex numbers A and B,

$$\begin{aligned} |A+B|^2 + |A-B|^2 &= \left(|A|^2 + |B|^2 - 2|A||B|\cos\theta\right) + \left(|A|^2 + |B|^2 + 2|A||B|\cos\theta\right) \\ &= 2\left(|A|^2 + |B|^2\right) \end{aligned} \tag{10.20}$$

where θ is the angle from A to B. Hence we have

$$\sum_{\mathbf{z}'\in\{0,1\}^n}\left|\sum_{\mathbf{z}\in\{0,1\}^n} C_{\mathbf{z}}(-1)^{\mathbf{z}\cdot\mathbf{z}'}\right|^2 = \sum_{\substack{\mathbf{z}'\in\{0,1\}^n\\ \text{s.t. } z'_n=0}}\left|\sum_{\mathbf{z}\in\{0,1\}^n} C_{\mathbf{z}}(-1)^{\mathbf{z}\cdot\mathbf{z}'}\right|^2 + \sum_{\substack{\mathbf{z}'\in\{0,1\}^n\\ \text{s.t. } z'_n=1}}\left|\sum_{\mathbf{z}\in\{0,1\}^n} C_{\mathbf{z}}(-1)^{\mathbf{z}\cdot\mathbf{z}'}\right|^2$$
$$= \sum_{\mathbf{y}'\in\{0,1\}^{n-1}}\left|\sum_{\mathbf{y}\in\{0,1\}^{n-1}} C_{\mathbf{y}0}(-1)^{\mathbf{y}\cdot\mathbf{y}'} + \sum_{\mathbf{y}\in\{0,1\}^{n-1}} C_{\mathbf{y}1}(-1)^{\mathbf{y}\cdot\mathbf{y}'}\right|^2 +$$
$$\sum_{\mathbf{y}'\in\{0,1\}^{n-1}}\left|\sum_{\mathbf{y}\in\{0,1\}^{n-1}} C_{\mathbf{y}0}(-1)^{\mathbf{y}\cdot\mathbf{y}'} - \sum_{\mathbf{y}\in\{0,1\}^{n-1}} C_{\mathbf{y}1}(-1)^{\mathbf{y}\cdot\mathbf{y}'}\right|^2$$
$$= 2\sum_{\mathbf{y}'\in\{0,1\}^{n-1}}\left(\left|\sum_{\mathbf{y}\in\{0,1\}^{n-1}} C_{\mathbf{y}0}(-1)^{\mathbf{y}\cdot\mathbf{y}'}\right|^2 + \left|\sum_{\mathbf{y}\in\{0,1\}^{n-1}} C_{\mathbf{y}1}(-1)^{\mathbf{y}\cdot\mathbf{y}'}\right|^2\right),$$

where in the last line we apply (10.20). The claim holds by induction. □

We then have the following reduction.

Lemma 10.30. *Let* $K > 1$ *and* $\theta \in [0, 2\pi)$*. Then*

$$\text{FACTOR-K-STRONGSIMIQP}_{1,2}(\theta) \leqslant_T \text{FACTOR-K}^2\text{-NORMIQPISING}(2\theta).$$

Proof. If all qubits in the input to FACTOR-K-STRONGSIMIQP$_{1,2}(\theta)$ are measured, then the result follows from Lemma 10.28. Otherwise, without loss of generality we assume the first $n-s$ qubits are measured. Let C, $I = [n-s]$ and $\mathbf{y}' \in \{0,1\}^{n-s}$ be the input to FACTOR-K-STRONGSIMIQP$_{1,2}(\theta)$. We use (10.12), (10.18), and the first line of (10.19):

$$\Pr_{C;I}(\mathbf{Y}' = \mathbf{y}') = \sum_{\mathbf{z}'\in\{0,1\}^s} \Pr_C(\mathbf{Y} = \mathbf{y}'\mathbf{z}')$$
$$= 2^{-2n}\sum_{\mathbf{z}'\in\{0,1\}^s}\left|\sum_{\sigma:V\to\{0,1\}}\left(e^{i2\theta}\right)^{m(\sigma)}\left(\prod_{l=n-s+1}^{n}(-1)^{z'_{l-(n-s)}\sigma(v_l)}\left(e^{-i2\theta}\right)^{p_l\sigma(v_l)}\right)\right.$$
$$\left.\left(\prod_{k=1}^{n-s}(-1)^{y'_k\sigma(v_k)}\left(e^{-i2\theta}\right)^{p_k\sigma(v_k)}\right)\right|^2$$
$$= 2^{-2n}\sum_{\mathbf{z}'\in\{0,1\}^s}\left|\sum_{\mathbf{z}\in\{0,1\}^s} Q_{\mathbf{z}}(-1)^{\mathbf{z}\cdot\mathbf{z}'}\right|^2, \tag{10.21}$$

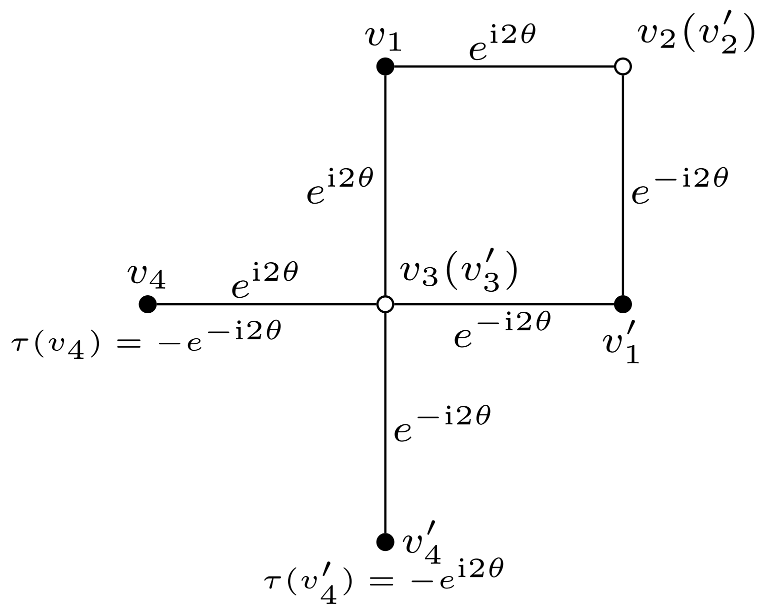


Figure 10.6: The equivalent Ising instance to the circuit in Figure 10.4, if qubits 2 and 3 are unmeasured.

where for $\mathbf{z} \in \{0,1\}^s$, $Q_{\mathbf{z}}$ is the contribution of assigning z_{l-n+s} to v_l without the possible -1 external field, that is,

$$Q_{\mathbf{z}} = \prod_{l=n-s+1}^{n} \left(e^{-i2\theta}\right)^{z_{l-n+s}p_l} \sum_{\substack{\sigma:V\to\{0,1\} \text{ such that} \\ \text{for } n-s+1\leqslant l\leqslant n, \sigma(v_l)=z_{l-n+s}}} \left(e^{i2\theta}\right)^{m(\sigma)} \cdot \prod_{k=1}^{n-s} (-1)^{y_k'\sigma(v_k)} \left(e^{-i2\theta}\right)^{p_k\sigma(v_k)}.$$

Apply Claim 10.29 on (10.21):

$$\Pr_{C;I}(\mathbf{Y}' = \mathbf{y}') = 2^{-2n+s} \sum_{\mathbf{z}\in\{0,1\}^s} |Q_{\mathbf{z}}|^2 . \tag{10.22}$$

Moreover we have

$$|Q_{\mathbf{z}}|^2 = \left| \sum_{\substack{\sigma:V\to\{0,1\} \text{ such that} \\ \text{for } n-s+1\leqslant l\leqslant n, \sigma(v_l)=z_{l-n+s}}} \left(e^{i2\theta}\right)^{m(\sigma)} \prod_{k=1}^{n-s} (-1)^{y_k'\sigma(v_k)} \left(e^{-i2\theta}\right)^{p_k\sigma(v_k)} \right|^2 .$$

We construct the following instance of FACTOR-K^2-NORMIQPISING(2θ). We first construct

$G_C = (V, E)$ with edge interaction $e^{i2\theta}$ as before. The vertex set $\{v_j\}$ contains one vertex for each of the n qubits. For each gate R_θ on two qubits j, k we add edge (j, k) with edge interaction $e^{i2\theta}$ to G_C. Now make a copy $G'_C = (V', E')$ such that the edge interaction is $\overline{e^{i2\theta}} = e^{-i2\theta}$. Let $\varphi_{C;I}$ be this edge interaction function. Then we identify vertices v_l with v'_l for all $n - s + 1 \leqslant l \leqslant n$. Let U be the set of these identified vertices and let $V_1 = V - U$ and $V'_1 = V' - U$. The external field $\tau = \tau_{C;I,\mathbf{y}'}$ is defined as follows: for any $v \in U$, $\tau(v) = 1$; for any $v_j \in V_1$, $\tau(v_j) = e^{-i(2p_j\theta)}(-1)^{y'_j}$; and for any $v'_j \in V'_1$, $\tau(v'_j) = \overline{\tau(v_j)} = e^{i(2p_j\theta)}(-1)^{y'_j}$. Informally, this instance was formed by putting G_C and its complement together and identifying vertices that correspond to unmeasured qubits. Note that if two vertices in U are connected by an edge, then they are actually connected by two edges, and the product of the two edge interactions is 1. We therefore remove all edges with both endpoints in U. Call the resulting graph H_C. One can verify that $(H_C, \varphi_{C;I}, \tau_{C;I,\mathbf{y}'})$ is a valid instance of FACTOR-K^2-NORMIQPISING(2θ). An example of the construction is given in Figure 10.6.

Fix an assignment $\mathbf{z} \in \{0,1\}^s$ on U. The contribution $Z_\mathbf{z}$ to $Z_{\text{Ising}}(H_C; \varphi_{C;I}, \tau_{C;I,\mathbf{y}'})$ can be counted in two independent parts, V and V'. Hence we have

$$\begin{aligned} Z_\mathbf{z} &= \left(\sum_{\sigma_1 : V_1 \to \{0,1\}} \left(e^{i2\theta}\right)^{m_*(\sigma_1, \mathbf{z})} \prod_{j=1}^{n-s} \tau(v_j)^{\sigma(v_j)} \right) \cdot \left(\sum_{\sigma'_1 : V'_1 \to \{0,1\}} \left(e^{-i2\theta}\right)^{m'_*(\sigma'_1, \mathbf{z})} \prod_{j=1}^{n-s} \overline{\tau(v_j)}^{\sigma(v'_j)} \right) \\ &= \left| \sum_{\sigma_1 : V_1 \to \{0,1\}} \left(e^{i2\theta}\right)^{m_*(\sigma_1, \mathbf{z})} \prod_{j=1}^{n-s} \tau(v_j)^{\sigma(v_j)} \right|^2, \end{aligned}$$

where given the configurations σ_1 (or σ'_1), $m_*(\sigma_1, \mathbf{z})$ (or $m'_*(\sigma'_1, \mathbf{z})$) is the number of monochromatic edges with at least one endpoint in V (or V'). Recall that $\tau(v_j) = e^{-i(2p_j\theta)}(-1)^{y'_j}$. Comparing $Z_\mathbf{z}$ to $|Q_\mathbf{z}|^2$, the only difference is that in $|Q_\mathbf{z}|^2$, $e^{i2\theta}$ is raised to the number of monochromatic edges in the whole V instead of V_1. However for any monochromatic edge in U, its contribution is independent from the configuration σ, and hence can be moved outside of the sum. All such terms are cancelled after taking the norm. This implies $Z_\mathbf{z} = |Q_\mathbf{z}|^2$. Therefore (10.22) can be rewritten as

$$\begin{aligned}\Pr_{C;I}(\mathbf{Y}'=\mathbf{y}') &= 2^{-2n+s}\sum_{\mathbf{z}\in\{0,1\}^s} Z_{\mathbf{z}}\\ &= 2^{-2n+s} Z_{\text{Ising}}(H_C;\varphi_{C;I},\tau_{C;I,\mathbf{y}'}) = 2^{-2n+s}|Z_{\text{Ising}}(H_C;\varphi_{C;I},\tau_{C;I,\mathbf{y}'})|. \qquad (10.23)\end{aligned}$$

The lemma follows from the above equation. □

Remark In fact, the construction of H_C can be further simplified. If $v \in V$ and $v' \in V'$ connect to some $u \in U$, we can replace edges (u, v) and (u, v') by a new edge (v, v') with an Ising interaction $\frac{2}{e^{i4\theta}+e^{-i4\theta}}$. (In case $e^{i4\theta} + e^{-i4\theta} = 0$ this interaction is equality and we identify v with v'.) Therefore we can reduce an instance of FACTOR-K-STRONGSIMIQP$_{1,2}(\theta)$ to an Ising model of size linear in $|I|$, the number of measured qubits. If $|I| = O(\log n)$, then the reduced Ising instance is tractable and so is the simulation. This matches known strong simulation results (see [BJS11, Theorem 3.4], the remark following that theorem and also [She10].)

The reduction also works in the other direction when $e^{i\theta}$ is a root of unity.

Theorem 10.31. *Let* $e^{i\theta}$ *be a root of unity and let* $K > 1$*. Then*

$$\text{FACTOR-K-NORMIQPISING}(2\theta) \ \equiv_T \ \text{FACTOR-}K^{1/2}\text{-STRONGSIMIQP}_{1,2}(\theta).$$

Proof. Lemma 10.30 implies a reduction from the right hand side to the left hand side. In the rest of the proof we show the other direction. As $e^{i\theta}$ is a root of unity, there exists a positive integer t such that $e^{-i2\theta} = e^{i2t\theta}$. Given an instance (G, φ, τ) of FACTOR-K-NORMIQPISING(2θ), we may replace each edge of interaction $e^{-i2\theta}$ by t parallel edges of weight $e^{i2\theta}$. Moreover, we may assume the external field is of the form $\tau(v_j) = (-1)^{a_j}\left(e^{-i2\theta}\right)^{b_j}$ for the same reason.

We construct an **IQP**$_{1,2}(\theta)$ circuit C on $n = |V|$ qubits. For each edge $(v_j, v_k) \in E$, we add a quantum gate R_θ on qubits j and k. For each $1 \leqslant j \leqslant n$, we add b_j many quantum gate P_θ on qubits j and let the output $y_j = 1$ on qubit j if a_j is odd. By Lemma 10.28 we see that $2^{2n}\Pr_C(\mathbf{Y} = \mathbf{y}) = \left|Z_{\text{Ising}}(G; e^{i2\theta}, \tau)\right|^2$. □

Suppose the Ising instance in the proof of Theorem 10.31 has no external field and has a constant edge interaction $e^{i2\theta}$. Then it is not hard to see that above construction does not rely on $e^{i\theta}$ being a root of unity and works for general θ. Hence we have the following lemma.

Lemma 10.32. *Let* $e^{i\theta} \in \mathbb{C}$ *and* $K > 1$*. Then*

$$\textsc{Factor-K-NormIsing}(e^{i\theta}) \leqslant_T \textsc{Factor-K}^{1/2}\textsc{-StrongSimIQP}_{1,2}(\theta/2).$$

It is easy to see that Theorem 10.27 follows from Lemma 10.32 and Corollary 10.17.

In a related result, Bremner et al. [BJS11, Corollary 3.3] showed that weakly simulating **IQP** with multiplicative error implies that the polynomial hierarchy collapses to the third level. More precisely, their result is the following. Suppose C is an $\mathbf{IQP}_{1,2}(\pi/8)$ circuit on n qubits. If there exists a classical randomized polynomial time procedure to sample a binary string $\mathbf{Z}$ of length n, such that for every string $\mathbf{y} \in \{0,1\}^n$ and any constant $1 \leqslant K < \sqrt{2}$,

$$\Pr_C(\mathbf{Y} = \mathbf{y})/K \leqslant \Pr(\mathbf{Z} = \mathbf{y}) \leqslant K \Pr_C(\mathbf{Y} = \mathbf{y}),$$

then the polynomial hierarchy collapses to the third level. The usual measure for determining the quality of a sampling procedure is total variation distance, which is weaker than "multiplicative error". So the result in [BJS11] does not rule out weak simulation with small variation distance. To see this, note that, if the multiplicative error is K, then obviously the total variation distance is at most $K - 1$. On the other hand, consider two distributions supported by two n-bit Boolean strings. A sample from the first distribution is obtained uniformly choosing each of the n bits. A sample from the second distribution is obtained by uniformly choosing each of the first $n - 1$ bits. The last bit is 1 if all other bits are 0, and is chosen uniformly otherwise. The total variation distance is 2^{-n}, but the multiplicative error is infinity at the all 0 string. Note that the complexity implication "polynomial hierarchy collapses to the third level" is apparently weaker than the consequence of strong simulation from Theorem 10.27, which is $\mathbf{FP} = \#\mathbf{P}$.

Strong simulation is also studied with respect to other classes of quantum circuits, see for example [JV14]. The allowable error is usually taken to be additive and exponentially small, instead of the constant factors that we have studied here. For example, [JV14] requires that the output be computed with k bits of precision in an amount of time that is polynomial in both k and the size of the input. Additive error is quite different from multiplicative error. Also, the amount of accuracy is important. Lemma 10.8 shows that there is no difference between a constant factor and an FPRAS scenario, in which the error is allowed to be a factor of $1 \pm 1/R$ for a unary input R. On the other hand, achieving a multiplicative error of $1 \pm 1/\exp(R)$ is an entirely different matter.

10.7 Complex Ising with External Fields

At last, we turn our attention to Ising models with an external field $\lambda \neq 1$. To obtain our hardness results, we need an (exponential) lower bound on the relevant partition functions, which will be developed in the next section. We will use some standard techniques from Diophantine approximation. After that, we will extend our hardness results.

Lower Bounds on Partition Functions

Suppose we have two edge weights y and y' that are close. It is easy to bound the distance between $Z_{\text{Ising}}(G; y)$ and $Z_{\text{Ising}}(G; y')$ additively, but not multiplicatively. To convert an absolute error into a relative error, one needs some lower bound on the partition function. However, when the edge interaction y is negative or complex, it is possible that the partition function vanishes. Assuming that it doesn't vanish, we would like to know how close to zero could it get. When y is rational, an exponential lower bound is easy to obtain by a simple granularity argument, but the argument is more difficult when y is not rational. In this section we give an exponential lower bound which is valid when y is an algebraic number. The techniques that we use are standard in transcendental number theory, see e.g. [Bug04].

We begin with some basic definitions from [Bug04]. For a polynomial with complex coefficients

$$P(x) = \sum_{i=0}^{n} a_i x^i = a_n \prod_{i=1}^{n} (x - \alpha_i),$$

the (naive) *height* of $P(x)$ is defined as $H(P) := \max_i\{|a_i|\}$. A more advanced tool, its *Mahler measure*, is defined as

$$M(P) := |a_n| \prod_{i=1}^{n} \max\{1, |\alpha_i|\}.$$

There is a standard inequality relating these two measures. It is proved for complex polynomials in [Bug04, Lemma A.2]. For completeness, we include the proof (following [Bug04]) for the case in which $P(x)$ is a real polynomial, which is all that we require.

Lemma 10.33. *Let* $P(x)$ *be a non-zero real polynomial of degree* n*. Then* $M(P) \leqslant \sqrt{n+1}\ H(P)$*.*

Proof. First apply Jensen's formula on $P(x)$ and on the unit circle in the complex plane,

$$M(P) = \exp\left\{\int_0^1 \log|P(e^{2i\pi t})|dt\right\}.$$

The convexity of exponential functions implies

$$M(P) \leqslant \int_0^1 |P(e^{2i\pi t})|dt \leqslant \left(\int_0^1 |P(e^{2i\pi t})|^2 dt\right)^{1/2},$$

where the second inequality follows by the Cauchy-Schwarz inequality writing $P(x)$ as $f(x)g(x)$ where $g(x) = 1$. The inner integral yields

$$\begin{aligned}\int_0^1 |P(e^{2i\pi t})|^2 dt &= \int_0^1 \left(\left(\sum_{j=0}^{n} a_j \cos(j \cdot 2\pi t)\right)^2 + \left(\sum_{j=0}^{n} a_j \sin(j \cdot 2\pi t)\right)^2\right) dt \\ &= \sum_{i=0}^{n} a_i^2 + 2\int_0^1 \sum_{0\leqslant j<k\leqslant n} a_j a_k (\cos(j\cdot 2\pi t)\cos(k\cdot 2\pi t) + \sin(j\cdot 2\pi t)\sin(k\cdot 2\pi t))dt \\ &= \sum_{i=0}^{n} a_i^2 + 2\sum_{0\leqslant j<k\leqslant n} a_j a_k \int_0^1 \cos((j-k)\cdot 2\pi t)dt = \sum_{i=0}^{n} a_i^2.\end{aligned}$$

The claim holds as $M(P) \leqslant \left(\sum_{i=0}^{n} a_i^2\right)^{1/2} \leqslant \sqrt{n+1}\ H(P)$. □

Let $y \in \mathbb{C}$ be an algebraic number and its minimal polynomial over $\mathbb{Z}$ is $P_y(x)$. The degree of $P_y(x)$ is called the *degree* of y and $H(P_y)$ is called the *height* of y, also denoted $H(y)$.

We also need the following notion of *resultants*.

Definition 10.34. *Let* $P(x) = a_n \prod_{i=1}^{n}(x - \alpha_i)$ *and* $Q(x) = b_m \prod_{i=1}^{m}(x - y_i)$ *be two non-constant polynomials. The resultant of* $P(x)$ *and* $Q(x)$ *is defined as*

$$\mathrm{Res}(P, Q) = a_n^m b_m^n \prod_{1 \leqslant i \leqslant n} \prod_{1 \leqslant i \leqslant m} (\alpha_i - y_j).$$

It is a standard result that $\mathrm{Res}(P, Q)$ is an integer polynomial in the coefficients of $P(x)$ and $Q(x)$. The resultant is also the determinant of the so-called Sylvester matrix. In particular, when $P(x)$ and $Q(x)$ are integer polynomials, $\mathrm{Res}(P, Q)$ is always an integer, as the Sylvester matrix is an integer matrix in this case. Moreover, we can rewrite the resultant as follows:

$$\mathrm{Res}(P, Q) = a_n^m \prod_{1 \leqslant i \leqslant n} Q(\alpha_i) = (-1)^{mn} b_m^n \prod_{1 \leqslant j \leqslant m} P(y_j).$$

Now we are ready to give a lower bound for any integer polynomial evaluated at an algebraic number. It is a standard result in algebraic number theory. For completeness we provide a proof here and the treatment is from [Bug04, Theorem A.1].

Lemma 10.35. *Let* $P(x)$ *be an integer polynomial of degree* n*, and* $y \in \mathbb{C}$ *be an algebraic number of degree* d*. Then either* $P(y) = 0$ *or*

$$|P(y)| \geqslant C_y^{-n} \left((n+1)\, H(P)\right)^{-d+1}.$$

where $C_y > 1$ *is an effectively computable constant that only depends on* y*.*

Proof. Assume $P(y) \neq 0$. Let $Q(x) = b_d \prod_{i=1}^{d}(x - y_i)$ be the minimal polynomial of y over $\mathbb{Z}$ with $y_1 = y$.

Suppose there is an $j \neq 1$ such that $P(y_j) = 0$. As $Q(x)$ is the minimal polynomial of y, none of y_j could be a rational number. Hence there is an automorphism of the splitting field of $Q(x)$ that maps y_j to y. Applying this automorphism on both sides of $P(y_j) = 0$, we get $P(y) = 0$. Contradiction!

Hence we have $P(y_i) \neq 0$ for all i and the resultant of $P(x)$ and $Q(x)$ is non-zero. Since

Res(P, Q) is an integer, we have

$$1 \leqslant |\operatorname{Res}(P, Q)| = |b_d|^n \prod_{1 \leqslant i \leqslant d} |P(y_i)|.$$

Clearly, by triangle inequality we have $|P(y_i)| \leqslant (n+1)\,H(P)(\max\{1, |y_i|\})^n$. It implies,

$$\begin{aligned} 1 &\leqslant |P(y)||b_d|^n \left((n+1)\,H(P)\right)^{d-1} \prod_{2 \leqslant i \leqslant d} (\max\{1, |y_i|\})^n \\ &= |P(y)| \left((n+1)\,H(P)\right)^{d-1} \left(\frac{M(Q)}{\max\{1, |y|\}}\right)^n \\ &\leqslant |P(y)| \left((n+1)\,H(P)\right)^{d-1} \left(\sqrt{d+1}\,H(y)\right)^n \end{aligned}$$

where the last inequality follows from Lemma 10.33. Therefore we have

$$|P(y)| \geqslant \left((n+1)\,H(P)\right)^{-d+1} \left(\sqrt{d+1}\,H(y)\right)^{-n}.$$

Let $C_y = \sqrt{d+1}\,H(y)$ and the lemma holds. □

Lemma 10.36. *Let* G *be a graph and* $y \in \mathbb{C}$ *a non-zero algebraic number of degree* d*. There exists a positive constant* $C > 1$ *depending only on* y *such that if* $Z_{Ising}(G; y) \neq 0$*, then* $|Z_{Ising}(G; y)| > C^{-m}$*, where* m *is the number of edges in* G*.*

Proof. Given a graph G, first suppose that G is not connected, G_i's are the components of G. Then $Z_{\text{Ising}}(G; y) = \prod_i Z_{\text{Ising}}(G_i; y)$. It is easy to see that if the claim holds for all components it hold for G as well. Therefore in the following we may assume G is connected. Then $m \geqslant n-1$ where n is the number of vertices.

We can rewrite $Z_{\text{Ising}}(G; y)$ as a polynomial in y as follows,

$$P(y) = Z_{\text{Ising}}(G; y) = \sum_{i=0}^{m} C_j y^j,$$

where C_j is the number of configurations such that there are exactly j many monochromatic edges. Notice that $\sum_{j=0}^{m} C_j = 2^n$, we have $H(P) \leqslant 2^n$. Assume $P(y) \neq 0$. Apply Lemma 10.35

and we obtain

$$\begin{aligned}|P(y)| &\geqslant C_y^{-m}\left((m+1)\,H(P)\right)^{-d+1} \\ &\geqslant (m+1)^{-d+1} C_y^{-m} 2^{-(d-1)n},\end{aligned}$$

where $C_y > 1$ is a constant depending only on y. As $m \geqslant n-1$, the right hand side decays exponentially in m and the lemma follows. □

Lemma 10.37. *Let G be a graph and $y, z \in \mathbb{C}$ two roots of unity. Let n be the number of vertices in G and m the number of edges. There exists a positive constant $C > 1$ depending only on y and z such that if $Z_{Ising}(G; y, z) \neq 0$, then $|Z_{Ising}(G; y, z)| > C^{-m}$.*

Proof. As in the previous lemma we may assume G is connected and $m \geqslant n-1$. Suppose y is of order d_1 and z order d_2. Let d be the least common multiple of d_1 and d_2. Then there exists a root of unity w of order d such that $y = w^{t_1}$ and $z = w^{t_2}$.

Given a graph G, we can rewrite $Z_{\text{Ising}}(G; y, z)$ as a polynomial in y and z as follows,

$$Z_{\text{Ising}}(G; y, z) = \sum_{k=0}^{n}\sum_{j=0}^{m} C_{j,k} y^j z^k,$$

where $C_{j,k}$ is the number of configurations such that there are exactly j many monochromatic edges and k many 1 vertices. Let

$$P(w) = Z_{\text{Ising}}(G; y, z) = \sum_{k=0}^{n}\sum_{j=0}^{m} C_{j,k} w^{t_1 j + t_2 k} = \sum_{\ell=0}^{t_1 m + t_2 n} C'_\ell w^\ell,$$

where $C'_\ell = \sum_{t_1 j + t_2 k = \ell} C_{j,k}$. Notice that $\sum_{\ell=0}^{t_1 m + t_2 n} C'_\ell = \sum_{k=0}^{n}\sum_{j=0}^{m} C_{j,k} = 2^n$, we have $H(P) \leqslant 2^n$. Assume $P(w) \neq 0$. Apply Lemma 10.35 and we obtain

$$\begin{aligned}|P(w)| &\geqslant C_w^{-t_1 m - t_2 n}\left((t_1 m + t_2 n + 1)\,H(P)\right)^{-d+1} \\ &\geqslant (t_1 m + t_2 n + 1)^{-d+1} C_w^{-t_1 m - t_2 n} 2^{-(d-1)n},\end{aligned}$$

where $C_w > 1$ is a constant depending only on w. As $m \geqslant n-1$, the right hand side decays exponentially in m and the lemma follows. □

Hardness Results

In this section we will show hardness results when both the edge interaction and external field are roots of unity. We first consider the external field -1.

Lemma 10.38. *Let* $K > 1$ *and* $y \in \mathbb{C}$ *be an algebraic complex number such that* $y \neq \pm 1$*. Then we have* FACTOR-K-NONZERO-NORMISING$(y) \leqslant_T$ FACTOR-K-NONZERO-NORMISING$(y, -1)$.

Proof. We first argue that a binary equality can be implemented. Consider a 2-stretch with the edge interaction y and external field -1. It is easy to calculate that the interaction matrix is $\begin{bmatrix} y^2-1 & 0 \\ 0 & 1-y^2 \end{bmatrix}$. Then do a 2-thickening. The resulting matrix is $\begin{bmatrix} (y^2-1)^2 & 0 \\ 0 & (1-y^2)^2 \end{bmatrix}$. Up to a constant of $(y^2-1)^2$ this is equality.

Suppose $G = (V, E)$ is an input to FACTOR-K-NONZERO-NORMISING(y). We introduce a new vertex v' for every vertex $v \in V$. Connect v and v' via this equality gadget, that is, first a 2-stretch and then a 2-thickening. Hence the external field on v is cancelled with this construction. The reduction follows. □

Next we consider the case when an real edge interaction can be implemented. If the norm of the interaction is less than 1, then we can cancel out the external field.

Lemma 10.39. *Let* $K > 1$ *and* $K' > 1$*. Let* y *and* z *be two roots of unity and* $z \neq \pm 1$*. Suppose some real number* $w \in (-1, 1)$ *as an edge interaction is implementable for the Ising model with edge interaction* y *and external field* z*. Then we have* FACTOR-K-NONZERO-NORMISING$(y) \leqslant_T$ FACTOR-(KK')-NONZERO-NORMISING(y, z).

Proof. Let $G = (V, E)$ be an input to FACTOR-K-NONZERO-NORMISING(y). Assume $Z_{\text{Ising}}(G; y) \neq 0$ as otherwise we are done. Suppose $|V| = n$, $|E| = m$, and $V = \{v_i | 1 \leqslant i \leqslant n\}$.

Suppose $w = 0$, which means we can implement inequality. For each vertex v_i, we introduce a new vertex v_i' and connect v_i and v_i' by the inequality. It is easy to verify that if v_i is assigned 0, the weight from v_i and v_i' together is z; when v_i is assigned 1, the weight is also z. Hence the external field is effectively cancelled and the reduction follows.

Otherwise assume $w \neq 0$, that is $w \in (-1, 0) \cup (0, 1)$. For each vertex v_i, we introduce a new vertex v_i', and add $2t$ many new edges between v_i and v_i', where t is a positive integer which we will choose later. By assumption we can implement the edge interaction w and we put it on all new edges. Let $V' = \{v_i' | 1 \leqslant i \leqslant n\}$ and we get a new graph $G' = (V \cup V', E')$.

For each vertex v_i, the contribution of v_i and v_i' combined is $w^{2t} + z$ when v_i is assigned 0 and $z(1+w^{2t}z)$ when v_i is assigned 1. Let $\lambda = \frac{z(1+w^{2t}z)}{w^{2t}+z}$. Notice that $w^{2t}+z \neq 0$ as $|w| < 1 = |z|$. We have

$$Z_{\text{Ising}}(G'; y, z) = (w^{2t} + z)^n \sum_{\sigma: V \to \{0,1\}} y^{m(\sigma)} \lambda^{n_1(\sigma)},$$

where $m(\sigma)$ is the number of monochromatic edges in E under σ and $n_1(\sigma)$ is the number of vertices in V that are assigned 1.

Let $Z := \left| \frac{Z_{\text{Ising}}(G';y,z)}{(w^{2t}+z)^n} - Z_{\text{Ising}}(G; y) \right|$. We want to show that Z is exponentially small. Apply the triangle inequality:

$$\begin{aligned}
|Z| &= \left| \sum_{\sigma: V \to \{0,1\}} y^{m(\sigma)} (\lambda^{n_1(\sigma)} - 1) \right| \leqslant \sum_{\sigma: V \to \{0,1\}} \left| y^{m(\sigma)} (\lambda^{n_1(\sigma)} - 1) \right| \\
&= \sum_{\sigma: V \to \{0,1\}} \left| \lambda^{n_1(\sigma)} - 1 \right| = \sum_{j=0}^{n} \binom{n}{j} \left| \lambda^j - 1 \right|, \qquad (10.24)
\end{aligned}$$

where we used the fact that $|y| = 1$. Let $\alpha = \lambda - 1 = \frac{z(1+w^{2t}z)}{w^{2t}+z} - 1 = \frac{w^{2t}(z^2-1)}{w^{2t}+z}$. As $z^2 - 1 \neq 0$ and $w^{2t} + z \neq 0$, $|\alpha|$ is decreasing exponentially in t. We may pick a positive integer $t = O(\log n)$ such that $ne|\alpha| < 1$. Applying the triangle inequality again for each $0 \leqslant j \leqslant n$, we get

$$\begin{aligned}
|\lambda^j - 1| &= |\sum_{l=1}^{j} \binom{j}{l} \alpha^l| \leqslant \sum_{l=1}^{j} \binom{j}{l} |\alpha^l| \\
&= (|\alpha| + 1)^j - 1 \leqslant (|\alpha| + 1)^n - 1 \\
&= \sum_{l=1}^{n} \binom{n}{l} |\alpha|^l \leqslant \sum_{l=1}^{n} \left(\frac{ne|\alpha|}{l} \right)^l \\
&\leqslant n^2 e |\alpha|, \qquad (10.25)
\end{aligned}$$

as $\left(\frac{ne|\alpha|}{l}\right)^{l}$ is decreasing in l. Plugging (10.25) into (10.24) we have

$$|Z| \leqslant \sum_{j=0}^{n} \binom{n}{j} n^2 e|\alpha| = e2^n n^2 |\alpha|. \tag{10.26}$$

Since $Z_{\text{Ising}}(G; y) \neq 0$, by Lemma 10.36, there exists a constant $C_y > 1$ such that $|Z_{\text{Ising}}(G; y)| > C_y^{-|E|}$. Since $|\alpha|$ is decreasing exponentially in t, by (10.26), we may pick an integer t that is polynomial in n (and sufficiently large with respect to K') such that

$$|Z| < \frac{K'-1}{K'} C_y^{-|E|} < \frac{K'-1}{K'} |Z_{\text{Ising}}(G; y)|. \tag{10.27}$$

By the definition of $|Z|$ and again the triangle inequality we get

$$\frac{1}{K'} = 1 - \frac{K'-1}{K'} \leqslant \frac{|Z_{\text{Ising}}(G'; y, z)|}{|w^{2t} + z|^n |Z_{\text{Ising}}(G; y)|} \leqslant 1 + \frac{K'-1}{K'} \leqslant K'.$$

This finishes the proof. □

A similar proof works when the implementable real field has a larger than 1 norm. Basically when this is the case we may power the external field z. If z is a root of unity then we could power it to 1.

Lemma 10.40. *Let* $K > 1$ *and* $K' > 1$. *Let* y *and* z *be two roots of unity and* $z \neq \pm 1$. *Suppose some real number* $w \in (-\infty, -1) \cup (1, \infty)$ *as an edge interaction is implementable for the Ising model with edge interaction* y *and external field* z. *Then we have* Factor-K-Nonzero-NormIsing(y, z^r) $\leqslant_T$ Factor-(KK')-Nonzero-NormIsing(y, z) *for any positive integer* r.

Proof. Let $G = (V, E)$ be an input to Factor-K-Nonzero-NormIsing(y, z^r). Assume that $Z_{\text{Ising}}(G; y, z^r) \neq 0$ as otherwise we are done. Suppose $|V| = n$, $|E| = m$, and $V = \{v_i | 1 \leqslant i \leqslant n\}$.

For each vertex v_i, we introduce $r-1$ many new vertices $v_{i,j}$, and add $2t$ many new edges between v_i and each $v_{i,j}$, where $j \in [r-1]$ and t is a positive integer which we will choose later. By assumption we can implement the edge interaction w and we put it on all new edges. Let $V' = \{v_{i,j} | 1 \leqslant i \leqslant n, 1 \leqslant j \leqslant r-1\}$ and we get a new graph $G' = (V \cup V', E')$.

For each vertex v_i, the contribution of v_i and all $v_{i,j}$ combined is $\left(w^{2t}+z\right)^{r-1}$ when v_i is assigned 0 and $z\left(1+w^{2t}z\right)^{r-1}$ when v_i is assigned 1. Let $\lambda = \frac{z\left(1+w^{2t}z\right)^{r-1}}{\left(w^{2t}+z\right)^{r-1}}$. Notice that $w^{2t}+z \neq 0$ as $|w| > 1 = |z|$. We have

$$Z_{\text{Ising}}(G';y,z) = \left(w^{2t}+z\right)^{n(r-1)} \sum_{\sigma:V\to\{0,1\}} y^{m(\sigma)}\lambda^{n_1(\sigma)},$$

where $m(\sigma)$ is the number of monochromatic edges in E under σ and $n_1(\sigma)$ is the number of vertices in V that are assigned 1.

Let $Z := \left|\frac{Z_{\text{Ising}}(G';y,z)}{\left(w^{2t}+z\right)^{n(r-1)}} - Z_{\text{Ising}}(G;y,z^r)\right|$. As the previous proof we show that Z is exponentially small. Apply the triangle inequality:

$$\begin{aligned}|Z| &= \left|\sum_{\sigma:V\to\{0,1\}} y^{m(\sigma)}(\lambda^{n_1(\sigma)} - z^{rn_1(\sigma)})\right| \leqslant \sum_{\sigma:V\to\{0,1\}} \left|y^{m(\sigma)}(\lambda^{n_1(\sigma)} - z^{rn_1(\sigma)})\right| \\ &= \sum_{\sigma:V\to\{0,1\}} \left|\lambda^{n_1(\sigma)} - z^{rn_1(\sigma)}\right| = \sum_{j=0}^{n} \binom{n}{j} \left|\lambda^j - z^{rj}\right|, \end{aligned} \tag{10.28}$$

where we used the fact that $|y| = 1$. Let $\alpha = \lambda - z^r = \frac{z\left(1+w^{2t}z\right)^{r-1}}{\left(w^{2t}+z\right)^{r-1}} - z^r = z\left((z+\mu)^{r-1} - z^{r-1}\right)$, where $\mu = \frac{1+w^{2t}z}{w^{2t}+z} - z = \frac{1-z^2}{w^{2t}+z} \neq 0$. As $z^2 - 1 \neq 0$ and $|w| > 1$, $|\mu|$ decreases exponentially in t. Pick a large enough integer t so that $|\mu| < 1$. Hence $|\alpha| = |z||(z+\mu)^{r-1} - z^{r-1}| = |\sum_{j=1}^{r-1}\binom{r-1}{j}\mu^j z^{r-1-j}| \leqslant \sum_{j=1}^{r-1}\binom{r-1}{j}|\mu^j| < |\mu|2^{r-1}$ by the triangle inequality. As $|\mu|$ decreases exponentially in t, so does $|\alpha|$.

Notice that $|\lambda| = |z^r + \alpha| \leqslant |z|^r + |\alpha| = 1 + |\alpha|$. Pick t large so that $|\alpha| < 1$. Applying the triangle inequality again for each $0 \leqslant j \leqslant n$, we get

$$\begin{aligned}|\lambda^j - z^{rj}| &= |\lambda - z^r|\left|\sum_{l=0}^{j-1}\lambda^l z^{r(j-1-l)}\right| \leqslant |\alpha|\left(\sum_{l=0}^{j-1}\left|\lambda^l z^{r(j-1-l)}\right|\right) \\ &= |\alpha|\left(\sum_{l=0}^{j-1}|\lambda|^l\right) \leqslant |\alpha|\left(\sum_{l=0}^{j-1}(1+|\alpha|)^l\right) \\ &< |\alpha|\left(\sum_{l=0}^{j-1}2^l\right) < 2^j|\alpha| \leqslant 2^n|\alpha|, \end{aligned} \tag{10.29}$$

as $|z| = 1$. Plugging (10.29) into (10.28) we have

$$|Z| < \sum_{j=0}^{n} \binom{n}{j} 2^n |\alpha| = 4^n |\alpha|. \tag{10.30}$$

Since $Z_{\text{Ising}}(G; y, z^r) \neq 0$, by Lemma 10.37, there exists a constant $C_{y,z^r} > 1$ such that

$$|Z_{\text{Ising}}(G; y, z^r)| > C_{y,z^r}^{-|E|}.$$

Since $|\alpha|$ is decreasing exponentially in t, by (10.30), we may pick an integer t that is polynomial in n (and sufficiently large with respect to K') such that

$$|Z| < \frac{K'-1}{K'} C_{y,z^r}^{-|E|} < \frac{K'-1}{K'} |Z_{\text{Ising}}(G; y, z^r)|. \tag{10.31}$$

By the definition of $|Z|$ and again the triangle inequality we get

$$\frac{1}{K'} = 1 - \frac{K'-1}{K'} \leqslant \frac{|Z_{\text{Ising}}(G'; y, z)|}{|w^{2t} + z|^{n(r-1)} |Z_{\text{Ising}}(G; y, z^r)|} \leqslant 1 + \frac{K'-1}{K'} \leqslant K'.$$

This finishes the proof. □

We will show how to implement a real edge interaction in the next lemma. Unless the norm of the new interaction is 1, the hardness holds due to the previous two lemmas. The failure cases are indeed tractable.

Lemma 10.41. *Let* $K > 1$. *Let* y *and* z *be two roots of unity such that* $y \notin \{1, -1, i, -i\}$ *and* $z \notin \{1, -1\}$. *Then* FACTOR-K-NONZERO-NORMISING(y, z) *is* #**P**-*hard.*

Proof. Let $y = e^{i\theta}$ and $z = e^{i\varphi}$ and $\theta, \varphi \in [0, 2\pi)$. Then $\theta \notin \{0, \pi/2, \pi, 3\pi/2\}$ and $\varphi \notin \{0, \pi\}$.

Since y is a root of unity, there exists an integer power of y that equals y^{-1}. Hence we can implement y^{-1} by thickenings. Then we implement a real interaction $w(\theta, \varphi)$ by the following gadget. We replace every edge by two parallel edges: one is a 2-stretch with interaction y and the other is also a 2-stretch but with y^{-1}. Then we calculate the effective edge interaction. When both endpoints are assigned 0, the contribution is $(y^2+z)(1/y^2+z) = 1+z^2+z(y^2+1/y^2)$. When both endpoints are assigned 1, the contribution is $(y^2z+1)(z/y^2+1) = 1+z^2+z(y^2+1/y^2)$ as well. When one endpoint is assigned 0 and the other 1, the contribution is $y(1+z)\cdot(1+z)/y = (1+z)^2$. Hence effectively on this edge the interaction is of the Ising type and its weight is $w(\theta,\varphi) = \frac{1+z^2+z(y^2+1/y^2)}{(1+z)^2}$.

We claim $w(\theta,\varphi) \in \mathbb{R}$. This is because

$$\begin{aligned} w(\theta,\varphi) &= \frac{1+z^2+z(y^2+1/y^2)}{(1+z)^2} = 1 + \frac{z(y^2+1/y^2-2)}{(1+z)^2} \\ &= 1 + \frac{(y-1/y)^2}{z+1/z+2} = 1 + \frac{-4\sin^2\theta}{2\cos\varphi+2} \\ &= 1 - \frac{\sin^2\theta}{\cos^2\frac{\varphi}{2}}. \end{aligned}$$

Notice that $\cos\frac{\varphi}{2} \neq 0$ as $\varphi \neq 0, \pi$. If $|w| < 1$, then we are done by combining Lemma 10.39 and Corollary 10.17. Otherwise if $|w| > 1$, the lemma follows from Lemma 10.40 by powering z to 1, and Corollary 10.17.

The failure case is $|w(\theta,\varphi)| = 1$ and hence $\sin^2\theta = 2\cos^2\frac{\varphi}{2}$ or $\sin\theta = 0$. Notice that $\sin\theta = 0$ implies $y = \pm 1$ which contradicts to our assumption. It is easy to implement y^2, which has argument 2θ. We then repeat the construction. If $|w(2\theta,\varphi)| \neq 1$, then it is reduced to previous cases. Otherwise $|w(2\theta,\varphi)| = 1$, implying that $\sin^2 2\theta = 2\cos^2\frac{\varphi}{2} = \sin^2\theta$ or $\sin 2\theta = 0$. The latter case is impossible as $\theta \notin \{0, \pi/2, \pi, 3\pi/2\}$. Hence $\sin^2 2\theta = \sin^2\theta$. It is easy to solve that $\theta \in \{\pi/3, 2\pi/3, 4\pi/3, 5\pi/3\}$ as $\theta \neq 0, \pi$. Therefore $2\cos^2\frac{\varphi}{2} = \sin^2\theta = 3/4$. However $\cos^2\frac{\varphi}{2} = 3/8$ has no solution of φ that is a rational fraction of π, which is contradicting to z being a root of unity. This finishes the proof. □

Lemma 10.42. *Let* $K > 1$*. Let* $y = \pm i$ *and* z *be a root of unity that is not one of* $\{1, -1, i, -i\}$*. Then* FACTOR-K-NONZERO-NORMISING(y, z) *is #**P**-hard.*

Proof. Let $y = e^{i\theta}$ and $z = e^{i\varphi}$ where $\theta, \varphi \in [0, 2\pi)$. As $y = \pm i$, we have $\theta \in \{\pi/2, 3\pi/2\}$ and $z \notin \{1, -1, i, -i\}$ implies $\varphi \notin \{0, \pi/2, \pi, 3\pi/2\}$. We use the same $w(\theta,\varphi) \in \mathbb{R}$ construction as in the proof of Lemma 10.41. If $|w(\theta,\varphi)| = 0$ then $\cos^2\frac{\varphi}{2} = 1$. This implies $\varphi/2 \in \{0, \pi\}$ contradicting to $\varphi \notin \{0, \pi/2, \pi, 3\pi/2\}$. If $|w(\theta,\varphi)| = 1$ then $\cos^2\frac{\varphi}{2} = 1/2$. This implies $\varphi/2 \in \{\pi/4, 3\pi/4, 5\pi/4, 7\pi/4\}$ also contradicting to $\varphi \notin \{0, \pi/2, \pi, 3\pi/2\}$. Hence we can implement a real edge interaction $w(\theta,\varphi)$ such that $|w(\theta,\varphi)| \neq 0, 1$.

Notice that $w(\theta,\varphi) = 1 - \frac{\sin^2\theta}{\cos^2\frac{\varphi}{2}} = 1 - 1/\cos^2\frac{\varphi}{2} < 0$. If $w(\theta,\varphi) \in (-1,0)$, then we adopt the construction in the proof of Lemma 10.39 to cancel the external field of z. Hence we can reduce Factor-K-Nonzero-NormIsing$(w(\theta,\varphi))$ to Factor-(KK')-Nonzero-NormIsing(y,z) for any constant $K' > 1$. The #**P**-hardness follows from Corollary 10.14.

Otherwise $w(\theta,\varphi) \in (-\infty,-1)$, then we use Lemma 10.40 to power up the external field of z. Instead of powering z to 1, we would like to pick a positive integer r such that $w(\theta, r\varphi) \in (-1,0)$, which reduces to the previous case. This is equivalent to $\frac{1}{2} < \cos^2\frac{r\varphi}{2} < 1$, which, in turn, is equivalent to $r\varphi \in (0,\pi/2) \cup (3/2\pi, 2\pi)$ modulo 2π. Suppose $\varphi = \frac{2a\pi}{b}$ where a, b are two co-prime positive integers and $b = 3$ or $b \geqslant 5$ since $z \notin \{1,-1,i,-i\}$. Assume $b \geqslant 5$ first. As a, b are co-prime, there exist two integers l_1 and l_2 such that $l_1 a + l_2 b = 1$ and $l_1 > 0$. Let $r = l_1$ and we have $r\varphi/2 = \frac{2al_1\pi}{b} = \frac{2\pi}{b} - 2l_2\pi$. This choice of r meets the requirement since $\frac{2\pi}{b} \in (0,\pi/2)$.

The case left is when $b = 3$, in which case $\varphi \in \{2\pi/3, 4\pi/3\}$. We reduce Factor-K-Nonzero-NormIsing$(y,-z)$ to Factor-K-Nonzero-NormIsing(y,z). This suffices due to $\arg(-z) = \varphi + \pi$, which is one of the previous cases.

Suppose $G = (V,E)$ is an input to Factor-K-Nonzero-NormIsing$(y,-z)$. Introduce a new vertex v' for each vertex $v \in V$. Since $y = \pm i$, there exists a positive integer t such that $y^t = -1$. Connect v and v' by t many new edges. We can calculate that the effective field of v in the new graph (with respect to interaction y and field z) is $\frac{z-z^2}{z-1} = -z$. This finishes our proof. $\square$

We can now prove Theorem 10.5.

Proof of Theorem 10.5. If $y = \pm 1$, then we can replace every edge interaction by two unary constraints. Hence the problem is tractable for any external field. Consider next the case where $y = \pm i$. If $z \in \{1,-1,i,-i\}$, the algorithm is from [CLX14]. Otherwise, the hardness is from Lemma 10.42. Finally, for the rest of the proof, we consider the case where $y \notin \{1,-1,i,-i\}$. For $z = 1$, the hardness follows from Corollary 10.17. For $z = -1$, the hardness is obtained by combining Lemma 10.38 and Corollary 10.17. Otherwise $z \notin \{1,-1\}$, and the hardness follows from Lemma 10.41. $\square$

Bibliography

[Bax82] Rodney J. Baxter. *Exactly solved models in statistical mechanics*. Academic press London, 1982.

[BD07] Andrei A. Bulatov and Víctor Dalmau. Towards a dichotomy theorem for the counting constraint satisfaction problem. *Information and Computation*, 205(5):651–678, 2007.

[BDG+09] Andrei A. Bulatov, Martin E. Dyer, Leslie Ann Goldberg, Markus Jalsenius, and David Richerby. The complexity of weighted Boolean #CSP with mixed signs. *Theor. Comput. Sci.*, 410(38-40):3949–3961, 2009.

[BDG+13] Andrei A. Bulatov, Martin E. Dyer, Leslie Ann Goldberg, Mark Jerrum, and Colin McQuillan. The expressibility of functions on the boolean domain, with applications to counting CSPs. *J. ACM*, 60(5):32, 2013.

[BG05] Andrei Bulatov and Martin Grohe. The complexity of partition functions. *Theor. Comput. Sci.*, 348(2):148–186, 2005.

[BG08] Antar Bandyopadhyay and David Gamarnik. Counting without sampling: Asymptotics of the log-partition function for certain statistical physics models. *Random Struct. Algorithms*, 33(4):452–479, 2008.

[BGK+07] Mohsen Bayati, David Gamarnik, Dimitriy A. Katz, Chandra Nair, and Prasad Tetali. Simple deterministic approximation algorithms for counting matchings. In *STOC*, pages 122–127, 2007.

[BJS11] Michael J. Bremner, Richard Jozsa, and Dan J. Shepherd. Classical simulation of commuting quantum computations implies collapse of the polynomial hierarchy. *Proc. R. Soc. A*, 467(2126):459–472, 2011.

[BO92] Thomas Brylawski and James Oxley. The Tutte polynomial and its applications. In Neil White, editor, *Matriod Applications*, pages 123–225. Cambridge University Press, 1992.

[BR91] Richard A. Brualdi and Herbert J. Ryser. *Combinatorial Matrix Theory*. Cambridge University Press, 1991.

[Bug04] Yann Bugeaud. *Approximation by Algebraic Numbers*. Cambridge Tracts in Mathematics. Cambridge University Press, 2004.

[Cay49] Arthur Cayley. Sur les déterminates gauches. *Journal für die reine und angewandte Mathematik*, 38:93–96, 1849.

[CC07] Jin-Yi Cai and Vinay Choudhary. Some results on matchgates and holographic algorithms. *Int. J. Software and Informatics*, 1(1):3–36, 2007.

[CCGL12] Jin-Yi Cai, Xi Chen, Heng Guo, and Pinyan Lu. Inapproximability after uniqueness phase transition in two-spin systems. In *COCOA*, pages 336–347, 2012.

[CCL09] Jin-Yi Cai, Vinay Choudhary, and Pinyan Lu. On the theory of matchgate computations. *Theory Comput. Syst.*, 45(1):108–132, 2009.

[CCL13] Jin-Yi Cai, Xi Chen, and Pinyan Lu. Graph homomorphisms with complex values: A dichotomy theorem. *SIAM J. Comput.*, 42(3):924–1029, 2013.

[CCLL10] Jin-Yi Cai, Xi Chen, Richard J. Lipton, and Pinyan Lu. On tractable exponential sums. In *FAW*, pages 148–159. Springer Berlin Heidelberg, 2010.

[CDG+15] Xi Chen, Martin E. Dyer, Leslie Ann Goldberg, Mark Jerrum, Pinyan Lu, Colin McQuillan, and David Richerby. The complexity of approximating conservative counting CSPs. *J. Comput. Syst. Sci.*, 81(1):311–329, 2015.

[CFGW15] Jin-Yi Cai, Zhiguo Fu, Heng Guo, and Tyson Williams. A Holant dichotomy: Is the FKT algorithm universal? In *FOCS*, 2015. To appear. *CoRR*, abs/1505.02993.

[CG14] Jin-Yi Cai and Aaron Gorenstein. Matchgates revisited. *Theory Comput.*, 10(7):167–197, 2014.

[CGG+14] Jin-Yi Cai, Andreas Galanis, Leslie Ann Goldberg, Heng Guo, Mark Jerrum, Daniel Štefankovič, and Eric Vigoda. #BIS-hardness for 2-spin systems on bipartite bounded degree graphs in the tree non-uniqueness region. In *RANDOM*, pages 582–595, 2014.

[CGW13] Jin-Yi Cai, Heng Guo, and Tyson Williams. A complete dichotomy rises from the capture of vanishing signatures (extended abstract). In *STOC*, pages 635–644. ACM, 2013. *CoRR*, abs/1204.6445.

[CGW14] Jin-Yi Cai, Heng Guo, and Tyson Williams. Holographic algorithms beyond matchgates. In *ICALP*, pages 271–282. Springer Berlin Heidelberg, 2014. *CoRR*, abs/1307.7430.

[CHL12] Jin-Yi Cai, Sangxia Huang, and Pinyan Lu. From Holant to #CSP and back: Dichotomy for Holant^c problems. *Algorithmica*, 64(3):511–533, 2012.

[CK12] Jin-Yi Cai and Michael Kowalczyk. Spin systems on k-regular graphs with complex edge functions. *Theor. Comput. Sci.*, 461:2–16, 2012.

[CK13] Jin-Yi Cai and Michael Kowalczyk. Partition functions on k-regular graphs with $\{0, 1\}$-vertex assignments and real edge functions. *Theor. Comput. Sci.*, 494:63–74, 2013.

[CKW12] Jin-Yi Cai, Michael Kowalczyk, and Tyson Williams. Gadgets and anti-gadgets leading to a complexity dichotomy. In *ITCS*, pages 452–467. ACM, 2012.

[CL10] Jin-Yi Cai and Pinyan Lu. On symmetric signatures in holographic algorithms. *Theory Comput. Syst.*, 46(3):398–415, 2010.

[CL11a] Jin-Yi Cai and Pinyan Lu. Holographic algorithms: From art to science. *J. Comput. Syst. Sci.*, 77(1):41–61, 2011.

[CL11b] Jin-Yi Cai and Pinyan Lu. Signature theory in holographic algorithms. *Algorithmica*, 61(4):779–816, 2011.

[CLX10] Jin-Yi Cai, Pinyan Lu, and Mingji Xia. Holographic algorithms with matchgates capture precisely tractable planar #CSP. In *FOCS*, pages 427–436, 2010.

[CLX11a] Jin-Yi Cai, Pinyan Lu, and Mingji Xia. Computational complexity of Holant problems. *SIAM J. Comput.*, 40(4):1101–1132, 2011.

[CLX11b] Jin-Yi Cai, Pinyan Lu, and Mingji Xia. Dichotomy for Holant* problems of Boolean domain. In *SODA*, pages 1714–1728, 2011.

[CLX12] Jin-Yi Cai, Pinyan Lu, and Mingji Xia. Holographic reduction, interpolation, and hardness. *Computational Complexity*, 21(4):573–604, 2012.

[CLX13] Jin-Yi Cai, Pinyan Lu, and Mingji Xia. Holographic algorithms by Fibonacci gates. *Linear Algebra Appl.*, 438(2):690–707, 2013.

[CLX14] Jin-Yi Cai, Pinyan Lu, and Mingji Xia. The complexity of complex weighted Boolean #CSP. *J. Comput. System Sci.*, 80(1):217–236, 2014.

[DDVM11] G. De las Cuevas, W. Dür, M. Van den Nest, and M. A. Martin-Delgado. Quantum algorithms for classical lattice models. *New Journal of Physics*, 13(9):093021, 2011.

[DFJ02] Martin E. Dyer, Alan M. Frieze, and Mark Jerrum. On counting independent sets in sparse graphs. *SIAM J. Comput.*, 31(5):1527–1541, 2002.

[DFK91] Martin E. Dyer, Alan M. Frieze, and Ravi Kannan. A random polynomial time algorithm for approximating the volume of convex bodies. *J. ACM*, 38(1):1–17, 1991.

[DG00a] Martin E. Dyer and Catherine S. Greenhill. The complexity of counting graph homomorphisms. *Random Struct. Algorithms*, 17(3-4):260–289, 2000.

[DG00b] Martin E. Dyer and Catherine S. Greenhill. On markov chains for independent sets. *J. Algorithms*, 35(1):17–49, 2000.

[DGGJ03] Martin E. Dyer, Leslie Ann Goldberg, Catherine S. Greenhill, and Mark Jerrum. The relative complexity of approximate counting problems. *Algorithmica*, 38(3):471–500, 2003.

[DGJ09] Martin E. Dyer, Leslie Ann Goldberg, and Mark Jerrum. The complexity of weighted Boolean CSP. *SIAM J. Comput.*, 38(5):1970–1986, 2009.

[DGJ10] Martin E. Dyer, Leslie Ann Goldberg, and Mark Jerrum. An approximation trichotomy for Boolean #CSP. *J. Comput. Syst. Sci.*, 76(3-4):267–277, 2010.

[DGL$^+$12] Jan Draisma, Dion C. Gijswijt, László Lovász, Guus Regts, and Alexander Schrijver. Characterizing partition functions of the vertex model. *J. Algebra*, 350:197–206, 2012.

[DP91] Christopher T. J. Dodson and Timothy Poston. *Tensor Geometry*, volume 130 of *Graduate Texts in Mathematics*. Springer-Verlag, second edition, 1991.

[Edm65] Jack Edmonds. Paths, trees, and flowers. *Canad. J. Math.*, 17:449–467, 1965.

[Fab08] John Faben. The complexity of counting solutions to generalised satisfiability problems modulo k. *CoRR*,, abs/0809.1836, 2008.

[FJ15] John Faben and Mark Jerrum. The complexity of parity graph homomorphism: an initial investigation. *Theory of Computing*, 11:35–57, 2015.

[FLS07] Michael Freedman, László Lovász, and Alexander Schrijver. Reflection positivity, rank connectivity, and homomorphism of graphs. *J. Amer. Math. Soc.*, 20(1):37–51, 2007.

[FM13] Keisuke Fujii and Tomoyuki Morimae. Quantum commuting circuits and complexity of Ising partition functions. *CoRR*,, abs/1311.2128, 2013.

[For01] G. David Forney Jr. Codes on graphs: normal realizations. *IEEE Transactions on Information Theory*, 47(2):520–548, 2001.

[Geo11] Hans-Otto Georgii. *Gibbs Measures and Phase Transitions*, volume 9 of *De Gruyter Studies in Mathematics*. de Gruyter, Berlin, second edition, 2011.

[GG14] Leslie Ann Goldberg and Heng Guo. The complexity of approximating complex-valued Ising and Tutte partition functions. *CoRR,*, abs/1409.5627, 2014.

[GGJT10] Leslie Ann Goldberg, Martin Grohe, Mark Jerrum, and Marc Thurley. A complexity dichotomy for partition functions with mixed signs. *SIAM J. Comput.*, 39(7):3336–3402, 2010.

[GGR14] Andreas Göbel, Leslie Ann Goldberg, and David Richerby. Counting homomorphisms to cactus graphs modulo 2. In *STACS*, pages 350–361, 2014.

[GGR15] Andreas Göbel, Leslie Ann Goldberg, and David Richerby. Counting homomorphisms to square-free graphs, modulo 2. *CoRR,*, abs/1501.07539, 2015.

[GGS+14] Andreas Galanis, Qi Ge, Daniel Stefankovic, Eric Vigoda, and Linji Yang. Improved inapproximability results for counting independent sets in the hard-core model. *Random Struct. Algorithms*, 45(1):78–110, 2014.

[GHLX11] Heng Guo, Sangxia Huang, Pinyan Lu, and Mingji Xia. The complexity of weighted Boolean #CSP modulo k. In *STACS*, pages 249–260, 2011.

[GJ07] Leslie Ann Goldberg and Mark Jerrum. The complexity of ferromagnetic Ising with local fields. *Combinatorics, Probability & Computing*, 16(1):43–61, 2007.

[GJ08] Leslie Ann Goldberg and Mark Jerrum. Inapproximability of the Tutte polynomial. *Inf. Comput.*, 206(7):908–929, 2008.

[GJ12a] Leslie Ann Goldberg and Mark Jerrum. Approximating the partition function of the ferromagnetic Potts model. *J. ACM*, 59(5):25, 2012.

[GJ12b] Leslie Ann Goldberg and Mark Jerrum. A counterexample to rapid mixing of the Ge-Štefankovič process. *Electron. Commun. Probab.*, 17(5):1–6, 2012.

[GJ14] Leslie Ann Goldberg and Mark Jerrum. The complexity of computing the sign of the Tutte polynomial. *SIAM J. Comput.*, 43(6):1921–1952, 2014.

[GJM15] Leslie Ann Goldberg, Mark Jerrum, and Colin McQuillan. Approximating the partition function of planar two-state spin systems. *J. Comput. Syst. Sci.*, 81(1):330–358, 2015.

[GJP03] Leslie Ann Goldberg, Mark Jerrum, and Mike Paterson. The computational complexity of two-state spin systems. *Random Struct. Algorithms*, 23(2):133–154, 2003.

[GK07] David Gamarnik and Dmitriy Katz. Correlation decay and deterministic FPTAS for counting list-colorings of a graph. In *SODA*, pages 1245–1254, 2007.

[GL15] Heng Guo and Pinyan Lu. Uniqueness, spatial mixing, and approximation for ferromagnetic 2-spin systems. 2015. *Submitted*.

[GLLY15] Heng Guo, Liang Li, Pinyan Lu, and Yitong Yin. Approximate counting via correlation decay up to uniqueness in two-state anti-ferromagnetic spin systems. 2015. *Manuscript*.

[GLV13] Heng Guo, Pinyan Lu, and Leslie G. Valiant. The complexity of symmetric Boolean parity Holant problems. *SIAM J. Comput.*, 42(1):324–356, 2013.

[GŠ12] Qi Ge and Daniel Štefankovič. A graph polynomial for independent sets of bipartite graphs. *Combinatorics, Probability & Computing*, 21(5):695–714, 2012.

[GŠV12] Andreas Galanis, Daniel Štefankovič, and Eric Vigoda. Inapproximability of the partition function for the antiferromagnetic Ising and Hard-Core models. *CoRR*,, abs/1203.2226, 2012.

[GŠV14] Andreas Galanis, Daniel Štefankovič, and Eric Vigoda. Inapproximability for antiferromagnetic spin systems in the tree non-uniqueness region. In *STOC*, pages 823–831, 2014.

[GŠVY14] Andreas Galanis, Daniel Štefankovič, Eric Vigoda, and Linji Yang. Ferromagnetic Potts model: Refined #BIS-hardness and related results. In *RANDOM*, pages 677–691, 2014.

[GW13] Heng Guo and Tyson Williams. The complexity of planar Boolean #CSP with complex weights. In *ICALP*, pages 516–527. Springer Berlin Heidelberg, 2013. *CoRR*, abs/1212.2284.

[HL12] Sangxia Huang and Pinyan Lu. A dichotomy for real weighted Holant problems. In *CCC*, pages 96–106. IEEE Computer Society, 2012. Full version available at `http://www.csc.kth.se/~sangxia/papers/2012-ccc.pdf`.

[Isi25] Ernst Ising. Beitrag zür theorie des ferromagnetismus. *Zeitschrift für Physik*, 31(1):253–258, 1925.

[JLR00] Svante Janson, Tomasz Łuczak, and Andrzej Ruciński. *Random Graphs*, volume 45 of *Wiley Series in Discrete Mathematics and Optimization*. John Wiley & Sons, 2000.

[JS89] Mark Jerrum and Alistair Sinclair. Approximating the permanent. *SIAM J. Comput.*, 18(6):1149–1178, 1989.

[JS93] Mark Jerrum and Alistair Sinclair. Polynomial-time approximation algorithms for the Ising model. *SIAM J. Comput.*, 22(5):1087–1116, 1993.

[JSV04] Mark Jerrum, Alistair Sinclair, and Eric Vigoda. A polynomial-time approximation algorithm for the permanent of a matrix with nonnegative entries. *J. ACM*, 51(4):671–697, 2004.

[JV14] Richard Jozsa and Marrten Van den Nest. Classical simulation complexity of extended Clifford circuits. *Quantum Info. Comput.*, 14(7&8):633–648, 2014.

[JVV86] Mark Jerrum, Leslie G. Valiant, and Vijay V. Vazirani. Random generation of combinatorial structures from a uniform distribution. *Theor. Comput. Sci.*, 43:169–188, 1986.

[JVW90] François Jaeger, Dirk L. Vertigan, and Dominic J. A. Welsh. On the computational complexity of the Jones and Tutte polynomials. *Math. Proc. Cambridge Philos. Soc.*, 108(1):35–53, 1990.

[Kas61] P. W. Kasteleyn. The statistics of dimers on a lattice. *Physica*, 27:1209–1225, 1961.

[Kas67] P. W. Kasteleyn. Graph theory and crystal physics. In F. Harary, editor, *Graph Theory and Theoretical Physics*, pages 43–110. Academic Press, London, 1967.

[Kel85] Frank P. Kelly. Stochastic models of computer communication systems. *Journal of the Royal Statistical Society. Series B (Methodological)*, 47(3):379–395, 1985.

[Kow10] Michael Kowalczyk. *Dichotomy theorems for Holant problems*. PhD thesis, University of Wisconsin—Madison, 2010. http://cs.nmu.edu/~mkowalcz/research/main.pdf.

[Lad75] Richard E. Ladner. On the structure of polynomial time reducibility. *J. ACM*, 22(1):155–171, 1975.

[Las88] Michel Las Vergnas. On the evaluation at $(3, 3)$ of the Tutte polynomial of a graph. *J. Comb. Theory, Ser. B*, 45(3):367–372, 1988.

[Lie67] Elliott H. Lieb. Residual entropy of square ice. *Phys. Rev.*, 162(1):162–172, 1967.

[LL15a] Jingcheng Liu and Pinyan Lu. FPTAS for #BIS with degree bounds on one side. In *STOC*, pages 549–556, 2015.

[LL15b] Jingcheng Liu and Pinyan Lu. FPTAS for counting monotone CNF. In *SODA*, pages 1531–1548, 2015.

[LLL14] Chengyu Lin, Jingcheng Liu, and Pinyan Lu. A simple FPTAS for counting edge covers. In *SODA*, pages 341–348, 2014.

[LLY12] Liang Li, Pinyan Lu, and Yitong Yin. Approximate counting via correlation decay in spin systems. In *SODA*, pages 922–940, 2012.

[LLY13] Liang Li, Pinyan Lu, and Yitong Yin. Correlation decay up to uniqueness in spin systems. In *SODA*, pages 67–84, 2013.

[LLZ14a] Jingcheng Liu, Pinyan Lu, and Chihao Zhang. The complexity of ferromagnetic two-spin systems with external fields. In *RANDOM*, pages 843–856, 2014.

[LLZ14b] Jingcheng Liu, Pinyan Lu, and Chihao Zhang. FPTAS for counting weighted edge covers. In *ESA*, pages 654–665, 2014.

[LMN13] J. M. Landsberg, Jason Morton, and Serguei Norine. Holographic algorithms without matchgates. *Linear Algebra Appl.*, 438(2):782–795, 2013.

[LS81] Elliott H. Lieb and Alan D. Sokal. A general Lee-Yang theorem for one-component and multicomponent ferromagnets. *Comm. Math. Phys.*, 80(2):153–179, 1981.

[LWZ14] Pinyan Lu, Menghui Wang, and Chihao Zhang. FPTAS for weighted Fibonacci gates and its applications. In *ICALP*, pages 787–799, 2014.

[LY52] Tsung-Dao Lee and Chen-Ning Yang. Statistical theory of equations of state and phase transitions. II. Lattice gas and Ising model. *Phys. Rev.*, 87(3):410–419, 1952.

[LY13] Pinyan Lu and Yitong Yin. Improved FPTAS for multi-spin systems. In *RANDOM*, pages 639–654, 2013.

[MM13] Jason Morton and Susan Margulies. Polynomial-time solvable #CSP problems via algebraic models and Pfaffian circuits. *CoRR*,, abs/1311.4066, 2013. To appear in Journal of Symbolic Computation.

[Mor11] Jason Morton. Pfaffian circuits. *CoRR*,, abs/1101.0129, 2011.

[MS13] Elchanan Mossel and Allan Sly. Exact thresholds for Ising-Gibbs samplers on general graphs. *Annals of Probability*, 41(1):294–328, 2013.

[MSW07] Fabio Martinelli, Alistair Sinclair, and Dror Weitz. Fast mixing for independent sets, colorings, and other models on trees. *Random Struct. Algorithms*, 31(2):134–172, 2007.

[MWW09] Elchanan Mossel, Dror Weitz, and Nicholas Wormald. On the hardness of sampling independent sets beyond the tree threshold. *Probability Theory and Related Fields*, 143(3-4):401–439, 2009.

[Ons44] Lars Onsager. Crystal statistics. I. A two-dimensional model with an order-disorder transition. *Phys. Rev.*, 65(3-4):117–149, 1944.

[PB83] J. Scott Provan and Michael O. Ball. The complexity of counting cuts and of computing the probability that a graph is connected. *SIAM J. Comput.*, 12(4):777–788, 1983.

[Pet91] Julius Petersen. Die theorie der regulären graphs. *Acta Mathematica*, 15(1):193–220, 1891.

[RST+13] Ricardo Restrepo, Jinwoo Shin, Prasad Tetali, Eric Vigoda, and Linji Yang. Improved mixing condition on the grid for counting and sampling independent sets. *Probability Theory and Related Fields*, 156(1-2):75–99, 2013.

[SB09] Dan J. Shepherd and Michael J. Bremner. Temporally unstructured quantum computation. *Proc. R. Soc. A*, 465(2105):1413–1439, 2009.

[Sch13] Alexander Schrijver. Characterizing partition functions of the spin model by rank growth. *Indag. Math. (N.S.)*, 24(4):1018–1023, 2013.

[She10] Dan Shepherd. Binary matroids and quantum probability distributions. *CoRR*,, abs/1005.1744, 2010.

[Sly10] Allan Sly. Computational transition at the uniqueness threshold. In *FOCS*, pages 287–296, 2010.

[Sok05] Alan D. Sokal. The multivariate Tutte polynomial (alias Potts model) for graphs and matroids. In *Surveys in combinatorics 2005*, volume 327 of *London Math. Soc. Lecture Note Ser.*, pages 173–226. Cambridge Univ. Press, 2005.

[SS14] Allan Sly and Nike Sun. The computational hardness of counting in two-spin models on d-regular graphs. *The Annals of Probability*, 42(6):2383–2416, 2014.

[SSŠY15] Alistair Sinclair, Piyush Srivastava, Daniel Štefankovič, and Yitong Yin. Spatial mixing and the connective constant: Optimal bounds. In *SODA*, pages 1549–1563, 2015.

[SST12] Alistair Sinclair, Piyush Srivastava, and Marc Thurley. Approximation algorithms for two-state anti-ferromagnetic spin systems on bounded degree graphs. In *SODA*, pages 941–953, 2012.

[SSY13] Alistair Sinclair, Piyush Srivastava, and Yitong Yin. Spatial mixing and approximation algorithms for graphs with bounded connective constant. In *FOCS*, pages 300–309, 2013.

[TF61] H. N. V. Temperley and M. E. Fisher. Dimer problem in statistical mechanics—an exact result. *Philosophical Magazine*, 6:1061–1063, 1961.

[Thu10] Marc Thurley. The complexity of partition functions on Hermitian matrices. *CoRR*,, abs/1004.0992, 2010.

[Val79a] Leslie G. Valiant. The complexity of computing the permanent. *Theor. Comput. Sci.*, 8(2):189–201, 1979.

[Val79b] Leslie G. Valiant. The complexity of enumeration and reliability problems. *SIAM J. Comput.*, 8(3):410–421, 1979.

[Val02a] Leslie G. Valiant. Expressiveness of matchgates. *Theor. Comput. Sci.*, 289(1):457–471, 2002.

[Val02b] Leslie G. Valiant. Quantum circuits that can be simulated classically in polynomial time. *SIAM J. Comput.*, 31(4):1229–1254, 2002.

[Val06] Leslie G. Valiant. Accidental algorthims. In *FOCS*, pages 509–517. IEEE Computer Society, 2006.

[Val08] Leslie G. Valiant. Holographic algorithms. *SIAM J. Comput.*, 37(5):1565–1594, 2008.

[Val10] Leslie G. Valiant. Some observations on holographic algorithms. In *LATIN*, pages 577–590. Springer Berlin Heidelberg, 2010.

[Ver91] Dirk L. Vertigan. *On the computational complexity of Tutte, Jones, Homfly and Kauffman invariants.* PhD thesis, University of Oxford, 1991.

[Ver05] Dirk Vertigan. The computational complexity of Tutte invariants for planar graphs. *SIAM J. Comput.*, 35(3):690–712, 2005.

[Vig00] Eric Vigoda. Improved bounds for sampling colorings. *J. Math. Phys.*, 41:1555, 2000.

[VV86] Leslie G. Valiant and Vijay V. Vazirani. NP is as easy as detecting unique solutions. *Theor. Comput. Sci.*, 47(3):85–93, 1986.

[Wei06] Dror Weitz. Counting independent sets up to the tree threshold. In *STOC*, pages 140–149, 2006.

[Wel93] Dominic Welsh. *Complexity: Knots, Colourings and Countings*. London Mathematical Society Lecture Note Series. Cambridge University Press, 1993.

[WJ08] Martin J. Wainwright and Michael I. Jordan. Graphical models, exponential families, and variational inference. *Foundations and Trends in Machine Learning*, 1(1-2):1–305, 2008.

[Xia11] Mingji Xia. Holographic reduction: A domain changed application and its partial converse theorems. *Int. J. Software and Informatics*, 5(4):567–577, 2011.

[Yan52] Chen-Ning Yang. The spontaneous magnetization of a two-dimensional Ising model. *Phys. Rev.*, 85(5):808–816, 1952.

[YL52] Chen-Ning Yang and Tsung-Dao Lee. Statistical theory of equations of state and phase transitions. I. Theory of condensation. *Phys. Rev.*, 87(3):404–409, 1952.

[Ziv82] Abraham Ziv. Relative distance—an error measure in round-off error analysis. *Math. Comp.*, 39(160):563–569, 1982.