



# Contents

|            |   |           |
|------------|---|-----------|
| <b>I</b>   | <b>Introduction</b>   | <b>1</b>  |
| <b>1</b>   | <b>Introduction</b>   | <b>1</b>  |
| <b>2</b>   | <b>Technical overview</b>   | <b>4</b>  |
| 2.1        | The MAX-CUT problem . . . . .   | 4         |
| 2.2        | The QUANTUM MAX-CUT problem . . . . .                                 | 5         |
| 2.3        | Semidefinite programming relaxations . . . . .                        | 6         |
| 2.4        | Rounding algorithms . . . . .   | 7         |
| 2.5        | SDP and algorithmic gaps . . . . .                                    | 8         |
| 2.6        | Our results . . . . .   | 9         |
| 2.7        | Proof overview . . . . .  | 10        |
| <b>3</b>   | <b>Conclusion and open questions</b>                                  | <b>12</b> |
| <b>II</b>  | <b>A vector-valued Borell’s inequality</b>                            | <b>14</b> |
| <b>4</b>   | <b>The spherical case</b>   | <b>16</b> |
| 4.1        | Spherical harmonics . . . . .   | 17        |
| 4.2        | Eigenvalues of the noise operator . . . . .                           | 18        |
| <b>5</b>   | <b>The full-dimensional case</b>                                      | <b>21</b> |
| 5.1        | The induced noise operator . . . . .                                  | 22        |
| 5.2        | The positive- $\rho$ case . . . . .                                   | 24        |
| <b>6</b>   | <b>Dimension reduction</b>  | <b>25</b> |
| 6.1        | Outline of the dimension reduction . . . . .                          | 25        |
| 6.2        | Technicalities . . . . .  | 26        |
| 6.3        | The first-order conditions . . . . .                                  | 30        |
| 6.4        | Spatial perturbations . . . . .                                       | 34        |
| 6.5        | Second variation . . . . .  | 35        |
| 6.6        | The index form for translations . . . . .                             | 36        |
| 6.7        | The case of negative $\rho$ . . . . .                                 | 38        |
| <b>III</b> | <b>Hardness of QUANTUM MAX-CUT</b>                                    | <b>39</b> |
| <b>7</b>   | <b>Preliminaries</b>  | <b>39</b> |
| 7.1        | Pauli matrices . . . . .  | 39        |
| 7.2        | On QUANTUM MAX-CUT and the Heisenberg model . . . . .                 | 39        |
| 7.3        | Alternative expressions for the QUANTUM MAX-CUT interaction . . . . . | 40        |
| 7.4        | Product states . . . . .  | 41        |
| 7.5        | Deriving the basic SDPs . . . . .                                     | 43        |
| 7.6        | Projection rounding . . . . .   | 45        |
| 7.7        | Fourier analysis on the hypercube . . . . .                           | 46        |

|           |  |           |
|-----------|--|-----------|
| <b>8</b>  | <b>Integrality gaps</b>                                  | <b>47</b> |
| 8.1       | The Gaussian graph as an integrality gap . . . . .       | 48        |
| 8.2       | Discretizing the Gaussian graph . . . . .                | 49        |
| <b>9</b>  | <b>Algorithmic gap for the BOV algorithm</b>             | <b>49</b> |
| <b>10</b> | <b>A dictator test for the product state value</b>       | <b>52</b> |
| <b>11</b> | <b>Unique Games hardness of QUANTUM MAX-CUT</b>          | <b>56</b> |
| <b>IV</b> | <b>Appendix</b>  | <b>61</b> |
| <b>A</b>  | <b>The non-commutative Sum of Squares hierarchy</b>      | <b>61</b> |
| A.1       | Sum-of-squares relaxations for MAX-CUT . . . . .         | 61        |
| A.2       | Sum-of-squares relaxations for QUANTUM MAX-CUT . . . . . | 62        |
| A.3       | Degree-two non-commutative Sum of Squares . . . . .      | 63        |
| <b>B</b>  | <b>Other Lemmas</b>                                      | <b>65</b> |
| B.1       | Cardinality Reduction . . . . .                          | 65        |
| B.2       | Proofs of Various Fourier Properties . . . . .           | 67        |
| B.3       | Proofs of Lipschitz Properties . . . . .                 | 67        |
| <b>C</b>  | <b>Rank-constrained MAX-CUT</b>                          | <b>70</b> |

## Part I

# Introduction

## 1 Introduction

Over the last 30 years, starting with the proof of the PCP theorem [AS98, ALM<sup>+</sup>98], researchers have gained a nearly complete understanding of the approximability of classical constraint satisfaction problems (CSPs), modulo the still-unproven Unique Games Conjecture (UGC) of Khot [Kho02]. Due to work of Raghavendra [Rag08, Rag09], we know that for each CSP, there is a “canonical algorithm” which achieves a certain approximation ratio  $\alpha > 0$ , and that it is NP-hard for any polynomial-time algorithm to do better than  $\alpha$ , assuming the UGC. This canonical algorithm is based on the “basic” *semidefinite programming (SDP)* relaxation of the CSP, which is itself derived from the second level of the *sum of squares (SoS)* hierarchy.

However, for the *quantum* analogue of CSPs, known as the *local Hamiltonian problem*, our understanding remains incomplete. The quantum analogue of the PCP theorem remains a conjecture, and there is no general theory of optimal algorithms, be they quantum or classical. In addition, there is a natural quantum analogue of the SoS hierarchy, but in spite of recent progress in analyzing it, many basic questions remain open, such as how best to round its solutions.

On top of this, approximation algorithms for the local Hamiltonian problem face an additional challenge not present for classical CSPs, namely what kind of quantum state should they output? The local Hamiltonian problem is an optimization problem over  $n$ -qubit states; however, actually outputting a general  $n$ -qubit state is infeasible on a classical computer, as it requires exponentially many bits to describe. Instead, algorithms typically output states from a subset of quantum states, called an *ansatz*, which can be efficiently represented on a classical computer. By far the most popular is the ansatz of *product states*, i.e. states of the form  $|\psi_1\rangle \otimes \cdots \otimes |\psi_n\rangle$ , which possess no entanglement but often give surprisingly good approximations to the optimal value [BH16]. In general, though, we lack a theory of optimal ansatzes for polynomial-time classical algorithms.

To study these questions, we focus on a special case of the local Hamiltonian problem known as QUANTUM MAX-CUT, which has been suggested as a useful testbed for designing approximation algorithms [GP19]. QUANTUM MAX-CUT is a natural maximization variant of the *anti-ferromagnetic Heisenberg XYZ model*, a classic family of QMA-complete [CM16, PM17] 2-local Hamiltonians first investigated by Heisenberg [Hei28] which models magnetic systems where interacting particles have opposing spins. The constraints in QUANTUM MAX-CUT involve pairs of qubits and, loosely speaking, enforce that the two qubits have opposing values in each of the Pauli  $X$ ,  $Y$ , and  $Z$  bases. As a result, it can be viewed as a quantum analogue of the classical MAX-CUT problem.

For the special case of finding the best product state in the QUANTUM MAX-CUT problem, an SDP algorithm of Briët, Oliveira, and Vallentin [BdOFV10] gives an approximation ratio of 0.956. For the more general case of finding the best (possibly entangled) state, Gharibian and Parekh [GP19] gave an algorithm with approximation ratio 0.498. Their algorithm, which uses the product state ansatz, is based on rounding the corresponding “basic” SDP, derived from the second level of the *noncommutative sum of squares (ncSoS)* hierarchy, a variant of the SoS hierarchy for optimization problems over matrices. Following this work, Anshu, Gosset, and Morenz [AGM20] gave an algorithm with an improved approximation ratio of 0.531. To achieve this, they used an ansatz which is more expressive than product states—tensor products of one- and two-qubit states—as it is known that product states cannot surpass approximation ratio 0.5. Parekh and Thompson [PT21a] then showed that a similar algorithm could be captured by the level-4 ncSoS hierarchy, with a slightly improved approximation ratio of 0.533.

The starting point of our work is a well-known statement from Gaussian geometry called *Borell’s inequality* [Bor85], which has played a central role in the inapproximability of the classical MAX-CUT problem ever since it was introduced to theoretical computer science in the work of Mossel, O’Donnell, and Oleszkiewicz [MOO10]. The earlier work of Khot, Kindler, Mossel, and O’Donnell [KKMO07] had shown that the Goemans-Williamson SDP algorithm [GW95] is the optimal polynomial-time approximation algorithm for MAX-CUT under the UGC, assuming an unproven statement they dubbed “Majority is Stablest” and left as a conjecture. One of the main results of [MOO10] was to supply a proof of this Majority is Stablest conjecture, which they did by translating it to Gaussian geometry and applying Borell’s inequality, thereby settling the UG-hardness of MAX-CUT. In addition, Borell’s inequality serves as an important technical ingredient in the analysis of the standard SDP integrality gap instance for MAX-CUT [FS02, OW08]. The first contribution of our work is to identify and conjecture a natural generalization of Borell’s inequality suitable for applications to QUANTUM MAX-CUT.

**Conjecture 1.1** (Vector-valued Borell’s inequality; negative  $\rho$  case). *Let  $k \leq n$  be positive integers and  $-1 \leq \rho \leq 0$ . Let  $f : \mathbb{R}^n \rightarrow S^{k-1}$ . In addition, let  $f_{\text{opt}} : \mathbb{R}^n \rightarrow S^{k-1}$  be defined by  $f_{\text{opt}}(x) = x_{\leq k} / \|x_{\leq k}\|$ , where  $x_{\leq k} = (x_1, \dots, x_k)$ . Then*

$$\mathbf{E}_{\mathbf{u} \sim_{\rho} \mathbf{v}} \langle f(\mathbf{u}), f(\mathbf{v}) \rangle \geq \mathbf{E}_{\mathbf{u} \sim_{\rho} \mathbf{v}} \langle f_{\text{opt}}(\mathbf{u}), f_{\text{opt}}(\mathbf{v}) \rangle,$$

where  $\mathbf{u} \sim_{\rho} \mathbf{v}$  means that  $\mathbf{u}$  and  $\mathbf{v}$  are  $\rho$ -correlated Gaussian vectors.

We also conjecture that the optimizers  $f_{\text{opt}}$  are unique, up to some natural symmetries; see [Conjecture 3.4](#) for the full statement.

The original Borell’s inequality is when  $k = 1$ ; in this case  $f_{\text{opt}}$  is just defined as  $f_{\text{opt}}(x) = \text{sgn}(x_1)$ , and the quantity  $\mathbf{E}_{\mathbf{u} \sim_{\rho} \mathbf{v}} [f(\mathbf{u})f(\mathbf{v})]$  is known as the *Gaussian noise stability* of  $f$ . When the sphere  $S^{k-1}$  is replaced by the probability simplex  $\Delta^{k-1}$ , the analogue of [Conjecture 1.1](#) is a well-known open problem—known as the “Peace Sign Conjecture” [IM12]—that was recently solved when  $\rho$  is positive but sufficiently close to zero [HT20].

Although we are unable to prove [Conjecture 1.1](#), we are able to give evidence in favor of it. For our first piece of evidence, we show that [Conjecture 1.1](#) is true when  $n = k$ , which we prove via a spectral argument. For our second piece of evidence, we instead consider the “positive  $\rho$  case” of  $0 \leq \rho \leq 1$ , in which we would like to identify the function which *maximizes* the expression  $\mathbf{E}_{\mathbf{u} \sim_{\rho} \mathbf{v}} \langle f(\mathbf{u}), f(\mathbf{v}) \rangle$  rather than minimizes it. In this case, it is clear that our conjectured optimizer  $f_{\text{opt}}(x) = x_{\leq k} / \|x_{\leq k}\|$  is *not* the maximizer, as any constant function will have value 1, but it is plausible that  $f_{\text{opt}}$  is still the maximizer among all functions which are mean-zero. For our second piece of evidence, then, we show a dimensionality reduction statement in the positive  $\rho$  case, implying that any maximizing mean-zero  $f$  is “intrinsically”  $k$ -dimensional; in other words, it suffices to only consider the  $n = k$  case. To show this, we extend the recent calculus of variations approach for proving Borell’s inequality due to Heilman and Tarter [HT20] to output dimensions of size  $k$  larger than 1. Unfortunately, since our reduction to the  $n = k$  case only holds for positive  $\rho$ , and our proof of the  $n = k$  case only holds for negative  $\rho$ , these two pieces do not combine to give a full proof of the conjecture. For more details, along with formal statements of these results, see [Part II](#).

Next, we show that the vector-valued Borell’s inequality allows us to address several questions related to the optimality of the SDP algorithms for QUANTUM MAX-CUT.

**Theorem 1.2** (Main results, informal). *Suppose that the vector-valued Borell’s inequality is true. Then the following statements hold.*

1. *There exists an integrality gap of 0.498 for the basic SDP, matching the rounding algorithm of Gharibian and Parekh [GP19]. This shows that the product state ansatz is optimal for the basic SDP. Combined with the work of Anshu, Gosset, and Morenz [AGM20], this also shows that the basic SDP does not achieve the optimal approximation ratio.*
2. *It is Unique-Games-hard (UG-hard) to compute a  $(0.956 + \epsilon)$ -approximation to the value of the best product state, matching an approximation algorithm due to Briët, Oliveira, and Vallentin [BdOFV10]. More generally, for any fixed  $k$ , it is UG-hard to outperform the approximation algorithm of Briët, Oliveira, and Vallentin [BdOFV10] for Rank- $k$  MAX-CUT ( $k = 3$  corresponding to the aforementioned 0.956-approximation).*
3. *It is UG-hard to compute a  $(0.956 + \epsilon)$ -approximation to the value of the best (possibly entangled) state.*

To our knowledge, this is the first constant-factor hardness of approximation result for a natural family of local Hamiltonians which does not already contain a hard-to-approximate classical CSP as a special case, modulo our conjectured vector-valued Borell’s inequality. We also show that our conjecture implies sharp inapproximability results for QUANTUM MAX-CUT with respect to product states. Finally, in a striking departure from classical CSPs, in which the level-2 SoS relaxation is optimal under the UGC, we show that our conjecture implies that the level-4 ncSoS relaxation for QUANTUM MAX-CUT strictly improves upon the level-2 relaxation. Our results highlight the importance of gaining a better understanding of the level-4 ncSoS relaxation, as well as the importance of settling [Conjecture 1.1](#). The relevance of [Conjecture 1.1](#) to our main results is discussed in more detail in [Section 2.7](#) below.

**Related work.** An earlier draft of this work incorrectly claimed a complete proof of the vector-valued Borell’s inequality. The bug was in the proof of the dimensionality reduction step, which we incorrectly claimed held for both positive *and* negative  $\rho$ . We thank Steve Heilman for pointing out the error in this proof. Subsequent to the original posting of this work, Parekh and Thompson gave an algorithm for QUANTUM MAX-CUT based on the level-4 ncSoS relaxation which achieves a 0.5-approximation and uses the product state ansatz [PT22]. This is optimal, as there are simple graphs of value 1 in which product states achieve value at most 0.5 [GP19]. Combined with our [Theorem 1.2](#) and assuming the vector-valued Borell’s inequality, this shows that level-4 ncSoS outperforms level-2, even when one is restricted to using the product state ansatz.

In the local Hamiltonian problem, one is given a set of constraints on a system of  $n$  qubits, and the objective is to find the “ground state energy,” i.e. the optimum energy of a state under these constraints. This is a central problem in both the fields of condensed matter physics and quantum computing, and the study of this problem has led to a rich exchange of ideas between the two; see [GHLS15] for an excellent survey on the topic. Its importance in condensed matter physics stems from the fact that many interesting real-world systems can be modeled as local Hamiltonians, with the ground state energy corresponding to the energy of the system at zero temperature. In quantum computing, it is the canonical QMA-complete problem [KSV02], and is thus intractable to solve exactly with a classical computer unless  $\text{BPP} = \text{QMA}$ .

Classical CSPs form a special case of the local Hamiltonian problem, and so the classical PCP theorem applies to local Hamiltonians as well. This means that given a general instance of the local Hamiltonian problem, it is NP-hard (though not necessarily NP-complete) to estimate its ground-state energy to a certain constant accuracy. The *quantum PCP conjecture* [AAV13] asserts that this task is in fact QMA-complete. One of the key differences between these two possibilities

is that every local Hamiltonian has an efficient quantum witness of its ground state energy, namely its  $n$ -qubit ground state, but it may not have an efficient *classical* (i.e. NP) witness, as the ground state need not be efficiently representable on a classical computer.

Researchers have designed classical approximation algorithms for various classes of 2-local Hamiltonian problems, the majority of which use the product state ansatz. These include algorithms for Hamiltonians whose local terms are positive semi-definite [GK12, HLP20, PT21b, PT22], traceless Hamiltonians [HM17, BGKT19, BBT09, PT21b], and fermionic Hamiltonians [BGKT19, HO22]. In condensed matter physics, the use of product states as an ansatz is widespread and known as *mean-field theory* [GHLS15]; in quantum chemistry, it is known as the *Hartree-Fock method* [BH16]. The ubiquity of this ansatz stems both from the ease with which it can be analyzed and from folklore that product states well-approximate ground states in some situations. This folklore was formalized in the work of Brandão and Harrow [BH16], which showed that product states give a good approximation to the ground states of local Hamiltonians whose interaction graphs are high degree or sufficiently good expanders, due to monogamy of entanglement. As a result, these Hamiltonians cannot serve as hard instances for the quantum PCP conjecture, as product states have an efficient classical description. They then used this structural result to design approximation algorithms for Hamiltonians with interaction graphs which are either planar, dense, or have low threshold rank. This is an example of how the study of approximation algorithms can shed light on the limitations of the quantum PCP conjecture.

To our knowledge, the only classical approximation algorithms which do not use the product state ansatz are the aforementioned algorithms of [AGM20, PT21a], which use tensor products of one- and two-qubit states (and which still rely heavily on the product state algorithms), and the intriguing recent algorithms of [AGM20, AGKS21], which use an ansatz consisting of states of the form  $|u\rangle = U|\psi\rangle$ , where  $|\psi\rangle$  is a product state and  $U$  is a low-depth quantum circuit. These two works give several settings in which product states can be “improved” by post-processing them with a low-depth quantum circuit; for example, the former work [AGM20] shows an algorithm using this ansatz which is guaranteed to outperform any product state algorithms on degree-3 and 4 instances of QUANTUM MAX-CUT.

## 2 Technical overview

In this section, we give a technical overview of our results. We begin with definitions of the problems we consider. Then, we state SDP relaxations for these problems and rounding algorithms for these SDPs that have been considered in the literature. Finally, we state our results formally and give an overview of our proofs.

### 2.1 The MAX-CUT problem

The classical analogue of the QUANTUM MAX-CUT problem is the MAX-CUT problem. The simplest of all nontrivial CSPs, this is the problem of partitioning the vertices of a graph into two sets in order to maximize the number of edges crossing the partition.

**Notation 2.1.** We use **boldface** to denote random variables.

**Definition 2.2** (Weighted graph). A *weighted graph*  $G = (V, E, w)$  is an undirected graph with weights on the edges specified by  $w : E \rightarrow \mathbb{R}^{\geq 0}$ . The weights specify a probability distribution on the edges, i.e.  $\sum_{e \in E} w(e) = 1$ . We write  $\mathbf{e} \sim E$  or  $(\mathbf{u}, \mathbf{v}) \sim E$  for a random edge sampled from this distribution. We generally use *graph* as shorthand for weighted graph.

**Definition 2.3** (MAX-CUT). Given a graph  $G = (V, E, w)$ , a *cut* is a function  $f : V \rightarrow \{\pm 1\}$ . The *value* of the cut is

$$\mathbf{E}_{(\mathbf{u}, \mathbf{v}) \sim E} \left[ \frac{1}{2} - \frac{1}{2} f(\mathbf{u}) f(\mathbf{v}) \right].$$

MAX-CUT is the problem of finding the value of the largest cut, i.e. the quantity

$$\text{MAX-CUT}(G) = \max_{f: V \rightarrow \{\pm 1\}} \mathbf{E}_{(\mathbf{u}, \mathbf{v}) \sim E} \left[ \frac{1}{2} - \frac{1}{2} f(\mathbf{u}) f(\mathbf{v}) \right].$$

MAX-CUT appears on Karp's original list of NP-complete problems [Kar72], and it is NP-hard to approximate to a factor better than  $\frac{16}{17}$  [TSSW00]. Goemans and Williams gave an algorithm using the basic SDP with approximation ratio 0.878 [GW95], and this SDP algorithm was later shown to be optimal by [KKMO07], at least assuming the UGC. Even without assuming the UGC, though, it is known that the SoS hierarchy is still the optimal algorithm among a large class of algorithms for MAX-CUT, namely those given by polynomial-size SDPs [LRS15].

## 2.2 The QUANTUM MAX-CUT problem

The main focus of this work is the QUANTUM MAX-CUT problem, a special case of the local Hamiltonian problem first introduced by Gharibian and Parekh in [GP19]. Although the local Hamiltonian problem is typically stated as a minimization problem, they instead defined QUANTUM MAX-CUT to be a maximization problem, as this makes it more convenient to study from an approximation algorithms perspective and resembles MAX-CUT. As stated earlier, QUANTUM MAX-CUT is a natural maximization variant of the anti-ferromagnetic Heisenberg XYZ model; we discuss this viewpoint, as well as the Heisenberg model in greater detail, in Section 7.2.

**Definition 2.4** (The QUANTUM MAX-CUT interaction). The QUANTUM MAX-CUT *interaction* is the 2-qubit operator  $h = \frac{1}{4}(I \otimes I - X \otimes X - Y \otimes Y - Z \otimes Z)$ .

Here  $X$ ,  $Y$ , and  $Z$  refer to the standard *Pauli matrices*. Intuitively, the QUANTUM MAX-CUT interaction, when applied to a pair of qubits, enforces that they are opposites in the  $X$ ,  $Y$ , and  $Z$  bases.

**Definition 2.5** (QUANTUM MAX-CUT). Let  $G = (V, E, w)$  be a graph known as the *interaction graph*. The corresponding instance of the QUANTUM MAX-CUT *problem* is the matrix which acts on  $(\mathbb{C}^2)^{\otimes V}$  given by

$$H_G = \sum_{(\mathbf{u}, \mathbf{v}) \in E} w_{\mathbf{u}, \mathbf{v}} \cdot h_{\mathbf{u}, \mathbf{v}} = \mathbf{E}_{(\mathbf{u}, \mathbf{v}) \sim E} h_{\mathbf{u}, \mathbf{v}}.$$

Here,  $h_{\mathbf{u}, \mathbf{v}} \in \mathbb{C}^{2^{|\mathbf{u}|} \times 2^{|\mathbf{v}|}}$  is shorthand for  $h_{\mathbf{u}, \mathbf{v}} \otimes I_{V \setminus \{\mathbf{u}, \mathbf{v}\}}$ , where  $h_{\mathbf{u}, \mathbf{v}}$  is the QUANTUM MAX-CUT interaction applied to the qubits  $\mathbf{u}$  and  $\mathbf{v}$ .

**Definition 2.6** (Energy). Let  $H_G$  be an instance of QUANTUM MAX-CUT. Given a state  $|\psi\rangle \in (\mathbb{C}^2)^{\otimes V}$ , its *value* or *energy* is the quantity  $\langle \psi | H_G | \psi \rangle$ . The *maximum energy* of  $H_G$ , also referred to as its *value*, is

$$\text{QMAX-CUT}(G) = \lambda_{\max}(H_G) = \max_{|\psi\rangle \in (\mathbb{C}^2)^{\otimes V}} \langle \psi | H_G | \psi \rangle.$$

To see one way QUANTUM MAX-CUT and MAX-CUT are related, if we let  $D_G$  be the diagonal matrix consisting of the diagonal entries of  $H_G$  in the computational basis, then  $\lambda_{\max}(D_G)$  is half the value of the maximum cut in  $G$ .

The second problem we consider is the special case of QUANTUM MAX-CUT when the optimization is only over product states, which has been a common approach in approximation algorithms for the local Hamiltonian problem.

**Definition 2.7** (Product state value). The *product state value* of  $H_G$  is

$$\text{PROD}(G) = \max_{\forall v \in V, |\psi_v\rangle \in \mathbb{C}^2} \langle \psi_G | H_G | \psi_G \rangle, \text{ where } |\psi_G\rangle = \otimes_{v \in V} |\psi_v\rangle.$$

There is an alternative expression for the product state value which we will find convenient to use. It is related to the *Bloch sphere* representation of qubits; see [Section 7.4](#) for a proof.

**Definition 2.8** (Balls and spheres). Given a dimension  $d \geq 1$ , the  $d$ -dimensional unit ball and sphere are given by  $B^d = \{x \in \mathbb{R}^d \mid \|x\| \leq 1\}$ , and  $S^{d-1} = \{x \in \mathbb{R}^d \mid \|x\| = 1\}$ , respectively.

**Proposition 2.9** (Rewriting the product state value).

$$\text{PROD}(G) = \max_{f: V \rightarrow S^2} \mathbf{E}_{(\mathbf{u}, \mathbf{v}) \sim E} \left[ \frac{1}{4} - \frac{1}{4} \langle f(\mathbf{u}), f(\mathbf{v}) \rangle \right].$$

Note the similarity to the definition of MAX-CUT ([Definition 2.3](#)): aside from the extra factor of  $\frac{1}{2}$ , the key distinction is that the function  $f$  has range  $S^2$  rather than  $S^0 = \{-1, 1\}$ . As a result, the product state value can be viewed as an additional quantum generalization of MAX-CUT. We may further generalize MAX-CUT by allowing  $f$  to have range  $S^{k-1}$ , yielding the rank-constrained version of MAX-CUT studied by Briët, Oliveira, and Vallentin [[BdOFV10](#)].

**Definition 2.10** (Rank- $k$  MAX-CUT).

$$\text{MAX-CUT}_k(G) = \max_{f: V \rightarrow S^{k-1}} \mathbf{E}_{(\mathbf{u}, \mathbf{v}) \sim E} \left[ \frac{1}{2} - \frac{1}{2} \langle f(\mathbf{u}), f(\mathbf{v}) \rangle \right].$$

Note that  $\text{MAX-CUT}_3(G) = 2 \cdot \text{PROD}(G)$ . Indeed, what we refer to as the BOV algorithm for the product state value was actually originally stated in [[BdOFV10](#)] as an algorithm for  $\text{MAX-CUT}_3(G)$ , though it applies equally well to both cases.

## 2.3 Semidefinite programming relaxations

A standard approach for solving MAX-CUT is through its SDP relaxation.

**Definition 2.11** (The MAX-CUT SDP). Let  $G = (V, E, w)$  be an  $n$ -vertex graph. The value of the MAX-CUT SDP is

$$\text{SDP}_{\text{MC}}(G) = \max_{f: V \rightarrow S^{n-1}} \mathbf{E}_{(\mathbf{u}, \mathbf{v}) \sim E} \left[ \frac{1}{2} - \frac{1}{2} \langle f(\mathbf{u}), f(\mathbf{v}) \rangle \right].$$

This is a relaxation because the optimal objective of the MAX-CUT SDP is at least as large as  $\text{MAX-CUT}(G)$  for all graphs  $G$ . It can be approximated in polynomial time, meaning one can compute an  $f: V \rightarrow S^{n-1}$  of value  $\text{SDP}_{\text{MC}}(G) - \epsilon$  in time  $\text{poly}(n) \cdot \log(1/\epsilon)$ <sup>1</sup>. In addition, it is equivalent to the level-2 SoS relaxation for MAX-CUT. (Note also that  $\text{SDP}_{\text{MC}}(G) = \text{MAX-CUT}_n(G)$ .)

There is a similar SDP relaxation for the product state value of QUANTUM MAX-CUT.

**Definition 2.12** (The product state SDP).

$$\text{SDP}_{\text{PROD}}(G) = \max_{f: V \rightarrow S^{n-1}} \mathbf{E}_{(\mathbf{u}, \mathbf{v}) \sim E} \left[ \frac{1}{4} - \frac{1}{4} \langle f(\mathbf{u}), f(\mathbf{v}) \rangle \right].$$

<sup>1</sup>The usual considerations show arbitrary additive approximation in polynomial time, e.g. see [[VB96](#)].

This is in fact the SDP relaxation given by level-2 of the SoS hierarchy applied to the product state value. Note that  $\text{SDP}_{\text{PROD}}(G) = \frac{1}{2}\text{SDP}_{\text{MC}}(G)$ .

Now we state the SDP relaxation we will use for the maximum energy of QUANTUM MAX-CUT, which is equivalent to the level-2 ncSoS relaxation of the maximum energy.

**Definition 2.13** (The QUANTUM MAX-CUT SDP).

$$\text{SDP}_{\text{QMC}}(G) = \max_{f:V \rightarrow S^{n-1}} \mathbf{E}_{(\mathbf{u}, \mathbf{v}) \sim E} \left[ \frac{1}{4} - \frac{3}{4} \langle f(\mathbf{u}), f(\mathbf{v}) \rangle \right].$$

Deriving this requires a bit more care than either of the MAX-CUT or product state SDPs. Indeed, it is not even obvious at first glance that this is a legitimate relaxation! We include a proof of this fact in [Section 7.5](#) and a discussion of the ncSoS hierarchy in [Appendix A](#). To our knowledge, we are the first to observe this particularly simple form of the QUANTUM MAX-CUT SDP. Note that  $\text{SDP}_{\text{QMC}}(G) = \frac{3}{2} \cdot \text{SDP}_{\text{MC}}(G) - \frac{1}{2}$ .

It may be surprising that these three very different optimization problems yield such similar SDPs. Indeed, the optimizing function  $f : V \rightarrow S^{n-1}$  is the same in each of them. In spite of this, however, the quality of the SDP relaxation is different in each case because each case features a different objective function to compare the SDP against.

## 2.4 Rounding algorithms

Most SDP algorithms work by computing the optimal SDP solution and then converting it to a solution to the original optimization problem in a process known as *rounding*. We will look at the standard Goemans-Williamson algorithm, used to round the MAX-CUT SDP, and two generalizations of this algorithm, used to round the product state and QUANTUM MAX-CUT SDPs.

**Halfspace rounding.** The Goemans-Williamson algorithm [[GW95](#)] uses a procedure called “halfspace rounding” to round the MAX-CUT SDP optimum  $f_{\text{SDP}} : V \rightarrow S^{n-1}$  into a random cut  $\mathbf{f} : V \rightarrow \{-1, 1\}$ . Halfspace rounding works as follows:

1. Sample a random vector  $\mathbf{z} = (z_1, \dots, z_n)$  from the  $n$ -dimensional Gaussian distribution.
2. For each  $u \in V$ , set  $\mathbf{f}(u)$  equal to

$$\mathbf{f}(u) := \text{sgn}(\langle \mathbf{z}, f_{\text{SDP}}(u) \rangle) = \frac{\langle \mathbf{z}, f_{\text{SDP}}(u) \rangle}{|\langle \mathbf{z}, f_{\text{SDP}}(u) \rangle|}.$$

Goemans and Williamson showed that for each  $u, v \in V$ ,

$$\mathbf{E}_{\mathbf{f}} \left[ \frac{1}{2} - \frac{1}{2} \mathbf{f}(u) \mathbf{f}(v) \right] \geq \alpha_{\text{GW}} \cdot \left[ \frac{1}{2} - \frac{1}{2} \langle f_{\text{SDP}}(u), f_{\text{SDP}}(v) \rangle \right], \quad (1)$$

where  $\alpha_{\text{GW}} = 0.878567$ . Taking an average over the edges in  $G$ , this shows that the expected value of  $\mathbf{f}$  is at least  $\alpha_{\text{GW}} \cdot \text{SDP}_{\text{MC}}(G)$ , and hence at least  $\alpha_{\text{GW}} \cdot \text{MAX-CUT}(G)$ . This shows that the Goemans-Williamson algorithm is an  $\alpha_{\text{GW}}$ -approximation algorithm.

**Projection rounding.** Briët, Oliveira, and Vallentin [[BdOFV10](#)] suggested a generalization of halfspace rounding, which we refer to as “projection rounding”, in order to round solutions of  $\text{SDP}_{\text{PROD}}(G)$ . The goal is to convert an SDP solution  $f_{\text{SDP}} : V \rightarrow S^{n-1}$  into a function  $\mathbf{f} : V \rightarrow S^2$ , which can then be converted into a product state via [Definition 7.7](#). Projection rounding works as follows:

1. Sample a random  $3 \times n$  matrix  $\mathbf{Z}$  consisting of  $3n$  i.i.d. standard Gaussians.
2. For each  $u \in V$ , set  $\mathbf{f}(u) = \mathbf{Z}f_{\text{SDP}}(u)/\|\mathbf{Z}f_{\text{SDP}}(u)\|_2$ .

Briët, Oliveira, and Vallentin showed that for each  $u, v \in V$ ,

$$\mathbf{E}_{\mathbf{f}}\left[\frac{1}{4} - \frac{1}{4}\langle \mathbf{f}(u), \mathbf{f}(v) \rangle\right] \geq \alpha_{\text{BOV}} \cdot \left[\frac{1}{4} - \frac{1}{4}\langle f_{\text{SDP}}(u), f_{\text{SDP}}(v) \rangle\right], \quad (2)$$

where  $\alpha_{\text{BOV}} = 0.956$ . With the same reasoning as above, they conclude the following.

**Theorem 2.14** (Performance of the BOV algorithm [BdOFV10]). *The Briët-Oliveira-Vallentin algorithm for the product state value achieves approximation ratio  $\alpha_{\text{BOV}}$ .*

Next, Gharibian and Parekh [GP19] used projection rounding to round solutions of  $\text{SDP}_{\text{QMC}}(G)$  into product states. Like [BdOFV10], this involves rounding  $f_{\text{SDP}} : V \rightarrow S^{n-1}$ , now the solution of  $\text{SDP}_{\text{QMC}}(G)$ , into  $\mathbf{f} : V \rightarrow S^2$ . They establish the inequality

$$\mathbf{E}_{\mathbf{f}}\left[\frac{1}{4} - \frac{1}{4}\langle \mathbf{f}(u), \mathbf{f}(v) \rangle\right] \geq \alpha_{\text{GP}} \cdot \left[\frac{1}{4} - \frac{3}{4}\langle f_{\text{SDP}}(u), f_{\text{SDP}}(v) \rangle\right], \quad (3)$$

where  $\alpha_{\text{GP}} = 0.498$ . Note that this inequality possesses an asymmetry not present in Eqs. (1) and (2): the coefficient of  $\mathbf{f}(u)\mathbf{f}(v)$  on the left-hand side is  $\frac{1}{4}$  but the coefficient of  $\langle f_{\text{SDP}}(u), f_{\text{SDP}}(v) \rangle$  on the right-hand side is  $\frac{3}{4}$ . This asymmetry comes about because they are solving the SDP relaxation for the maximum energy, which is an optimization over all states, but only rounding into the set of product states. Nevertheless, this yields the following theorem.

**Theorem 2.15** (Performance of the GP algorithm [GP19]). *The Gharibian-Parekh algorithm for QUANTUM MAX-CUT achieves approximation ratio  $\alpha_{\text{GP}}$ .*

## 2.5 SDP and algorithmic gaps

The quality of an SDP relaxation is traditionally measured through its *integrality gap*.

**Definition 2.16** (Integrality gap). Let  $\mathcal{P}$  denote a maximization problem and let  $\text{SDP}(\cdot)$  be a semidefinite programming relaxation for  $\mathcal{P}$ . Given an instance  $\mathcal{I}$  of  $\mathcal{P}$ , its *integrality gap* is the quantity

$$\text{GapSDP}(\mathcal{I}) = \frac{\text{OPT}(\mathcal{I})}{\text{SDP}(\mathcal{I})}.$$

The *integrality gap* of the SDP is defined to be the minimum integrality gap among all instances, i.e.

$$\inf_{\text{instances } \mathcal{I}} \{\text{GapSDP}(\mathcal{I})\}.$$

The integrality gap of an SDP serves as a bound on the approximation ratio of any algorithm based on rounding its solutions. This is because one typically analyzes a rounding algorithm by comparing the value of its solution to the value of the SDP, as done for the rounding algorithms in Section 2.4. We refer to this as the *standard analysis* of rounding algorithms. We therefore typically view a rounding algorithm as optimal if its worst-case performance matches the integrality gap.

Usually, though, one actually cares about how the rounded solution compares to the optimal value, not the SDP value. To show that one has given a tight analysis of an algorithm's approximation ratio, one must actually exhibit a matching *algorithmic gap*.

**Definition 2.17** (Algorithmic gap). Let  $\mathcal{P}$  denote a maximization problem. Let  $A$  be an approximation algorithm for  $\mathcal{P}$ , and let  $A(\mathcal{I})$  be the expected value of the solution it outputs on input  $\mathcal{I}$ . Given an instance  $\mathcal{I}$ , its *algorithmic gap* is the quantity

$$\text{Gap}_A(\mathcal{I}) = \frac{A(\mathcal{I})}{\text{OPT}(\mathcal{I})}.$$

The *algorithmic gap* of  $A$  is defined to be the minimum algorithmic gap among all instances, i.e.

$$\inf_{\text{instances } \mathcal{I}} \{\text{Gap}_A(\mathcal{I})\}.$$

## 2.6 Our results

Our first set of results are a pair of integrality gaps for the QUANTUM MAX-CUT and product state SDPs.

**Theorem 2.18** (Integrality gap for the QUANTUM MAX-CUT SDP). *Assuming [Conjecture 1.1](#), the QUANTUM MAX-CUT semidefinite program  $\text{SDP}_{\text{QMC}}(G)$  has integrality gap  $\alpha_{\text{GP}}$ .*

Assuming the vector-valued Borell’s inequality, this matches the approximation ratio of the GP algorithm and shows that projection rounding is optimal for  $\text{SDP}_{\text{QMC}}(G)$ . In addition, it shows that product states are the optimal ansatz for this SDP, as the GP algorithm outputs product states. Finally, it implies that the GP algorithm is strictly worse than the algorithms of [\[AGM20, PT21a\]](#), and that level-4 of the ncSoS hierarchy strictly improves upon level-2 of the ncSoS hierarchy.

**Theorem 2.19** (Integrality gap for product state SDP). *Assuming [Conjecture 1.1](#), the product state semidefinite program  $\text{SDP}_{\text{PROD}}(G)$  has integrality gap  $\alpha_{\text{BOV}}$ .*

This matches the approximation ratio of the BOV algorithm and shows that projection rounding is optimal for  $\text{SDP}_{\text{PROD}}(G)$ , assuming [Conjecture 1.1](#). Next, we show an algorithmic gap for the product state SDP. This shows that the “standard analysis” of the BOV algorithm is sharp, and so its approximation ratio is  $\alpha_{\text{BOV}}$  exactly. We note that this result is unconditional, and therefore not reliant on [Conjecture 1.1](#).

**Theorem 2.20** (Algorithmic gap for product state SDP). *The Briët-Oliveira-Vallentin algorithm has algorithmic gap  $\alpha_{\text{BOV}}$ .*

Finally, we prove a Unique Games-hardness result for the product state value. This uses the standard framework of [\[KKMO07, Rag08\]](#) for translating SDP integrality gaps into inapproximability results. We apply this framework to the integrality gap from [Theorem 2.19](#).

**Theorem 2.21** (Inapproximability of the product state value). *Assuming [Conjecture 1.1](#) and the Unique Games Conjecture, it is NP-hard to approximate  $\text{PROD}(G)$  to within a factor of  $\alpha_{\text{BOV}} + \epsilon$ , for all  $\epsilon > 0$ .*

This shows that the BOV algorithm is optimal, assuming [Conjecture 1.1](#) and the UGC. Next, we observe that the QUANTUM MAX-CUT instances which occur in this proof have interaction graphs of high degree. Hence, by [\[BH16\]](#) their product state value is roughly identical to their maximum energy. As a consequence, we also derive a Unique-Games hardness result for the maximum energy.

**Theorem 2.22** (Inapproximability of QUANTUM MAX-CUT). *Assuming [Conjecture 1.1](#) and the Unique Games Conjecture, it is NP-hard to approximate  $\text{QMAX-CUT}(G)$  to within a factor of  $\alpha_{\text{BOV}} + \epsilon$ , for all  $\epsilon > 0$ .*

This is our one result which is *not* tight, to our knowledge, as the best known approximation for QUANTUMMAX-CUT achieves approximation 0.533 [PT21a], which is less than  $\alpha_{\text{BOV}} = 0.956$ . The difficulty is that  $\text{SDP}_{\text{QMC}}(G)$  is not an optimal SDP, as it is outperformed by the algorithms of [AGM20, PT21a], and so we cannot convert an integrality gap for it into a UG-hardness result.

We also generalize our results for the product state value to hold for MAX-CUT $_k$  for any fixed  $k$ . See Appendix C for more details.

## 2.7 Proof overview

Our proof of Theorems 2.18 and 2.19 is inspired by a well-known integrality gap construction for the MAX-CUT SDP due to Feige and Schechtman [FS02] which achieves an integrality gap of  $\alpha_{\text{GW}}$ . We will begin with an overview of this construction and a related construction called the ‘‘Gaussian graph’’, and then we will discuss how to modify these constructions to give integrality gaps for the two QUANTUM MAX-CUT SDPs.

**Integrality gap for the MAX-CUT SDP.** The construction of the Feige-Schechtman MAX-CUT integrality gap is motivated by the following two desiderata.

1. Given an optimal solution  $f_{\text{SDP}} : V \rightarrow S^{n-1}$  to the MAX-CUT SDP, halfspace rounding outputs a random solution  $\mathbf{f}$  with expected value exactly equal to  $\alpha_{\text{GW}} \cdot \text{SDP}_{\text{MC}}(G)$ .
2. This random solution  $\mathbf{f}$  is always an optimal cut, i.e. it has value MAX-CUT $(G)$ .

As we saw in Section 2.4, the random solution  $\mathbf{f}$  output by halfspace rounding has expected value at least  $\alpha_{\text{GW}} \cdot \text{SDP}_{\text{MC}}(G)$ . Hence, if Item 1 were not true, one of these solutions would have value strictly bigger than  $\alpha_{\text{GW}} \cdot \text{SDP}_{\text{MC}}(G)$ , contradicting  $G$  being an integrality gap. Likewise, if Item 2 were not true, there would exist a cut with value at least  $\alpha_{\text{GW}} \cdot \text{SDP}_{\text{MC}}(G)$ .

Now we use Item 1 to derive a constraint on  $f_{\text{SDP}}$ . Recall from Equation (1) that the value of  $\mathbf{f}$  is at least  $\alpha_{\text{GW}}$  times the value of  $f_{\text{OPT}}$  in the SDP for each edge  $(u, v)$ . In other words,

$$\mathbf{E}_{\mathbf{f}}\left[\frac{1}{2} - \frac{1}{2}\mathbf{f}(u)\mathbf{f}(v)\right] \geq \alpha_{\text{GW}} \cdot \left[\frac{1}{2} - \frac{1}{2}\langle f_{\text{SDP}}(u), f_{\text{SDP}}(v) \rangle\right], \quad (4)$$

This means that if Item 1 is true, Equation (4) must be satisfied with equality, for each edge  $(u, v)$ . To see what this implies, we first recall the proof of Equation (4). Letting  $\rho_{u,v}$  denote the inner product  $\rho_{u,v} = \langle f_{\text{SDP}}(u), f_{\text{SDP}}(v) \rangle$ , there is an exact formula for the left-hand side, namely

$$\mathbf{E}_{\mathbf{f}}\left[\frac{1}{2} - \frac{1}{2}\mathbf{f}(u)\mathbf{f}(v)\right] = \frac{\arccos(\rho_{u,v})}{\pi}.$$

See [GW95] for a proof of this fact. Then Equation (4) follows as a consequence of the statement

$$\min_{-1 \leq \rho \leq 1} \frac{\arccos(\rho)/\pi}{\frac{1}{2} - \frac{1}{2}\rho} = \alpha_{\text{GW}}.$$

There is in fact a *unique* minimizer of this expression, which we write as  $\rho_{\text{GW}} \approx -0.69$ . As a result, Equation (4) is satisfied with equality if and only if  $\rho_{u,v} = \rho_{\text{GW}}$ . Thus, Item 1 implies that  $\rho_{u,v} = \rho_{\text{GW}}$  for each edge  $(u, v)$ .

Motivated by this, Feige and Schechtman consider the  $n$ -dimensional sphere graph  $\mathcal{S}_{\rho_{\text{GW}}}^{n-1}$ . This is an infinite graph with vertex set  $S^{n-1}$  in which two vertices  $u, v \in S^{n-1}$  are connected whenever

$\langle u, v \rangle \approx \rho_{\text{GW}}$ . There is a natural SDP embedding of the sphere graph  $f_{\text{SDP}} : S^{n-1} \rightarrow S^{n-1}$ , in which  $f_{\text{SDP}}(u) = u$ , for each  $u \in S^{n-1}$ . It has value

$$\mathbf{E}_{(\mathbf{u}, \mathbf{v}) \sim E} \left[ \frac{1}{2} - \frac{1}{2} \langle f_{\text{SDP}}(\mathbf{u}), f_{\text{SDP}}(\mathbf{v}) \rangle \right] = \mathbf{E}_{(\mathbf{u}, \mathbf{v}) \sim E} \left[ \frac{1}{2} - \frac{1}{2} \langle \mathbf{u}, \mathbf{v} \rangle \right] \approx \frac{1}{2} - \frac{1}{2} \rho_{\text{GW}}.$$

Thus,  $\text{SDP}_{\text{MC}}(\mathcal{S}^{n-1}) \gtrsim \frac{1}{2} - \frac{1}{2} \rho_{\text{GW}}$ . In addition, this graph satisfies our two desiderata:

1. By construction,  $\langle f_{\text{SDP}}(u), f_{\text{SDP}}(v) \rangle \approx \rho_{\text{GW}}$  for each edge  $(u, v)$ . Thus, hyperplane rounding will produce a random cut  $\mathbf{f}$  with average value  $\approx \alpha_{\text{GW}} \cdot (\frac{1}{2} - \frac{1}{2} \rho_{\text{GW}})$ .
2. Each cut  $\mathbf{f}$  is of the form  $\mathbf{f}(u) = \text{sgn}(\langle \mathbf{z}, f_{\text{SDP}}(u) \rangle) = \text{sgn}(\langle \mathbf{z}, u \rangle)$ , where  $\mathbf{z}$  is a random Gaussian. By rotational symmetry, all of these cuts have the same value, and in particular they have the same value as the case when  $\mathbf{z} = e_1$ , i.e. the cut  $f_{\text{opt}}(u) = \text{sgn}(u_1)$ . The main technical argument of [FS02] is that this is in fact the optimal cut. In other words, for every function  $f : S^{n-1} \rightarrow \{-1, 1\}$ ,

$$\mathbf{E}_{(\mathbf{u}, \mathbf{v}) \sim E} [f(\mathbf{u})f(\mathbf{v})] \geq \mathbf{E}_{(\mathbf{u}, \mathbf{v}) \sim E} [f_{\text{opt}}(\mathbf{u})f_{\text{opt}}(\mathbf{v})]. \quad (5)$$

Together, these imply that  $\mathcal{S}_{\rho_{\text{GW}}}^{n-1}$  has integrality gap  $\alpha_{\text{GW}}$ .

**Moving to the Gaussian graph.** We will actually use a second, related construction of the integrality gap called the *Gaussian graph*, which will turn out to be more convenient to analyze in our case. It is defined as follows.

**Definition 2.23** ( $\rho$ -correlated Gaussian graph). Let  $n$  be a positive integer and  $-1 \leq \rho \leq 1$ . We define the  $\rho$ -correlated Gaussian graph to be the infinite graph  $\mathcal{G}_{\rho}^n$  with vertex set  $\mathbb{R}^n$  in which a random edge  $(\mathbf{u}, \mathbf{v})$  is distributed as two  $\rho$ -correlated Gaussian vectors.

For large  $n$ , the Gaussian graph, when scaled by a factor of  $\frac{1}{\sqrt{n}}$ , behaves like the sphere graph. For example, if  $\mathbf{u}$  is a random Gaussian vector, then  $\frac{1}{\sqrt{n}}\mathbf{u}$  is close to a unit vector with high probability. In addition, if  $(\mathbf{u}, \mathbf{v})$  is a random edge, then  $\langle \frac{1}{\sqrt{n}}\mathbf{u}, \frac{1}{\sqrt{n}}\mathbf{v} \rangle \approx \rho$  with high probability. As a result, an argument similar to above shows that it has an integrality gap of  $\alpha_{\text{GW}}$ .

The optimal SDP assignment is the function  $f_{\text{SDP}}(u) = u/\|u\|$ . Hyperplane rounding produces a cut of the form  $\mathbf{f}(u) = \text{sgn}(\langle \mathbf{z}, f_{\text{SDP}}(u) \rangle)$ , which is equivalent (up to rotation) to the assignment  $f_{\text{opt}}(u) = \text{sgn}(u_1)$ . The analogous statement to Equation (5) above that  $f_{\text{opt}}$  is optimal is a well-known result in Gaussian geometry known as *Borell's inequality* or *Borell's isoperimetric theorem*. There is a version of this theorem for positive and negative  $\rho$ , but we will only need to state the negative  $\rho$  case as the integrality gap only requires the  $\rho_{\text{GW}} \approx -0.69$  case.

**Theorem 2.24** (Borell's isoperimetric theorem, negative  $\rho$  case). *Let  $n$  be a positive integer and  $-1 \leq \rho \leq 0$ . Let  $f : \mathbb{R}^n \rightarrow \{-1, 1\}$ . In addition, let  $f_{\text{opt}} : \mathbb{R}^n \rightarrow \{-1, 1\}$  be defined by  $f_{\text{opt}}(x) = \text{sgn}(x_1)$ . Then*

$$\mathbf{E}_{\mathbf{u} \sim_{\rho} \mathbf{v}} [f(\mathbf{u})f(\mathbf{v})] \geq \mathbf{E}_{\mathbf{u} \sim_{\rho} \mathbf{v}} [f_{\text{opt}}(\mathbf{u})f_{\text{opt}}(\mathbf{v})].$$

**Integrality gap for QUANTUM MAX-CUT.** The integrality gaps we design for the product state and QUANTUM MAX-CUT SDPs are constructed along very similar lines to the MAX-CUT integrality gaps. In both cases, our goal is to show that projection rounding is the optimal rounding algorithm, which motivates us to study the conditions in which projection rounding performs worst. As

we saw in [Section 2.4](#), projection rounding applied to the product state and QUANTUM MAX-CUT SDPs yields a random function  $\mathbf{f} : V \rightarrow S^2$  satisfying the following conditions, for each edge  $(u, v)$ :

$$\mathbf{E}_{\mathbf{f}}\left[\frac{1}{4} - \frac{1}{4}\langle \mathbf{f}(u), \mathbf{f}(v) \rangle\right] \geq \alpha_{\text{BOV}} \cdot \left[\frac{1}{4} - \frac{1}{4}\langle f_{\text{SDP}}(u), f_{\text{SDP}}(v) \rangle\right], \quad (6)$$

$$\mathbf{E}_{\mathbf{f}}\left[\frac{1}{4} - \frac{3}{4}\langle \mathbf{f}(u), \mathbf{f}(v) \rangle\right] \geq \alpha_{\text{GP}} \cdot \left[\frac{1}{4} - \frac{3}{4}\langle f_{\text{SDP}}(u), f_{\text{SDP}}(v) \rangle\right], \quad (7)$$

respectively. As in the case of MAX-CUT, the left-hand sides of these equations depend only on the inner product  $\rho_{u,v} = \langle f_{\text{SDP}}(u), f_{\text{SDP}}(v) \rangle$ . This was shown by [\[BdOFV10\]](#), who gave the following exact expression for this quantity:

$$\mathbf{E}_{\mathbf{f}}\langle \mathbf{f}(u), \mathbf{f}(v) \rangle = \frac{2}{3} \left( \frac{\Gamma(2)}{\Gamma(3/2)} \right)^2 \rho_{u,v} \cdot {}_2F_1(1/2, 1/2; 5/2; \rho_{u,v}^2), \quad (8)$$

where  ${}_2F_1(\cdot, \cdot; \cdot; \cdot)$  is the Gaussian hypergeometric function. Thus, one can compute the approximation ratios  $\alpha_{\text{BOV}}$  and  $\alpha_{\text{GP}}$  by finding the “worst case” values of  $\rho_{u,v}$ . For [Equation \(6\)](#), this is  $\rho_{\text{BOV}} \approx -0.584$ ; for [Equation \(7\)](#), this is  $\rho_{\text{GP}} \approx -0.97$ .

This suggests finding a graph for the product state SDP and the QUANTUM MAX-CUT SDP in which  $\rho_{u,v} = \rho_{\text{BOV}}$  and  $\rho_{u,v} = \rho_{\text{GP}}$  for each edge  $(u, v)$ , respectively. Again, we will use the  $\rho$ -correlated Gaussian graph, where again the optimum SDP assignment is the function  $f_{\text{SDP}}(u) = u/\|u\|$ . Projection rounding produces a solution of the form  $\mathbf{f}(u) = \mathbf{Z}u/\|\mathbf{Z}u\|$ , where  $\mathbf{Z}$  is a random  $3 \times n$  Gaussian matrix. When  $n$  is large, this is roughly equivalent to projecting  $u$  onto a random 3-dimensional subspace, in which case it is equivalent (up to rotation) to  $f_{\text{opt}}(u) = (u_1, u_2, u_3)/\|(u_1, u_2, u_3)\|$ .

We must now show that  $f_{\text{opt}}$  is indeed the optimal solution. In the case of the product state value, we must show that it is the best among all product states; equivalently, among all functions  $f : \mathbb{R}^n \rightarrow S^2$ . On the other hand, for the case of the ground state energy, we must show that it is the best among all quantum states, which need not be product. Fortunately, the Gaussian graph is of high (in fact, infinite) degree, and so by [\[BH16\]](#) the optimal state is a product state. The optimality of  $f_{\text{opt}}$  then follows from our conjectured vector-valued analogue of Borell’s isoperimetric theorem ([Conjecture 1.1](#), in the case  $k = 3$ ).

### 3 Conclusion and open questions

In this work, we have made progress on understanding the approximability of QUANTUM MAX-CUT. However, there are many interesting questions which remain open, such as finding the optimal approximation ratio. We list these below.

1. Most obviously, is the vector-valued Borell’s inequality true?
2. Does there exist an algorithmic gap instance for the GP algorithm with algorithmic gap  $\alpha_{\text{GP}}$ ? We believe there is but were unable to find one. The key difficulty seems to be that an algorithmic gap instance should be a low-degree graph with an entangled maximum energy state. Otherwise, the optimizing state would be close to a product state, and in this case the GP algorithm matches the  $\alpha_{\text{BOV}}$  approximation ratio of the BOV algorithm. But finding explicit examples of families of QUANTUM MAX-CUT instances with entangled maximum energy states for which we can even compute their optimum value is a difficult problem, and only a few such examples are known (see [\[Fra17, Men13\]](#)).

3. Can we perform an optimal analysis of the level-4 ncSoS relaxation of QUANTUM MAX-CUT? This would involve designing a rounding algorithm and finding an integrality gap which matches its performance, as well as identifying the optimal ansatz, which would need to be more powerful than product states. Inspired by the work of [AGM20], Parekh and Thompson [PT21a] have considered tensor products of one- and two-qubit states, but it is unclear whether this is the optimal ansatz.
4. Could the level-4 ncSoS relaxation actually be optimal for QUANTUM MAX-CUT?
5. The BOV algorithm is essentially the optimal algorithm for QUANTUM MAX-CUT on high-degree graphs. What about low-degree graphs? Can we design improved approximation algorithms in this case as well? Recently, for example, Anshu, Gosset, Morenz Korol and Soleimanifar have demonstrated a way to improve the objective value of a given product state assuming the graph is low degree [AGKS21].
6. Are there *quantum* approximation algorithms for QUANTUM MAX-CUT whose approximation ratios we can analyze?
7. Can we prove a hardness of approximation result for QUANTUM MAX-CUT which improves upon our [Theorem 2.22](#)? This would involve showing a reduction from the Unique Games problem which outputs low-degree instances of QUANTUM MAX-CUT. This is because high-degree instances have maximum energy states which can be approximated by product states, and so the BOV algorithm produces an  $\alpha_{\text{BOV}}$ -approximation in this case. However, traditional Unique Games reductions typically produce high-degree instances, and so it seems like new techniques might be needed. This is related to our difficulty in producing algorithmic gap instances with entangled maximum energy states, and so designing algorithmic gap instances might be a good first step.
8. Is it QMA-hard to approximate QUANTUM MAX-CUT? Proving this unconditionally would also prove the quantum PCP conjecture, putting it beyond the range of current techniques. But it might be possible to show this *assuming* the quantum PCP conjecture is true.
9. Since the Heisenberg model has an analytic solution only for certain graphs (see for example the “Bethe Ansatz” [Bet31]), physicists use a set of heuristic algorithms for approximating the ground state of Hamiltonians from the Heisenberg model [Blo30, AA88, FBB79]. Can we find a rigorous theoretical justification to support the success of these heuristics in practice?
10. How well do the techniques used to design approximation algorithms for QUANTUM MAX-CUT carry over to other families of Hamiltonians? One natural family is the set of Hamiltonians in which each local term is a projective matrix. An  $\alpha$ -approximation for this case gives an  $\alpha$ -approximation for any Hamiltonian with positive-semidefinite local terms. This was considered in [PT21b] where it was shown that a rounding algorithm akin to that used in [GP19] also applies to this case. This was followed up by an approach in [PT22] that applies to more general ansatzes. However, the analysis in these works is not tight so proving the actual performance of the algorithm is an open question.

## Acknowledgments

We would like to thank Anurag Anshu, Srinivasan Arunachalam, and Penghui Yao for their substantial contributions to this paper. We would also like to thank Steve Heilman for pointing out a bug in a previous draft of the paper.

## Part II

# A vector-valued Borell's inequality

In this part, we discuss our main conjecture, [Conjecture 1.1](#), in detail and prove various partial results.

We are primarily concerned with vector-valued functions,  $f : \mathbb{R}^n \rightarrow B^k$ , where  $B^k$  is the unit ball in  $\mathbb{R}^k$ . Writing  $(f_1, \dots, f_k) = f$  for the coordinate functions of  $f$ , note that if  $f : \mathbb{R}^n \rightarrow B^k$  then each  $f_i$  takes values in  $[-1, 1]$  and is therefore square-integrable under the standard Gaussian measure  $\gamma$  (i.e.  $f_i \in L^2(\gamma)$ ). Generally if a function  $f$  is square integrable on a space  $A$  with measure  $\eta$  we will denote this as  $f \in L^2(A, \eta)$  or as  $f \in L^2(\eta)$  if the space is clear from context. We say that a vector-valued function  $f : \mathbb{R}^n \rightarrow \mathbb{R}^k$  is square-integrable (i.e. belongs to  $L^2(\gamma)$ ) if each  $f_i \in L^2(\gamma)$ .

**Gaussian variables.** We will make use of standard notations for multivariate Gaussian random variables: If  $\mu \in \mathbb{R}^n$  and  $\Sigma \in \mathbb{R}^{n \times n} \succeq 0$  then  $\mathbf{x} \sim \mathcal{N}(\mu, \Sigma)$  if  $\mathbf{x}$  is a multivariate normal satisfying  $\mathbf{E}[\mathbf{x}_i] = \mu_i$  and  $\mathbf{E}[(\mathbf{x}_i - \mu_i)(\mathbf{x}_j - \mu_j)] = \Sigma_{i,j}$ .

**Definition 3.1** ( $\rho$ -correlated Gaussians). Let  $\mathbf{x}$  and  $\mathbf{y}$  be Gaussian random variables taking values in  $\mathbb{R}^n$  and  $-1 \leq \rho \leq 1$  be a parameter. The variables  $\mathbf{x}$  and  $\mathbf{y}$  are  $\rho$ -correlated Gaussians, denoted  $\mathbf{x} \sim_\rho \mathbf{y}$ , if

$$(\mathbf{x}, \mathbf{y}) \sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} I_n & \rho I_n \\ \rho I_n & I_n \end{bmatrix}\right),$$

where  $\mathbf{0}$  is a vector of zeros of the appropriate size and  $I_n$  is the  $n \times n$  identity matrix.

**Definition 3.2** (Noise operator for vector-valued functions). Let  $f : \mathbb{R}^n \rightarrow \mathbb{R}^k$  be a square-integrable vector-valued function and  $-1 \leq \rho \leq 1$  be a parameter. Then the *Gaussian noise operator*,  $U_\rho$  is defined as

$$U_\rho f(x) = \mathbf{E}_{\mathbf{x} \sim_\rho \mathbf{y}} [f(\mathbf{y}) \mid \mathbf{x} = x].$$

We observe that if  $f = (f_1, \dots, f_k)$  then  $U_\rho f = (U_\rho f_1, \dots, U_\rho f_k)$ .

**Definition 3.3** (Noise stability for vector-valued functions). Let  $f : \mathbb{R}^n \rightarrow \mathbb{R}^k$  be a square-integrable vector-valued function and  $-1 \leq \rho \leq 1$  be a parameter. Then the *noise stability* of  $f$  at  $\rho$  is

$$\mathbf{Stab}_\rho[f] = \mathbf{E}_{\mathbf{x} \sim_\rho \mathbf{y}} \langle f(\mathbf{x}), f(\mathbf{y}) \rangle = \mathbf{E}_{\mathbf{x} \sim \mathcal{N}(0,1)^n} \langle f(\mathbf{x}), U_\rho f(\mathbf{x}) \rangle.$$

The following vector-valued version of Borell's inequality (for negative  $\rho$ ) is our main conjecture.

**Conjecture 3.4.** Let  $1 \leq k \leq n$  be positive integers and let  $-1 \leq \rho \leq 0$ . For all measurable functions  $f : \mathbb{R}^n \rightarrow B^k$ ,

$$\mathbf{Stab}_\rho[f] \geq \mathbf{Stab}_\rho[f_{\text{opt}}]$$

where  $f_{\text{opt}}(x) = x_{\leq k} / \|x_{\leq k}\|$  and  $x_{\leq k} = (x_1, \dots, x_k)$ .

Moreover, this optimizer is unique up to orthogonal transformations: for any  $f : \mathbb{R}^n \rightarrow B^k$  for which  $\mathbf{Stab}_\rho[f] = \mathbf{Stab}_\rho[f_{\text{opt}}]$ , there is an orthogonal  $n \times n$  matrix  $M$  such that  $f(x) = f_{\text{opt}}(Mx)$  almost surely.

Note that this recovers Theorem 2.24 in the case that  $k = 1$ , and also that the assumption  $f : \mathbb{R}^n \rightarrow B^k$  could be replaced by  $f : \mathbb{R}^n \rightarrow S^{k-1}$  without having much effect: since the optimal function  $f : \mathbb{R}^n \rightarrow B^k$  actually takes values in  $S^{k-1}$ , the optimal value is the same if we restrict the optimization to functions  $f : \mathbb{R}^n \rightarrow S^{k-1}$ .

Although our main application of Conjecture 3.4 involves negative  $\rho$ , we will also consider a version with positive  $\rho$ . In this case, one needs to add the assumption that  $\mathbf{E}[f] = 0$ ; without this constraint, constant functions like  $f(x) = (1, \dots, 0)$  are maximizers.

**Conjecture 3.5.** *Let  $1 \leq k \leq n$  be positive integers and let  $0 \leq \rho \leq 1$ . For all measurable functions  $f : \mathbb{R}^n \rightarrow B^k$  satisfying  $\mathbf{E}[f(\mathbf{x})] = 0$ ,*

$$\mathbf{Stab}_\rho[f] \leq \mathbf{Stab}_\rho[f_{\text{opt}}]$$

where  $f_{\text{opt}}(x) = x_{\leq k} / \|x_{\leq k}\|$  and  $x_{\leq k} = (x_1, \dots, x_k)$ .

Moreover, this optimizer is unique up to orthogonal transformations: for any  $f : \mathbb{R}^n \rightarrow B^k$  for which  $\mathbf{Stab}_\rho[f] = \mathbf{Stab}_\rho[f_{\text{opt}}]$ , there is an orthogonal  $n \times n$  matrix  $M$  such that  $f(x) = f_{\text{opt}}(Mx)$  almost surely.

## Proposed strategy

There are a few notable difficulties in establishing Conjecture 3.4, compared with the scalar-valued case. Recall in particular that Borell’s theorem is also known when the expectation of  $f$  is constrained: among all functions  $f : \mathbb{R}^n \rightarrow [-1, 1]$  with  $\mathbf{E}[f] = v \in [-1, 1]$ , the noise stability is minimized (for negative  $\rho$ ) by a linear threshold function with the appropriate expectation, i.e. a function of the form  $f_{\text{opt}}(x) = \text{sgn}(\langle a, x \rangle + b)$ . The formulation in Theorem 2.24 is then recovered just by noting that as a function of the constraint  $v \in [-1, 1]$ , the optimal noise stability is minimized when  $v = 0$ . In the vector-valued situation  $k \geq 2$ , we do not solve a constrained version of the problem. In fact, we do not even know of a good guess for the optimal stability among functions with  $\mathbf{E}[f] = v \neq 0$ . The guess that comes from naively extrapolating the  $k = 1$  solution,  $f(x) = (x_{\leq k} - a) / \|x_{\leq k} - a\|$  appears not to be optimal. This inconvenient fact rules out certain proof techniques that work for the scalar-valued case—specifically, recent approaches using  $\rho$ -convexity [MN15] and stochastic calculus [Eld15]—because if those methods had worked, they would have also shown optimality of  $f(x) = (x_{\leq k} - a) / \|x_{\leq k} - a\|$  in the constrained version. On the other hand, we also do not know how to exploit the older symmetrization-based approaches [Bor85] because of the difficulty of symmetrizing vector-valued functions.

We will propose a three-step strategy for Conjectures 3.4 and 3.5. The reason that we have been unable to complete the proof is that one of the three steps is only proven for positive  $\rho$ , while another step is only proven for negative  $\rho$ .

The first step is to consider noise stability for functions  $f : S^{n-1} \rightarrow S^{n-1}$  and we prove that (in both positive- $\rho$  and negative- $\rho$  cases) the function  $f(x) = x$  has optimal noise stability. The argument here is purely spectral: having shown that the eigenvectors of our noise operator (modified appropriately to live on the sphere) are spherical harmonics, we expand the function  $f$  in the basis of spherical harmonics and show that the optimal thing to do is to put all the “weight” on “level-1” coefficients. One interesting feature of this argument is that it doesn’t require  $f$  to take values in  $S^{n-1}$ : we show that  $f(x) = x$  has optimal noise stability among all functions  $f : S^{n-1} \rightarrow \mathbb{R}^n$  satisfying  $\mathbf{E}[\|f(\mathbf{x})\|^2] = 1$ . This step of the proof also works for  $0 < \rho \leq 1$ : in this range, the noise stability is *maximized*, among functions with  $\mathbf{E}[f] = 0$ , by  $f(x) = x$ .

The second step of the proof is to consider functions  $f : \mathbb{R}^n \rightarrow S^{n-1}$ . We do this by decomposing  $\mathbb{R}^n$  into radial “shells” and applying the spherical argument on each shell. This step requires  $-1 \leq \rho \leq 0$ .

The final step, which applies only to  $0 \leq \rho \leq 1$  is a kind of dimension reduction. Specifically, we show that if  $f : \mathbb{R}^n \rightarrow S^{k-1}$  has optimal stability then  $f$  is “essentially  $k$ -dimensional” in the sense that up to a change of coordinates there is a function  $g : \mathbb{R}^k \rightarrow S^{k-1}$  such that  $f(x) = g(x_1, \dots, x_k)$ . This essentially reduces the problem to the case of functions  $f : \mathbb{R}^k \rightarrow S^{k-1}$ , which was already handled in the second step. This step of the proof uses tools from the calculus of variations. Essentially, we show that if  $f : \mathbb{R}^n \rightarrow S^{k-1}$  is not essentially  $k$ -dimensional then it can be modified in a way that improves the noise stability. Arguments of this kind go back to McGonagle and Ross [MR15] in the setting of the Gaussian isoperimetric problem. They were developed in the vector-valued (but still isoperimetric) setting by Milman and Neeman [MN18], and then applied to noise stability by Heilman and Tarter [HT20].

Note that by combining the first two steps (which both apply for  $-1 \leq \rho \leq 0$ ) we can show that [Conjecture 3.4](#) holds in the case  $k = n$ .

## 4 The spherical case

Here we consider the case of  $f : S^{n-1} \rightarrow B^n$ . Since we want to work with Gaussian noise on  $\mathbb{R}^n$ , this shell decomposition imposes a specific noise operator on  $S^{n-1}$ . In this section, we will work with a more general noise operator that includes the ones we will need in later sections.

The “uniform” measure on the sphere will be denoted  $\omega$ : this is the unique rotationally invariant probability measure on  $S^{n-1}$ . If  $\mathbf{u} \sim \omega$ , we will write  $\tilde{\omega}$  for the distribution of  $\mathbf{u}_1$  (which is the same as the distribution of  $\langle v, \mathbf{u} \rangle$  for any  $v \in S^{n-1}$ ). Note that  $\tilde{\omega}$  is a density on  $[-1, 1]$ , with density

$$d\tilde{\omega}(t) = \frac{1}{Z_n} (1 - t^2)^{\frac{n-3}{2}} dt, \quad (9)$$

where  $Z_n$  is a normalizing constant. The proof of Lemma 4.17 in [FE12] depicts how  $d\tilde{\omega}$  arises from  $d\omega$ .

For a function  $g : [-1, 1] \rightarrow \mathbb{R}$  satisfying

$$\int_{-1}^1 |g(t)| d\tilde{\omega}(t) < \infty, \quad (10)$$

define the operator  $U_g$ , acting on functions  $f : S^{n-1} \rightarrow \mathbb{R}^k$  by

$$U_g f(u) = \int_{S^{n-1}} g(\langle u, v \rangle) f(v) d\omega(v). \quad (11)$$

To make the definition fully rigorous, note that the integrability condition (10) implies that if  $f$  is bounded then  $U_g f$  is defined pointwise. Then Jensen’s inequality implies that  $\|U_g f\|_{L^p(\omega)} \leq C \|f\|_{L^p(\omega)}$  for every bounded  $f$  – with  $C$  being the left hand side of (10) – and since bounded functions are dense in  $L^p(\omega)$  it follows that  $U_g$  can be uniquely extended to an operator  $L^2(\omega) \rightarrow L^2(\omega)$ .

We will be interested in non-negative  $g$ , and it might also be convenient to imagine  $g$  as integrating to 1; i.e., with  $\int_{-1}^1 g(t) d\tilde{\omega}(t) = 1$ . In this case  $U_g f(u)$  is an average of values of  $f$ , much like our Gaussian noise operator  $U_\rho$ .

The main result of this section is that if  $g$  is monotonic then the function  $f(x) = x$  is optimally stable.

**Theorem 4.1.** *If  $g : [-1, 1] \rightarrow [0, \infty)$  satisfies (10) and is non-decreasing then for every  $f : S^{n-1} \rightarrow \mathbb{R}^n$  with  $\mathbf{E}_{\mathbf{u} \sim \omega}[\|f(\mathbf{u})\|^2] = 1$ ,*

$$\mathbf{E}_{\mathbf{u} \sim \omega} \langle f(\mathbf{u}), U_g f(\mathbf{u}) \rangle \geq \mathbf{E}_{\mathbf{u} \sim \omega} \langle f_{\text{opt}}(\mathbf{u}), U_g f_{\text{opt}}(\mathbf{u}) \rangle,$$

where  $f_{\text{opt}}(\mathbf{u}) = u$ .

*On the other hand, if  $g$  is non-increasing then for every  $f : S^{n-1} \rightarrow \mathbb{R}^n$  with  $\mathbf{E}_{\mathbf{u} \sim \omega}[\|f(\mathbf{u})\|^2] = 1$  and  $\mathbf{E}_{\mathbf{u} \sim \omega}[f] = 0$ ,*

$$\mathbf{E}_{\mathbf{u} \sim \omega} \langle f(\mathbf{u}), U_g f(\mathbf{u}) \rangle \leq \mathbf{E}_{\mathbf{u} \sim \omega} \langle f_{\text{opt}}(\mathbf{u}), U_g f_{\text{opt}}(\mathbf{u}) \rangle.$$

## 4.1 Spherical harmonics

We prove Theorem 4.1 by decomposing each coordinate of the vector-valued function  $f$  into spherical harmonics. We suggest [FE12, DX13] as introductory references for spherical harmonics.

**Definition 4.2** (Spherical harmonics). *A homogeneous polynomial of degree  $d$  is a function  $p : \mathbb{R}^n \rightarrow \mathbb{R}$  expressible as a linear combination of degree- $d$  monomials. We define the sets  $\{\mathcal{H}_d : d = 0, 1, \dots\}$  by first setting  $\mathcal{H}_0$  to be the set of constant functions  $S^{n-1} \rightarrow \mathbb{R}$ , and then inductively defining  $\mathcal{H}_d$  to be the functions  $S^{n-1} \rightarrow \mathbb{R}$  that can be represented as homogeneous polynomials of degree  $d$ , and which are orthogonal to  $\bigoplus_{k=0}^{d-1} \mathcal{H}_k$ . The elements of  $\mathcal{H}_d$  are called degree- $d$  spherical harmonics.*

One subtlety (that will not be particularly important for us) is that distinct polynomials may give rise to the same function  $S^{n-1} \rightarrow \mathbb{R}$ ; for example, the constant function  $f(u) = 1$  can be written both as the constant polynomial 1 and as the degree-2 polynomial  $u_1^2 + \dots + u_n^2$ , which evaluates to 1 on the sphere. The name *spherical harmonics* comes from the fact that  $\mathcal{H}_d$  can be equivalently defined as the set of homogeneous degree- $d$  polynomials  $p$  that are *harmonic* in the sense that  $\sum_{i=1}^n \frac{\partial^2}{\partial x_i^2} p(x) = 0$ .

The first important thing about the spaces  $\mathcal{H}_d$  is that they form an orthogonal decomposition of  $L^2(S^{n-1}, \omega)$ : the  $\mathcal{H}_d$  are orthogonal in the sense that if  $f \in \mathcal{H}_d$  and  $g \in \mathcal{H}_{d'}$  for  $d \neq d'$  then  $\mathbf{E}_{\mathbf{u} \sim \omega}[f(\mathbf{u})g(\mathbf{u})] = 0$ ; they decompose  $L^2(S^{n-1}, \omega)$  in the sense that if  $H_d : L^2(S^{n-1}, \omega) \rightarrow \mathcal{H}_d$  is the orthogonal projection operator then  $f = \sum_{d \geq 0} H_d f$  for every  $f \in L^2(S^{n-1}, \omega)$  [DX13]. In this sense, the decomposition into spherical harmonics is analogous to the decomposition of a Boolean function into Fourier levels [O'D14], or the decomposition of a function in  $L^2(\mathbb{R}^n, \gamma)$  into Hermite levels [O'D14].

Then second important thing about the spaces  $\mathcal{H}_d$  is the Funk-Hecke formula, which essentially says that noise operators like our  $U_g$  act diagonally on spherical harmonics (much like the usual Boolean and Gaussian noise operators act diagonally on the Fourier and Hermite bases respectively).

**Theorem 4.3** (Funk-Hecke formula [Hec17]; see [DX13], Theorem 2.9). *For any  $g : [-1, 1] \rightarrow \mathbb{R}$  satisfying (10), let  $U_g$  be defined as in (11) (with  $k = 1$ ). Then for every  $d \in \{0, 1, \dots\}$  there exists a  $\lambda_d$  such that for every  $f \in \mathcal{H}_d$  and every  $u \in S^{n-1}$ ,*

$$U_g f(u) = \lambda_d f(u).$$

*In other words, spherical harmonics are the eigenvalues of  $U_g$  and the eigenvalues depend only on the degree  $d$ .*

One particularly nice feature of the Funk-Hecke formula is that because the eigenvalues depend only on the degree  $d$ , we can compute  $\lambda_d$  by choosing the most convenient  $f \in \mathcal{H}_d$  and  $u \in S^{n-1}$ . This leads us to the Gegenbauer polynomials, a family of univariate polynomials that capture the *zonal* spherical harmonics (see [DX13], Theorem 2.6), those depending only on one direction.

**Definition 4.4** (Gegenbauer polynomials). Let  $\alpha > -\frac{1}{2}$  and  $d$  be a nonnegative integer. The *Gegenbauer polynomial* with Gegenbauer index  $\alpha$  and degree  $d$  is a univariate, real polynomial denoted  $C_d^{(\alpha)}$ . The Gegenbauer polynomials correspond to the zonal spherical harmonics of interest to us when  $\alpha = \frac{n-2}{2}$  and  $n \geq 3$ , and we will henceforth make these assumptions on  $\alpha$  and  $n$ .

Gegenbauer polynomials may be defined recursively, using generating functions, in terms of the Gaussian hypergeometric function, or as special cases of other polynomials (see [AS72], Chapter 22). We will only need a few properties of them.

**Proposition 4.5** (Properties of the Gegenbauer polynomials).

1. ([AS72], 22.4.2) We have the following explicit formulas for low-degrees:

$$C_0^{(\alpha)}(t) = 1, \text{ and } C_1^{(\alpha)}(t) = 2\alpha t.$$

2. ([AS72], 22.14.2, [Sze39], Theorem 7.4.1) For  $-1 \leq t \leq 1$  and  $\alpha > 0$ ,

$$|C_d^{(\alpha)}(t)| \leq C_d^{(\alpha)}(1) = \frac{(2\alpha)_d}{d!},$$

with a strict inequality if  $d \geq 1$  and  $-1 < t < 1$ .

3. ([AS72], 22.13.2) For each integer  $d \geq 0$ , define the quantity

$$\text{ratio}_d(t) = \frac{C_d^{(\alpha)}(t)}{C_d^{(\alpha)}(1)} \cdot (1-t^2)^{\alpha-\frac{1}{2}}.$$

Then

$$\int \text{ratio}_d(t) dt = -\frac{2(1-t^2)^{\alpha+\frac{1}{2}}}{d(d+2\alpha)} \cdot \frac{C_{d-1}^{(\alpha+1)}(t)}{C_d^{(\alpha)}(1)}.$$

4. ([DX13], Theorem 2.6) For each integer  $d \geq 0$ , the function  $S^{n-1} \rightarrow \mathbb{R}$  defined by  $u \mapsto C_d^{(\alpha)}(u_1)$  belongs to  $\mathcal{H}_d$ .

## 4.2 Eigenvalues of the noise operator

The last property of Proposition 4.5 shows the relevance of Gegenbauer polynomials to the computation of  $\lambda_d$ : letting  $h(u) = C_d^{(\alpha)}(u_1)$ , we have

$$\lambda_d = \frac{U_g h(e_1)}{h(e_1)} = \frac{\mathbf{E}_{\mathbf{u} \sim \omega}[h(\mathbf{u})g(\langle \mathbf{u}, e_1 \rangle)]}{h(e_1)} = \mathbf{E}_{t \sim \tilde{\omega}} \left[ \frac{C_d^{(\alpha)}(t)}{C_d^{(\alpha)}(1)} g(t) \right].$$

Recalling the formula (9) for the density of  $\tilde{\omega}$ , we conclude:

**Corollary 4.6** (Eigenvalues of  $U_g$ ).

$$\lambda_d = \frac{1}{Z_n} \int_{-1}^1 \text{ratio}_d(t) g(t) dt.$$

**Remark 4.7** (The 2-dimensional case). As noted in [Definition 4.4](#), we have thus far required  $n \geq 3$ . When  $n = 2$ , spherical harmonics reduce to Fourier series. If  $t = \cos(\theta)$ , we may define  $C_d(t) := \cos(d\theta) = T_d(t)$ , where the latter is the degree- $d$  Chebyshev polynomial of the first kind. Analogues of the properties in [Proposition 4.5](#) hold in this case, allowing us to recover our results for  $n = 2$ .

Since we are interested in comparing  $\lambda_d$  as  $d$  varies, the factor  $\frac{1}{Z_n}$  is unimportant for us. The key bound that we will need essentially amounts to considering the case of the indicator function  $g = 1_{[-1,t]}$ , defined to be 1 on  $[-1, t]$  and 0 elsewhere.

**Lemma 4.8** (Key Gegenbauer lemma). *For each integer  $d \geq 0$ , define the quantity*

$$\nu_d(t) = \int_{-1}^t \text{ratio}_d(w) dw,$$

where  $-1 < t < 1$ . Then  $|\nu_d(t)| < -\nu_1(t)$  for  $d \geq 1$ . In addition,  $\nu_1(t) \leq 0$ .

*Proof.* By [Item 3](#) of the Gegenbauer properties,

$$\nu_d(t) = -\frac{2(1-t^2)^{\alpha+\frac{1}{2}}\alpha}{d(d+2\alpha)} \cdot \frac{C_{d-1}^{(\alpha+1)}(t)}{C_d^{(\alpha)}(1)}.$$

Using [Item 1](#), we can simplify the  $d = 1$  case as follows:

$$\nu_1(t) = -\frac{2(1-t^2)^{\alpha+\frac{1}{2}}\alpha}{(1+2\alpha)} \cdot \frac{C_0^{(\alpha+1)}(t)}{C_1^{(\alpha)}(1)} = -\frac{(1-t^2)^{\alpha+\frac{1}{2}}}{(1+2\alpha)}.$$

This is clearly  $\leq 0$ , as  $\alpha > -\frac{1}{2}$ . Finally, by [Item 2](#), we have the bound

$$\begin{aligned} |\nu_d(t)| &= \frac{2(1-t^2)^{\alpha+\frac{1}{2}}\alpha}{d(d+2\alpha)} \cdot \frac{|C_{d-1}^{(\alpha+1)}(t)|}{C_d^{(\alpha)}(1)} \\ &< \frac{2(1-t^2)^{\alpha+\frac{1}{2}}\alpha}{d(d+2\alpha)} \cdot \frac{C_{d-1}^{(\alpha+1)}(1)}{C_d^{(\alpha)}(1)} \\ &= \frac{2(1-t^2)^{\alpha+\frac{1}{2}}\alpha}{d(d+2\alpha)} \cdot \frac{(2\alpha+2)_{d-1} \cdot d!}{(d-1)! \cdot (2\alpha)_d} \\ &= \frac{(1-t^2)^{\alpha+\frac{1}{2}}}{(2\alpha+1)} = -\nu_1(t). \end{aligned}$$

This completes the proof. □

Once we have considered the case of  $g = 1_{[-1,t]}$ , all other monotonic cases follow simply by expressing monotonic functions as linear combinations of indicator functions:

**Corollary 4.9.** *In the setting of [Theorem 4.1](#), if  $g$  is non-increasing then  $\lambda_1 \leq 0$  and  $|\lambda_d| < -\lambda_1$  for all  $d \geq 2$ . On the other hand, if  $g$  is non-decreasing then  $\lambda_1 \geq 0$  and  $|\lambda_d| < \lambda_1$  for all  $d \geq 2$ .*

*Proof.* If  $g : [-1, 1] \rightarrow [0, \infty)$  is non-increasing then it can be written as a linear combination of non-increasing indicator functions: there is a measure  $\mu$  on  $[-1, 1]$  such that

$$g(t) = \int_{-1}^1 1_{[-1, s]}(t) d\mu(s).$$

By [Corollary 4.6](#) and Fubini's theorem,

$$\lambda_d = \frac{1}{Z_n} \int_{-1}^1 \int_{-1}^1 \text{ratio}_d(t) 1_{[-1, s]}(t) d\mu(s) dt = \frac{1}{Z_n} \int_{-1}^1 \nu_d(s) d\mu(s),$$

where  $\nu_d$  is defined as in [Lemma 4.8](#). The claim then follows from [Lemma 4.8](#).

For the case of non-decreasing  $g$ , note that  $\nu_d(1) = 0$  for all  $d \geq 1$ , for example because of [Item 4](#) and the fact that spherical harmonics of degree  $d \geq 1$  are orthogonal to constant functions (which are the spherical harmonics of degree 0). Then we represent  $g$  as a linear combination of non-decreasing indicator functions by choosing  $\mu$  such that

$$g(t) = \int_{-1}^1 1_{[s, 1]}(t) d\mu(s) = \int_{-1}^1 1 - 1_{[-1, s]}(t) d\mu(s);$$

and finally, we have

$$\lambda_d = \frac{1}{Z_n} \int_{-1}^1 \int_{-1}^1 \text{ratio}_d(t) (1 - 1_{[-1, s]}(t)) d\mu(s) dt = -\frac{1}{Z_n} \int_{-1}^1 \nu_d(s) d\mu(s),$$

and we conclude as before using [Lemma 4.8](#).  $\square$

Finally, [Theorem 4.1](#) follows from [Corollary 4.9](#) simply by decomposing the function  $f$  in spherical harmonics.

*Proof of [Theorem 4.1](#).* Assume first that  $g$  is non-increasing, and choose  $f : S^{n-1} \rightarrow \mathbb{R}^n$  with  $\mathbf{E}[\|f\|^2] = 1$ . Recall that  $H_d : L^2(S^{n-1}) \rightarrow \mathcal{H}_d$  is the orthogonal projection onto degree- $d$  spherical harmonics. We extend  $H_d$  to act on vector-valued functions coordinate-wise, so that if  $f_1, \dots, f_n$  are the coordinate functions of  $f$  then  $H_d f = (H_d f_1, \dots, H_d f_n)$ . We also have  $U_g f = (U_g f_1, \dots, U_g f_n)$ . Recall that  $f = \sum_{d \geq 0} H_d f$ ; then [Theorem 4.3](#) implies that

$$\mathbf{E}_{\mathbf{u} \sim \omega} \langle f(\mathbf{u}), U_g f(\mathbf{u}) \rangle = \mathbf{E}_{\mathbf{u} \sim \omega} \left\langle \sum_{d \geq 0} (H_d f)(\mathbf{u}), \sum_{d' \geq 0} (U_g H_{d'} f)(\mathbf{u}) \right\rangle = \sum_{d \geq 0} \lambda_d \mathbf{E}_{\mathbf{u} \sim \omega} [\|H_d f(\mathbf{u})\|^2],$$

where the cross-terms with  $d \neq d'$  vanished because of the orthogonality of spherical harmonics.

On the other hand, the orthogonality of the decomposition  $f = \sum_{d \geq 0} H_d f$  implies that

$$\sum_{d \geq 0} \lambda_d \mathbf{E}_{\mathbf{u} \sim \omega} [\|H_d f(\mathbf{u})\|^2] = \mathbf{E}_{\mathbf{u} \sim \omega} [\|f(\mathbf{u})\|^2] = 1.$$

Since (by [Corollary 4.9](#) and the fact that  $\lambda_0 = \mathbf{E}_{\mathbf{t} \sim \tilde{\omega}}[g(\mathbf{t})] \geq 0$ )  $\lambda_1$  is the most-negative of all eigenvalues,

$$\mathbf{E}_{\mathbf{u} \sim \omega} \langle f(\mathbf{u}), U_g f(\mathbf{u}) \rangle \geq \lambda_1.$$

Since  $f(u) = u$  is a degree-1 spherical harmonic, we get equality in this case. This completes the proof for non-increasing  $g$ .

When  $g$  is non-decreasing, the argument is the same except that the assumption  $\mathbf{E}[f] = 0$  implies that  $H_0 f = 0$ , and then [Corollary 4.9](#) implies that  $\lambda_1$  is the most positive among all remaining eigenvalues. Therefore,

$$\mathbf{E}_{\mathbf{u} \sim \omega} \langle f(\mathbf{u}), U_g f(\mathbf{u}) \rangle \leq \lambda_1,$$

and as before we have equality for  $f(u) = u$ .  $\square$

## 5 The full-dimensional case

Here we consider the case when  $f$  is an assignment from  $\mathbb{R}^n$  to the ball  $B^k$  when  $k = n$ , and we consider a correlation parameter  $-1 \leq \rho \leq 0$ .

**Theorem 5.1** (Vector-valued Borell’s inequality;  $n$ -dimensional outputs). *Let  $f : \mathbb{R}^n \rightarrow B^n$ . In addition, let  $f_{\text{opt}} : \mathbb{R}^n \rightarrow B^n$  be defined by  $f_{\text{opt}}(x) = x/\|x\|$ . Let  $-1 \leq \rho \leq 0$ . Then*

$$\mathbf{Stab}_\rho[f] \geq \mathbf{Stab}_\rho[f_{\text{opt}}].$$

Moreover, if  $\mathbf{Stab}_\rho[f] = \mathbf{Stab}_\rho[f_{\text{opt}}]$  then there is an orthogonal matrix  $M$  such that  $f(x) = f_{\text{opt}}(Mx)$  almost surely.

Our goal is to lower bound the expression

$$\mathbf{E}_{\mathbf{x} \sim_\rho \mathbf{y}} \langle f(\mathbf{x}), f(\mathbf{y}) \rangle.$$

Fourier analysis is a natural tool to bring to bear on this problem. Although it is natural to consider the Hermite polynomials since the expectation is “diagonal” in this set of polynomials (e.g. [O’D14], Proposition 11.33), the optimal  $f_{\text{opt}}$ ’s expansion in the Hermite basis of polynomials is complicated (see [PT21b] for a Hermite expansion yielding that of  $f_{\text{opt}}$ ). As a result, it is difficult to compare the value of  $f$  to the value of  $f_{\text{opt}}$  using the Hermite basis.

Instead, we reparameterize  $\mathbf{x}$  as  $\mathbf{x} = r \cdot \mathbf{u}$ , where  $r$  is the length (or radius) of  $\mathbf{x}$  and  $\mathbf{u}$  is the unit vector in the direction of  $\mathbf{x}$ . Similarly, we will reparameterize  $\mathbf{y}$  as  $\mathbf{y} = s \cdot \mathbf{v}$ . For each value  $r$  that the random variable  $r$  may take, we will think of  $f$  as specifying a separate function on the unit sphere  $S^{n-1}$ . We denote this function as  $f_r : S^{n-1} \rightarrow B^n$  and define it by

$$f_r(\mathbf{u}) := f(r \cdot \mathbf{u}) = f(\mathbf{x}).$$

Using this, we can rewrite our original expectation as

$$\mathbf{E}_{r,s} \mathbf{E}_{\mathbf{u},\mathbf{v}} \langle f_r(\mathbf{u}), f_s(\mathbf{v}) \rangle. \tag{12}$$

What is nice about this reparameterization is that it simplifies our optimizer  $f_{\text{opt}}$ . In particular, for each fixed  $r \geq 0$ ,  $(f_{\text{opt}})_r(u)$  is simply equal to  $u$ .

To analyze Equation (12), we first condition on fixed values of  $r, s \geq 0$ . This gives the expression

$$\mathbf{E}_{\mathbf{u},\mathbf{v}} \langle f_r(\mathbf{u}), f_s(\mathbf{v}) \rangle.$$

This is just an expectation involving two functions on the sphere (under a distribution on  $\mathbf{u}$  and  $\mathbf{v}$  described in the next section). If we take  $U_\rho^{r,s}$  as the standard Gaussian noise operator (Definition 3.2) conditioned on  $r, s$ , then we can then further rewrite the above expectation as

$$\mathbf{E}_{\mathbf{u}} \langle f_r(\mathbf{u}), U_\rho^{r,s} f_s(\mathbf{u}) \rangle, \tag{13}$$

where we may think of  $U_\rho^{r,s} f_s(u)$  as the average of  $f_s(\mathbf{v})$  over a random  $\mathbf{v}$ , conditioned on  $r, s$ , and  $\mathbf{u} = u$ . This noise operator turns out to fall into the setting that we considered in the previous section, and so applying Theorem 4.1 for each fixed  $r, s$  will allow us to prove Theorem 5.1.

## 5.1 The induced noise operator

If  $(\mathbf{x}, \mathbf{y})$  are  $\rho$ -correlated random variables then the probability density function (PDF) can be written as

$$G_\rho(x, y) = \frac{1}{A_\rho} e^{-\frac{\|\mathbf{x}\|^2 + \|\mathbf{y}\|^2 - 2\rho\langle \mathbf{x}, \mathbf{y} \rangle}{2(1-\rho^2)}} = \frac{1}{A_\rho} e^{-\frac{\|\mathbf{x}\|^2 + \|\mathbf{y}\|^2}{2(1-\rho^2)}} e^{\frac{\rho rs\langle \mathbf{u}, \mathbf{v} \rangle}{(1-\rho^2)}},$$

where  $A_\rho$  is a normalizing constant. When we reparameterize according to  $(r, s, u, v)$  we obtain:

$$G_\rho(r, s, u, v) = \frac{1}{A_\rho} (rs)^{n-1} e^{-\frac{r^2 + s^2}{2(1-\rho^2)}} e^{\frac{\rho rs\langle \mathbf{u}, \mathbf{v} \rangle}{(1-\rho^2)}},$$

where the  $(rs)^{n-1}$  factor arises from the change of variables.

We will be more interested, however, in the conditional distributions:

**Definition 5.2** (Conditioned correlated Gaussians). We denote by  $G_\rho^{r,s}(u, v)$  the PDF of  $(\mathbf{u}, \mathbf{v})$ , with respect to the measure  $\omega$ , conditioned on the values  $r, s$ . We write  $(\mathbf{u}, \mathbf{v}) \sim N_\rho^{r,s}$  for correlated random variables drawn from this distribution.

We denote by  $G_\rho^{r,s}(v | u)$  the PDF of  $\mathbf{v}$ , with respect to the measure  $\omega$ , conditioned on the values  $r, s$ , and  $u$ . This can be written as

$$G_\rho^{r,s}(v | u) = \frac{1}{A_\rho^{r,s}} e^{\frac{\rho rs\langle \mathbf{u}, \mathbf{v} \rangle}{(1-\rho^2)}}, \quad (14)$$

where  $A_\rho^{r,s}$  is a normalizing constant that depends on  $r, s$ , and  $\rho$ .

We note that (14) depends only on the quantity  $\langle \mathbf{u}, \mathbf{v} \rangle$ , and it is monotonically decreasing in this quantity because  $r, s \geq 0$  and  $\rho \leq 0$ .

**Definition 5.3** (Conditioned Gaussian noise operator). The *conditioned Gaussian noise operator* is an operator on  $L_2(S^{n-1}, \omega)$  which acts on a function  $f : S^{n-1} \rightarrow \mathbb{R}$  as:

$$U_\rho^{r,s} f(u) = \mathbf{E}_{(\mathbf{u}, \mathbf{v}) \sim N_\rho^{r,s}} [f(\mathbf{v}) | \mathbf{u} = u] = \int_{S^{n-1}} G_\rho^{r,s}(v | u) f(v) d\omega(v).$$

Note in particular that for every  $r, s \geq 0$  and for every  $-1 \leq \rho \leq 0$ ,  $U_\rho^{r,s}$  is a noise operator of the form (11), for the non-increasing function  $g(t) = \frac{1}{A_\rho^{r,s}} e^{\frac{\rho rst}{1-\rho^2}}$ . We have that for any  $u$ ,

$$1 = \int_{S^{n-1}} G_\rho^{r,s}(v | u) d\omega(v) = \int_{-1}^1 g(t) d\tilde{\omega}(t),$$

by the definitions of  $\omega$  and  $\tilde{\omega}$ , demonstrating that  $g \geq 0$  satisfies (10). Recalling from [Theorem 4.3](#) that the eigenfunctions of  $U_\rho^{r,s}$  are spherical harmonics, let  $\lambda_d^{r,s}$  be the eigenvalue corresponding to  $\mathcal{H}_d$ ; i.e.,  $U_\rho^{r,s} h = \lambda_d^{r,s} h$  for all  $h \in \mathcal{H}_d$ . Then [Corollary 4.9](#) implies that

$$\lambda_1^{r,s} \leq -|\lambda_d|^{r,s} \text{ for every } d \geq 0, \quad (15)$$

with equality only if  $d = 1$ . Moreover, the fact that  $G_\rho^{r,s}$  is a probability density implies that  $\lambda_0^{r,s} = 1$ .

The following lemma gives our main lower-bound. It shows that if  $f_r$  and  $f_s$  have mean zero, then the average inner product is lower-bounded by  $\lambda_1^{r,s}$ , exactly the value that the optimizer  $f_{\text{opt}}$  would achieve. However, when they are not mean-zero, they can outperform the optimizer; consider  $f_r = (1, 0, \dots, 0)$  and  $f_s = (-1, 0, \dots, 0)$ , which have average inner-product  $-1$ . To compensate for this, the lemma includes a correction factor depending on the means of  $f_r$  and  $f_s$ .

**Lemma 5.4** (Main lower bound).

$$\mathbf{E}_{(\mathbf{u}, \mathbf{v}) \sim N_\rho^{r,s}} \langle f_r(\mathbf{u}), f_s(\mathbf{v}) \rangle \geq \langle \mathbf{E}[f_r], \mathbf{E}[f_s] \rangle + \lambda_1^{r,s},$$

with equality if and only if there is an orthogonal matrix  $M$  so that  $f_r(u) = f_s(u) = Mu$ .

*Proof.* Recalling that  $H_d$  is the orthogonal projection onto  $\mathcal{H}_d$  and that  $U_\rho^{r,s}h = \lambda_d^{r,s}h$  for all  $h \in \mathcal{H}_d$ , we have

$$\begin{aligned} \mathbf{E}_{(\mathbf{u}, \mathbf{v}) \sim N_\rho^{r,s}} \langle f_r(\mathbf{u}), f_s(\mathbf{v}) \rangle &= \mathbf{E}_{\mathbf{u} \sim \omega} \langle f_r(\mathbf{u}), U_\rho^{r,s} f_s(\mathbf{u}) \rangle \\ &= \sum_{d \geq 0} \mathbf{E}_{\mathbf{u} \sim \omega} \langle f_r(\mathbf{u}), U_\rho^{r,s} H_d f_s(\mathbf{u}) \rangle \\ &= \sum_{d \geq 0} \lambda_d^{r,s} \mathbf{E}_{\mathbf{u} \sim \omega} \langle f_r(\mathbf{u}), H_d f_s(\mathbf{u}) \rangle \\ &= \sum_{d \geq 0} \lambda_d^{r,s} \mathbf{E}_{\mathbf{u} \sim \omega} \langle H_d f_r(\mathbf{u}), H_d f_s(\mathbf{u}) \rangle \end{aligned}$$

Recall that  $\lambda_0^{r,s} = 1$  and that  $H_0 f_r$  is the constant function  $\mathbf{E}_{\mathbf{u} \sim \omega}[f_r(\mathbf{u})]$ ; then

$$\mathbf{E}_{(\mathbf{u}, \mathbf{v}) \sim N_\rho^{r,s}} \langle f_r(\mathbf{u}), f_s(\mathbf{v}) \rangle = \langle \mathbf{E}[f_r], \mathbf{E}[f_s] \rangle + \sum_{d \geq 1} \lambda_d^{r,s} \mathbf{E}_{\mathbf{u} \sim \omega} \langle H_d f_r(\mathbf{u}), H_d f_s(\mathbf{u}) \rangle.$$

For the second term, Cauchy-Schwarz (twice) and (15) imply that

$$\sum_{d \geq 1} \lambda_d^{r,s} \mathbf{E}_{\mathbf{u} \sim \omega} \langle H_d f_r(\mathbf{u}), H_d f_s(\mathbf{u}) \rangle \geq - \sum_{d \geq 1} |\lambda_d^{r,s}| \sqrt{\mathbf{E}[\|H_d f_r\|^2] \mathbf{E}[\|H_d f_s\|^2]} \quad (16)$$

$$\geq \lambda_1^{r,s} \sum_{d \geq 1} \sqrt{\mathbf{E}[\|H_d f_r\|^2] \mathbf{E}[\|H_d f_s\|^2]} \quad (17)$$

$$\geq \lambda_1^{r,s} \sqrt{\sum_{d \geq 1} \mathbf{E}[\|H_d f_r\|^2] \cdot \sum_{d \geq 1} \mathbf{E}[\|H_d f_s\|^2]}$$

Finally, recalling that  $\sum_{d \geq 0} \mathbf{E}[\|H_d f_r\|^2] = \mathbf{E}[\|f_r\|^2] \leq 1$ , we have

$$\sum_{d \geq 1} \lambda_d^{r,s} \mathbf{E}_{\mathbf{u} \sim \omega} \langle H_d f_r(\mathbf{u}), H_d f_s(\mathbf{u}) \rangle \geq \lambda_1^{r,s}.$$

This completes the proof of the inequality.

The equality simply follows because if  $f_r(u) = f_s(u) = u$  then all inequalities in this proof are equalities. By the equality cases in (15), we have equality in (17) if and only if  $f_r$  and  $f_s$  are both affine functions:  $f_r(u) = \mathbf{E}[f_r] + M_r u$  for some  $n \times n$  matrix  $M_r$  and  $f_s(u) = \mathbf{E}[f_s] + M_s u$  for some  $n \times n$  matrix  $M_s$ . Then we have equality in (16) if and only if  $M_s$  is a non-negative scalar multiple of  $M_r$ . Because  $f_r(u)$  takes values in  $B^n$ , we must have  $\|M_r\|_{\text{op}} \leq 1$  and  $\|M_s\|_{\text{op}} \leq 1$ . But in order to have equality in  $\mathbf{E}[\|H_1 f_r\|^2] \leq 1$ , we must have  $\|M_r\|_2^2 = 1$ , and so  $M_r$  is an orthogonal matrix. Similarly  $M_s$  must be an orthogonal matrix, and since it is a non-negative multiple of  $M_r$  they must be equal. Finally,  $\mathbf{E}[\|f_r\|^2] = \|\mathbf{E}[f_r]\|^2 + \mathbf{E}[\|H_1 f_r\|^2] \leq 1$ , and so if  $\mathbf{E}[\|H_1 f_1\|^2] = 1$  then we must have  $\mathbf{E}[f_r] = 0$ ; similarly for  $\mathbf{E}[f_s]$ .  $\square$

Now we prove [Theorem 5.1](#). At a high-level, the correction term in [Lemma 5.4](#) showed that one can improve on the optimizer for fixed  $r$  and  $s$  by using nonzero means  $\mathbf{E}[f_r]$  and  $\mathbf{E}[f_s]$ . What we will now show is that although this is true for fixed  $r$  and  $s$ , when averaged over random  $\mathbf{r}$  and  $\mathbf{s}$  this correction term no longer helps. In other words, we will show that  $\mathbf{E}_{\mathbf{r},\mathbf{s}}\langle \mathbf{E}_{\mathbf{u}}[f_{\mathbf{r}}(\mathbf{u})], \mathbf{E}_{\mathbf{u}}[f_{\mathbf{s}}(\mathbf{u})] \rangle$  is nonnegative, and so it can only increase the average inner product.

*Proof of [Theorem 5.1](#).* Our goal is to lower-bound

$$\mathbf{E}_{\mathbf{x} \sim \rho \mathbf{y}} \langle f(\mathbf{x}), f(\mathbf{y}) \rangle = \mathbf{E}_{\mathbf{r},\mathbf{s}} \mathbf{E}_{\mathbf{u},\mathbf{v}} \langle f_{\mathbf{r}}(\mathbf{u}), f_{\mathbf{s}}(\mathbf{v}) \rangle. \quad (18)$$

Setting  $g(r) = \mathbf{E}_{\mathbf{u}} f_r(\mathbf{u})$ , [Lemma 5.4](#) implies that this is at least

$$(18) \geq \mathbf{E}_{\mathbf{r},\mathbf{s}} [\langle g(\mathbf{r}), g(\mathbf{s}) \rangle + \lambda_1^{r,s}] = \mathbf{E}_{\mathbf{r},\mathbf{s}} \langle g(\mathbf{r}), g(\mathbf{s}) \rangle + \mathbf{E}_{\mathbf{r},\mathbf{s}} \lambda_1^{r,s}.$$

The second term is exactly the value of our conjectured optimizer via [Lemma 5.4](#). As a result, it suffices to show that the first term is nonnegative. We will begin by rewriting it as

$$\mathbf{E}_{\mathbf{x} \sim \rho \mathbf{y}} \langle g(\|\mathbf{x}\|), g(\|\mathbf{y}\|) \rangle. \quad (19)$$

Consider the following method of drawing two  $\rho$ -correlated strings  $\mathbf{x}$  and  $\mathbf{y}$ : first, sample  $\mathbf{z}, \mathbf{z}', \mathbf{z}'' \sim \mathcal{N}(0, 1)^n$ . Next, set

$$\mathbf{x} = \rho' \cdot \mathbf{z} + \sqrt{1 - (\rho')^2} \cdot \mathbf{z}', \quad \mathbf{y} = -(\rho' \cdot \mathbf{z} + \sqrt{1 - (\rho')^2} \cdot \mathbf{z}''),$$

where  $\rho' = \sqrt{-\rho}$ . Then conditioned on  $\mathbf{z}$ ,  $\mathbf{x}$  and  $-\mathbf{y}$  are independent and identically distributed random variables. Hence, we can write

$$(19) = \mathbf{E}_{\mathbf{z}} \mathbf{E}_{\mathbf{x},\mathbf{y}} \langle g(\|\mathbf{x}\|), g(\|\mathbf{y}\|) \rangle = \mathbf{E}_{\mathbf{z}} (\mathbf{E}_{\mathbf{x}} g(\|\mathbf{x}\|), \mathbf{E}_{\mathbf{y}} g(\|\mathbf{y}\|)) = \mathbf{E}_{\mathbf{z}} (\mathbf{E}_{\mathbf{x}} g(\|\mathbf{x}\|), \mathbf{E}_{\mathbf{y}} g(\|-\mathbf{y}\|)).$$

Note that the last equality holds because  $\|-\mathbf{y}\| = \|\mathbf{y}\|$ . For each  $\mathbf{z}$ , the two terms in the inner product are equal, and so this is nonnegative. This completes the proof of the inequality.

To see the equality cases, recall that we applied the bound of [Lemma 5.4](#) for every  $r$  and  $s$ . If equality is attained in the inequality, we must have equality in [Lemma 5.4](#) for almost every  $r$  and  $s$ . It follows that the matrix  $M$  of [Lemma 5.4](#) must be independent of  $r$  and  $s$ , and the claimed characterization of equality cases follows.  $\square$

## 5.2 The positive- $\rho$ case

We assumed in this section that  $\rho \leq 0$ . In the case  $\rho > 0$ , the Gaussian noise model induces a spherical noise model of the form [\(11\)](#) with an *increasing* function  $g$ . By the results of [Section 4](#),

$$\lambda_1^{r,s} \geq |\lambda_d^{r,s}| \quad (20)$$

for all  $d \geq 2$ , and so [Lemma 5.4](#) may be extended to the  $\rho > 0$  case, with the opposite inequality. The problem comes from the first term on the right hand side of [Lemma 5.4](#); this term has a non-negative sign, which is in our favor when  $\rho < 0$  but against us when  $\rho > 0$ . It is possible that this non-negative term is cancelled out by the difference between the two sides of [\(??\)](#), but we were not able to show this.

## 6 Dimension reduction

We will eventually be concerned with 3-dimensional assignments to points which lie in a  $n$ -dimensional sphere,  $S^{n-1}$ . **Theorem 5.1** shows if we are allowed  $n$ -dimensional assignments,  $f_{\text{opt}}$  minimizes the noise stability for negative  $\rho$ . In this section, we will show that the optimization over  $k$ -dimensional assignments (for  $k \leq n$ ) reduces to the optimization over  $n$ -dimensional assignments, *but only for non-negative  $\rho$* . We do this by showing that optimally stable functions are “at most  $k$ -dimensional,” in the sense that they can be defined on  $\mathbb{R}^k$  and not on  $\mathbb{R}^n$ .

We say that  $f : \mathbb{R}^n \rightarrow B^k$  is optimally stable with parameter  $\rho \in [0, 1]$  if  $\rho > 0$  and  $\mathbf{E}_{\mathbf{x} \sim \rho \mathbf{y}}[\langle f(\mathbf{x}), f(\mathbf{y}) \rangle]$  is maximal among all functions  $f : \mathbb{R}^n \rightarrow B^k$  with  $\mathbf{E}_{\mathbf{x}}[f(\mathbf{x})] = 0$ .

**Theorem 6.1.** *For every  $n, k \geq 1$  and every  $\rho \in [0, 1]$ , there is an optimally stable function  $f$ . Moreover, if  $k \leq n$  and  $\rho \in (0, 1)$  then for every optimally stable function  $f$ , after a change of coordinates on  $\mathbb{R}^n$ ,  $f(x)$  depends only on  $x_1, \dots, x_k$ .*

Let’s address the existence part first, because it’s easier.

*Proof of existence in Theorem 6.1.* When  $\rho \in \{0, 1\}$ , existence is trivial because every function is optimally stable; from now on, assume  $\rho \in (0, 1)$ .

Choose an optimizing sequence  $f_n$ , i.e. a sequence of functions  $f_n : \mathbb{R}^n \rightarrow B^k$  such that  $\mathbf{E}[f_n] = 0$  and  $\mathbf{E}[\langle f(\mathbf{x}), f(\mathbf{y}) \rangle]$  converges to the optimal value. Since  $f_n$  are uniformly bounded in  $L^2(\gamma)$ , after passing to a subsequence we may assume that  $f_n$  converges weakly to, say,  $f$ . By testing weak convergence against a constant function, it follows that  $\mathbf{E}[f] = 0$ .

Recall that the noise operator  $U_\rho : L^2(\gamma) \rightarrow L^2(\gamma)$  is compact – for example, because it acts diagonally on the Hermite basis, with eigenvalues that converge to zero. It follows that  $U_\rho f_n$  converges strongly in  $L^2(\gamma)$  to  $U_\rho f$  and hence

$$\begin{aligned} \mathbf{E}_{\mathbf{x} \sim \rho \mathbf{y}}[\langle f_n(\mathbf{x}), f_n(\mathbf{y}) \rangle] &= \mathbf{E}_{\mathbf{x}}[\langle f_n(\mathbf{x}), U_\rho f_n(\mathbf{x}) \rangle] \\ &= \mathbf{E}[\langle f_n, U_\rho f_n - U_\rho f \rangle] + \mathbf{E}[\langle f_n, U_\rho f \rangle] \\ &\rightarrow \mathbf{E}[\langle f, U_\rho f \rangle], \end{aligned}$$

where the first term converged to zero because  $\|f_n\|$  is bounded and  $\|U_\rho f_n - U_\rho f\| \rightarrow 0$ , and the second term converged to  $\mathbf{E}[\langle f, U_\rho f \rangle]$  by the weak convergence of  $f_n$ . Since  $\mathbf{E}[\langle f, U_\rho f \rangle] = \mathbf{E}_{\mathbf{x} \sim \rho \mathbf{y}}[\langle f(\mathbf{x}), f(\mathbf{y}) \rangle]$ , the limit function  $f$  is optimally stable.  $\square$

For a differentiable function  $f : \mathbb{R}^n \rightarrow \mathbb{R}^k$ , we write  $Df(x)$  for the  $k \times n$  matrix of partial derivatives at the point  $x \in \mathbb{R}^n$ . For  $v \in \mathbb{R}^n$ , we will write  $D_v f(x) \in \mathbb{R}^k$  for the directional derivative of  $f$  in the direction  $v$ . Of course,  $D_v f(x)$  is just an abbreviation for  $(Df(x)) \cdot v$ .

### 6.1 Outline of the dimension reduction

The main idea behind the proof of **Theorem 6.1** is perturbative: we show that if the function depends on more than  $k$  coordinates, there is a perturbation  $\tilde{f}$  of  $f$  that satisfies  $\mathbf{E}[\tilde{f}] = 0$  but has a better noise stability. We will consider two families of perturbations: “value” perturbations of the form  $\tilde{f}(x) = f(x) + \epsilon \psi(x) + o(\epsilon)$ , and “spatial” perturbations of the form  $\tilde{f}(x) = f(x + \epsilon \Psi(x)) + o(\epsilon)$ ; our final perturbation will be a combination of these.

The perturbation  $\tilde{f}$  will never be written down very explicitly. In most of our analysis, we will rather consider a one-parameter family  $f_\epsilon$  of perturbations, and we will establish the *existence* of a good perturbation by studying the derivatives of  $f_\epsilon$  at  $\epsilon = 0$ .

There are many technical details, partly because we are considering an infinite-dimensional optimization problem (over all  $f : \mathbb{R}^n \rightarrow B^k$ ) and partly because a priori the optimal functions could be almost arbitrarily nasty. However, most of our arguments have simple analogues for finite-dimensional constrained optimization. In particular, suppose that we are trying to maximize a differentiable function  $\psi : \mathbb{R}^m \rightarrow \mathbb{R}$  while obeying the constraint  $g(x) = 0$ , for a differentiable  $g : \mathbb{R}^m \rightarrow \mathbb{R}^k$ . Classical Lagrangian theory for this problem implies that if  $x_0 \in \mathbb{R}^m$  is a maximizer and  $Dg(x_0)$  has rank  $k$  then there is some  $\lambda \in \mathbb{R}^k$  such that  $D\psi(x_0) = \lambda^T Dg(x_0)$ : if this were not the case, there would be a curve  $c : [-\delta, \delta] \rightarrow \mathbb{R}^m$  with  $c(0) = x_0$ ,  $g(c(t)) \equiv 0$ , and  $\frac{d}{dt}\big|_{t=0} \psi(c(t)) \neq 0$ , contradicting the maximality of  $x_0$ .

The classical theory extends to second-order (at least, if  $\psi$  and  $g$  are twice-differentiable): if  $x_0$  is a maximizer and  $Dg(x_0)$  has rank  $k$  then the matrix  $D^2\psi - \sum_i \lambda_i D^2g_i$  acts negatively on the kernel of  $Dg(x_0)$  (where  $\lambda = (\lambda_1, \dots, \lambda_k)$  is the one whose existence was guaranteed by the first-order theory). This is essentially the constrained-optimization analogue of the statement that a function has a negative-semidefinite Hessian at a maximizer, and it can be proven by showing that if it fails to hold then there is a curve  $c : [-\delta, \delta] \rightarrow \mathbb{R}^m$  with  $c(0) = c_0$ ,  $g(c(t)) \equiv 0$ , and  $\frac{d^2}{dt^2}\big|_{t=0} \psi(c(t)) > 0$ , contradicting the maximality of  $x_0$ .

To prove Theorem 6.1, we first find analogues of the first- and second-order variational principles above. For the first-order conditions, we show (Lemma 6.10) that there exists  $\lambda \in \mathbb{R}^k$  such that

$$|U_\rho f - \lambda/2|f = U_\rho f - \lambda/2. \quad (21)$$

For the second-order conditions, we show that for the same  $\lambda$  and for any nice enough vector field  $\Psi : \mathbb{R}^n \rightarrow \mathbb{R}^n$  satisfying  $\mathbf{E}[D_{\Psi(x)}f(\mathbf{x})] = 0$ ,

$$\mathbf{E}_{\mathbf{x} \sim_\rho \mathbf{y}} [\langle D_{\Psi(\mathbf{x})}f(\mathbf{x}), D_{\Psi(\mathbf{y})}f(\mathbf{y}) \rangle] - \mathbf{E}_{\mathbf{x}} [|U_\rho f - \lambda/2| \cdot |D_{\Psi(\mathbf{x})}f(\mathbf{x})|^2] \leq 0. \quad (22)$$

Note that the expression above is a quadratic function of the vector field  $\Psi$ , which can be thought of as a “direction” along which we perturb  $f$ . In particular, our second-order condition really says – as in the finite-dimensional case – that a certain quadratic form acts non-positively on a certain subspace.

Finally, we test (22) by substituting constant vector fields  $\Psi(x) \equiv v \in \mathbb{R}^n$ , and show that either

$$\mathbf{E}_{\mathbf{x} \sim_\rho \mathbf{y}} [\langle D_v f(\mathbf{x}), D_v f(\mathbf{y}) \rangle] - \mathbf{E}_{\mathbf{x}} [|U_\rho f - \lambda/2| \cdot |D_v f(\mathbf{x})|^2] > 0$$

or  $D_v f \equiv 0$ . Hence, for every  $v \in \mathbb{R}^n$ ,  $\mathbf{E}[D_v f(\mathbf{x})] = 0$  implies  $D_v f \equiv 0$ . The function  $v \mapsto \mathbf{E}[D_v f(\mathbf{x})]$  is linear, so if  $W \subset \mathbb{R}^n$  is its kernel then  $W$  has codimension at least  $k$ . After applying a change of variables so that  $\text{span}\{e_1, \dots, e_k\} \subseteq W^\perp$  the fact that  $D_v f \equiv 0$  for  $v \in W$  implies that  $f$  is a function only of  $x_1, \dots, x_k$ .

## 6.2 Technicalities

One problem with the outline above is that we wrote “ $D_v f$ ” several times, but no one told us that the optimal function  $f$  was differentiable.

We get around this difficulty by exploiting the “smoothness” of our objectives and constraints. For example, we don’t care so much about the derivatives of  $f$  as we do about how  $\mathbf{E}[f_\epsilon]$  changes as we vary  $\epsilon$ . But  $\mathbf{E}[f_\epsilon]$  has as many derivatives (in  $\epsilon$ ) as we wish, because we may write  $\mathbf{E}[f_\epsilon] = \int f(x + \epsilon\Psi(x) + o(\epsilon)) \frac{d\gamma}{dx} dx$  and then use a change of variables to pass the spatial perturbation onto the (very smooth) Gaussian density. Organizing the computations with this explicit change of

variables is tedious, so what we actually do is to first derive our perturbative formulas for smooth functions  $f$ , then integrate by parts to push the derivatives onto  $\frac{d\gamma}{dx}$ . We then get formulas that make sense for non-smooth  $f$ ; we show that they actually hold for non-smooth  $f$  by taking smooth approximations.

The rest of this section is about the integration-by-parts formulas and uniform approximations that make everything go through rigorously. In particular, we prove several non-smooth analogues of statements that are trivial for differentiable functions.

For a  $C^1$  vector field  $W$ , define

$$\operatorname{div}_\gamma W(x) = \operatorname{div} W(x) - \langle W(x), x \rangle.$$

Note that this satisfies the product rule  $\operatorname{div}_\gamma(fW) = f \operatorname{div}_\gamma W + \nabla_W f$  for  $C^1$  functions  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ . The point of this definition is the formula

$$\int \operatorname{div}_\gamma W d\gamma = 0$$

for compactly supported  $W$ . Using the product rule, this is equivalent to

$$\int f \operatorname{div}_\gamma W d\gamma = - \int \nabla_W f d\gamma \tag{23}$$

for compactly supported  $W$  and/or  $f$ . Now think of the left hand side as *defining* the derivative of  $f$  in a weak sense, noting that the left hand side makes sense for non-smooth  $f$ .

Because of the way (23) expresses derivatives of  $f$  in terms of derivatives of  $W$ , we will need to impose regularity on the vector fields  $W$  that we consider.

**Definition 6.2.** A vector field  $W$  is *tame* if it's bounded,  $C^\infty$ -smooth, and if its derivatives of all orders are bounded.

Next, we define our spatial perturbations and our main tool for approximating it by smooth functions: let  $W : \mathbb{R}^n \rightarrow \mathbb{R}^n$  be a tame vector field and let  $\{F_t : t \in \mathbb{R}\}$  be the flow along  $W$ , defined as the unique function satisfying  $F_0(x) = x$  and

$$\frac{dF_t(x)}{dt} = W(F_t(x))$$

for all  $t \in \mathbb{R}$  and  $x \in \mathbb{R}^n$ . Then  $F_t$  is a  $C^\infty$  diffeomorphism for all  $t$ . Given an optimal function  $f : \mathbb{R}^n \rightarrow \mathbb{R}^k$ , we may consider the competitor function  $\mathcal{S}_{t,W}f$  given by

$$(\mathcal{S}_{t,W}f)(x) = f(F_t^{-1}(x)).$$

It is well-known that functions in  $L^2(\gamma)$  can be approximated (for example, by truncating and mollifying) using smooth functions. The point here is that we can do this approximation in such a way that it also applies *uniformly in  $t$*  to the spatial perturbations  $\mathcal{S}_{t,W}f$ .

**Lemma 6.3.** *If  $f : \mathbb{R}^n \rightarrow \mathbb{R}^k$  is bounded and  $W$  is tame then there is a sequence uniformly bounded functions  $f_n \in C_c^\infty$  such that*

$$\sup_{t \in [-1,1]} \|\mathcal{S}_{t,W}f - \mathcal{S}_{t,W}f_n\|_{L^2(\gamma)} \rightarrow 0.$$

*Proof.* Using a change of variables, we can write

$$\|\mathcal{S}_{t,W}g\|_{L_2(\gamma)}^2 = \int \|g(x)\|^2 |DF_t(x)| \phi(F_t(x)) dx,$$

where  $|DF_t|$  denotes the Jacobian determinant of  $F_t$ . Now define  $\tilde{\phi}(x) = \sup_{t \in [-1,1]} |DF_t(x)| \phi(F_t(x))$ . Then  $\tilde{\phi}$  is integrable:  $|DF_t(x)|$  is uniformly bounded for  $t \in [-1, 1]$ ; also  $|F_t(x) - x|$  is uniformly bounded and so

$$\phi(F_t(x)) \leq C \exp(-(|x| - C)_+^2/2)$$

for some  $C$ , which is integrable. Define the finite measure  $d\tilde{\gamma} = \tilde{\phi} dx$ ; note that our definition of  $\tilde{\phi}$  ensures that

$$\|\mathcal{S}_{t,W}g\|_{L_2(\gamma)} \leq \|g\|_{L_2(\tilde{\gamma})}$$

for every  $t \in [-1, 1]$ .

Finally, take a uniformly bounded sequence of functions  $f_n \in C_c^\infty$  such that  $f_n \rightarrow f$  in  $L^2(\tilde{\gamma})$ . Then the claim follows, because

$$\sup_{t \in [-1,1]} \|\mathcal{S}_{t,W}f - \mathcal{S}_{t,W}f_n\|_{L_2(\gamma)} = \sup_{t \in [-1,1]} \|\mathcal{S}_{t,W}(f - f_n)\|_{L_2(\gamma)} \leq \|f - f_n\|_{L_2(\tilde{\gamma})}.$$

□

If  $f : \mathbb{R}^n \rightarrow S^{k-1}$  is differentiable then for any  $x, v \in \mathbb{R}^n$ ,  $D_v f(x)$  is tangent to  $S^{k-1}$  at  $f(x)$ ; or in other words,  $\langle D_v f, f \rangle \equiv 0$ . Here is an analogue for certain non-smooth  $f$ .

**Lemma 6.4.** *Suppose  $f : \mathbb{R}^n \rightarrow B^k$  is measurable and  $\psi : \mathbb{R}^n \rightarrow [0, \infty)$  is a bounded, Lipschitz function that is differentiable on  $\{\psi > 0\}$ . Assume that  $f(x) \in S^{k-1}$  whenever  $\psi(x) > 0$ , and that  $\psi f$  has a uniformly bounded derivative. Then, for every tame vector field  $W$ ,*

$$\sum_i \int f_i \operatorname{div}_\gamma(\psi f_i W) d\gamma = 0.$$

(To see why this is an analogue of the easy fact above, note that integration by parts shows that the left hand side is  $-\int \psi \langle D_W f, f \rangle d\gamma$  in the case of smooth  $f$ .)

*Proof.* Since  $\operatorname{div}_\gamma(\psi f_i W)$  is integrable, by the dominated convergence theorem it suffices to find uniformly bounded functions  $f^\epsilon$  converging pointwise to  $f$  such that

$$\lim_{\epsilon \rightarrow 0} \sum_i \int f_i^\epsilon \operatorname{div}_\gamma(\psi f_i W) d\gamma = 0. \quad (24)$$

Fix a constant  $C \geq 1$  large enough to be larger than the the uniform bound on  $\psi$  and the Lipschitz constants of both  $\psi$  and  $\psi f$ .

For  $\epsilon > 0$ , let  $\eta^\epsilon : [0, \infty) \rightarrow [0, \infty)$  be a  $C^1$  function satisfying

- $\eta^\epsilon(s) = s$  for  $s \geq \epsilon$ ,
- $\eta^\epsilon(s) \geq \frac{\epsilon}{2}$  for all  $s$ ,
- $(\eta^\epsilon)'(s) \leq 1$  for all  $s$ ,
- $(\eta^\epsilon)'(0) = 0$ .

Now define  $\psi^\epsilon = \eta^\epsilon \circ \psi$  and  $f^\epsilon = \frac{\psi}{\psi^\epsilon} f$ . Note that  $f^\epsilon = f$  whenever  $\psi \geq \epsilon$ . Moreover,  $\psi^\epsilon$  is  $\mathcal{C}^1$  with

$$|D_v \psi^\epsilon| \leq |(\eta^\epsilon)' \circ \psi| |D_v \psi| \leq C$$

for any unit vector  $v$ . Since  $\eta^\epsilon \geq \frac{\epsilon}{2}$ , it follows that  $f^\epsilon$  is  $\mathcal{C}^1$  with

$$|D_v f^\epsilon| \leq \frac{2}{\epsilon} |D_v(\psi f)| + |\psi f| \frac{|D_v \psi^\epsilon|}{\epsilon/2} \leq \frac{4C^2}{\epsilon} \quad (25)$$

for every unit vector  $v$ .

Since  $f^\epsilon = f$  whenever  $\psi \geq \epsilon$ , we have  $|f^\epsilon(x)| = 1$  on  $\{\psi \geq \epsilon\}$ . It follows then that  $\langle D_v f^\epsilon(x), f(x) \rangle \equiv 0$  on  $\{\psi \geq \epsilon\}$ . Therefore,

$$\begin{aligned} \sum_i \int f_i^\epsilon \operatorname{div}_\gamma(\psi f_i W) d\gamma &= - \int \langle D_W f^\epsilon, \psi f \rangle d\gamma \\ &= - \int_{\{\psi < \epsilon\}} \langle D_W f^\epsilon, \psi f \rangle d\gamma. \end{aligned}$$

By (25),

$$\left| \int_{\{\psi < \epsilon\}} \langle D_W f^\epsilon, \psi f \rangle d\gamma \right| \leq 4C^2 \int_{\{\psi < \epsilon\}} \frac{\psi |W|}{\epsilon} d\gamma \leq 4C^2 \int_{\{0 < \psi < \epsilon\}} |W| d\gamma.$$

Since  $|W|$  is uniformly bounded, the final bound converges to zero as  $\epsilon \rightarrow 0$ . This establishes (24) and thus completes the proof.  $\square$

Here's a simple bound on the derivatives of  $U_\rho g$  for any  $L^2$  function  $g$ .

**Lemma 6.5.** *For any  $g \in L^2(\gamma)$  and any  $-1 < \rho < 1$ ,  $U_\rho g$  is  $\mathcal{C}^\infty$  smooth and satisfies*

$$\mathbf{E}[\|\nabla^k U_\rho g\|^2] \leq C(\rho, k) \mathbf{E}[g^2]$$

for some constant  $C(\rho, k) < \infty$ , where  $\|\nabla^k g\|_2^2$  denotes the sum of squares of all  $k$ th order partial derivatives of  $g$ .

*Proof.* With the change of variables  $z = \rho x + \sqrt{1 - \rho^2} y$ , we can write

$$\begin{aligned} U_\rho g(x) &= (2\pi)^{-n/2} \int g(\rho x + \sqrt{1 - \rho^2} y) e^{-|y|^2/2} dy \\ &= (2\pi(1 - \rho^2))^{-n/2} \int g(z) e^{-\frac{|z - \rho x|^2}{2(1 - \rho^2)}} dz. \end{aligned}$$

This last formula is clearly differentiable in  $x$ .

To show the claimed bound, recall that if  $H_\alpha$  are the orthonormal Hermite functions (where  $\alpha$  is a multi-index) then  $\frac{\partial}{\partial x_i} H_\alpha = \sqrt{\alpha_i} H_{\alpha - e_i}$ . Hence, if  $g = \sum_\alpha H_\alpha \hat{g}_\alpha$  is the Hermite expansion of  $g$  then

$$\mathbf{E} \left[ \left( \frac{\partial}{\partial x_i} U_\rho g \right)^2 \right] = \sum_\alpha \alpha_i e^{-2\rho|\alpha|} \hat{g}_\alpha^2$$

and so

$$\mathbf{E} [\|\nabla U_\rho g\|^2] = \sum_\alpha |\alpha| e^{-2\rho|\alpha|} \hat{g}_\alpha^2$$

The claimed inequality for  $k = 1$  follows because, for  $x \geq 0$ ,  $x e^{-\rho x}$  is bounded by a constant depending on  $\rho$ ; for larger  $k$  it follows by induction on  $k$ .  $\square$

Here is an integrated-by-parts version of the obvious fact that if  $\mathbf{E}[|D_w f|^2] = 0$  then  $f(x)$  is “independent of  $w$ ” in the sense that  $f(x) = f(y)$  whenever  $x$  and  $y$  differ by a multiple of  $w$ .

**Lemma 6.6.** *For  $f \in L^2(\gamma)$ ,  $w \in \mathbb{R}^n$ , and  $0 < \rho < 1$ ,*

$$\sum_i \mathbf{E}[f_i \operatorname{div}_\gamma((D_w U_\rho f_i)w)] \leq 0,$$

with equality if and only if there is a function  $g : w^\perp \rightarrow \mathbb{R}$  with  $f(x) = g(\Pi_{w^\perp} x)$  almost surely.

*Proof.* Fix  $s < 1$ ; since  $U_s f$  is sufficiently smooth (e.g. by Lemma 6.5),

$$\begin{aligned} \sum_i \mathbf{E}[U_s f_i \operatorname{div}_\gamma((D_w U_\rho f_i)w)] &= - \sum_i \mathbf{E}[\langle D_w U_s f_i, D_w U_\rho f_i \rangle] \\ &= - \sum_i \mathbf{E}[\langle D_w U_s f_i, D_w U_{\sqrt{\rho/s}} U_{\sqrt{\rho s}} f_i \rangle] \\ &= - \sqrt{\frac{s}{\rho}} \sum_i \mathbf{E}[\langle D_w U_s f_i, U_{\sqrt{\rho/s}} D_w U_{\sqrt{\rho s}} f_i \rangle] \\ &= - \sqrt{\frac{s}{\rho}} \sum_i \mathbf{E}[\langle U_{\sqrt{\rho/s}} D_w U_s f_i, D_w U_{\sqrt{\rho s}} f_i \rangle] \\ &= - \sum_i \mathbf{E}[\langle D_w U_{\sqrt{\rho s}} f_i, D_w U_{\sqrt{\rho s}} f_i \rangle] \\ &= - \mathbf{E} \|D_w U_{\sqrt{\rho s}} f\|_2^2. \end{aligned}$$

Taking the limit as  $s \rightarrow 1$ , we obtain the identity

$$\sum_i \mathbf{E}[f_i \operatorname{div}_\gamma((D_w U_\rho f_i)w)] = - \sum_i \mathbf{E}[\langle D_w f_i, D_w U_\rho f_i \rangle] = - \mathbf{E} \|D_w U_{\sqrt{\rho}} f\|_2^2$$

for any  $f \in L^2(\gamma)$ . The non-positivity claim follows easily, and it is also clear that zero is attained if and only if  $U_{\sqrt{\rho}} f$  is independent of  $w$ . To see that the same is true for  $f$ , suppose without loss of generality that  $w = e_i$ . Then  $f$  is independent of  $w$  if and only if  $f$ 's Hermite coefficients  $\hat{f}_\alpha$  are zero whenever  $\alpha_i > 0$ . Since  $U_{\sqrt{\rho}}$  acts diagonally and non-degenerately on the Hermite basis,  $f$  is independent of  $w$  if and only if  $U_{\sqrt{\rho}} f$  is.  $\square$

### 6.3 The first-order conditions

To derive the first-order optimality condition (21), we introduce the “value” perturbations described in the outline. For  $f : \mathbb{R}^n \rightarrow \mathbb{R}^k$  and a vector field  $W : \mathbb{R}^n \rightarrow \mathbb{R}^k$ , define (for  $t \in \mathbb{R}$ )

$$(\mathcal{V}_{t,W} f)(x) = \tilde{N}(f(x) + tW(x)),$$

where  $\tilde{N}(x) = x / \max\{1, \|x\|\}$ .

**Lemma 6.7.** *For any measurable  $f : \mathbb{R}^n \rightarrow \mathbb{R}^k$  and any bounded, measurable vector field  $W : \mathbb{R}^n \rightarrow \mathbb{R}^k$ ,*

$$\left. \frac{d}{dt} \right|_{t=0} \mathbf{E}[\mathcal{V}_{t,W} f] = \mathbf{E}[W - \langle f, W \rangle_+ f 1_{\{\|f\|=1\}}],$$

where  $a_+ = \max\{a, 0\}$ .

*Proof.* If  $\|f(x)\| < 1$ ,  $\tilde{N}(f(x) + tW(x)) = f(x) + tW(x)$  for sufficiently small  $t$ . On the other hand, if  $\|f(x)\| = 1$  then Taylor expansion gives

$$\tilde{N}(f(x) + tW(x)) = f(x) + tW(x) - t\langle W(x), f(x) \rangle_+ f(x) + O(t^2) \quad (26)$$

for any  $x \in \mathbb{R}^n$ . The  $O(t^2)$  term is uniform in  $x$  because we assume  $W$  to be uniformly bounded, and hence

$$\frac{\mathbf{E}[\tilde{N}(f + tW) - f]}{t} = \mathbf{E}[W - \langle f, W \rangle_+ f 1_{\{\|f\|=1\}}] + O(t),$$

and the claim follows by taking the limit as  $t \rightarrow 0$ .  $\square$

It is important to note that we can perturb  $\mathbf{E}[f]$  in all possible directions; this is our analogue of the fact that for the finite-dimensional constrained-optimization theory to hold, the constraint function should have a full-rank derivative.

**Lemma 6.8.** *For any measurable  $f : \mathbb{R}^n \rightarrow B^k$ , there exists a set  $W_1, \dots, W_k$  of vector fields such that*

$$\left\{ \frac{d}{dt} \Big|_{t=0} \mathbf{E}[\mathcal{V}_{t, W_i} f] : i = 1, \dots, k \right\}$$

spans  $\mathbb{R}^k$ .

*Proof.* If  $\{x : \|f(x)\| < 1\}$  has positive measure, the claim is clear because for vector fields  $W$  supported on  $\{x : \|f(x)\| < 1\}$ ,  $\frac{d}{dt} \Big|_{t=0} \mathbf{E}[\mathcal{V}_{t, W} f] = \mathbf{E}[W]$ . From now on, we will assume that  $f : \mathbb{R}^n \rightarrow S^{k-1}$ .

First, choose  $v_0 \in S^{k-1}$  belonging to the support of  $f$ . For some sufficiently small  $\epsilon > 0$ , let  $A := \{x : |f(x) - v_0| < \epsilon\}$ , and note that  $A$  has positive measure. Let  $w_1, \dots, w_{k-1}$  be a basis for  $v_0^\perp$  and let  $w_k = -v_0$ ; then  $\{w_1, \dots, w_k\}$  spans  $\mathbb{R}^k$ . Define (for  $i = 1, \dots, k$ )  $W_i = w_i 1_A / \gamma(A)$ .

First, consider any  $i = 1, \dots, k-1$ . Since  $\langle W_i, v_0 \rangle \equiv 0$ , and since  $W_i = 0$  whenever  $|f(x) - v_0| \geq \epsilon$ ,

$$|\langle f(x), W_i(x) \rangle| \leq \epsilon |W_i(x)|$$

for every  $x$ . Therefore,

$$\tilde{w}_i := \frac{d}{dt} \Big|_{t=0} \mathbf{E}[\mathcal{V}_{t, W_i} f] = \mathbf{E}[W_i] - \mathbf{E}[\langle f, W_i \rangle_+ f] = w_i + O(\epsilon).$$

On the other hand, for  $i = k$  we have

$$|\langle f(x), W_k(x) \rangle + 1| \leq \epsilon |W_k(x)|,$$

meaning in particular that  $\langle f, W_k \rangle \leq 0$  pointwise as soon as  $\epsilon < 1$ . Therefore,

$$\tilde{w}_k := \frac{d}{dt} \Big|_{t=0} \mathbf{E}[\mathcal{V}_{t, W_k} f] = \mathbf{E}[W_k] = w_k.$$

Since  $\{\tilde{w}_1, \dots, \tilde{w}_k\}$  is an arbitrarily small perturbation of  $\{w_1, \dots, w_k\}$ , if  $\epsilon > 0$  is sufficiently small then  $\{\tilde{w}_1, \dots, \tilde{w}_k\}$  spans  $\mathbb{R}^k$ .  $\square$

The next step in establishing the first-order conditions is to show that it's enough to consider derivatives: if it's possible to improve the objective to first-order while preserving the constraints to first-order then it's also possible to improve the objective to first-order while preserving the constraints exactly.

**Lemma 6.9.** *If  $\rho \in (0, 1)$  and  $f : \mathbb{R}^n \rightarrow B^k$  is optimally stable then for every bounded, measurable vector field  $W$ ,  $\frac{d}{dt}\Big|_{t=0} \mathbf{E}[\mathcal{V}_{t,W}f] = 0$  implies*

$$\frac{d}{dt}\Big|_{t=0} \mathbf{E}_{\mathbf{x} \sim_{\rho} \mathbf{y}} [\langle (\mathcal{V}_{t,W}f)(\mathbf{x}), (\mathcal{V}_{t,W}f)(\mathbf{y}) \rangle] = 0.$$

*Proof.* Let  $W$  be any vector field with  $\frac{d}{dt}\Big|_{t=0} \mathbf{E}[\mathcal{V}_{t,W}f] = 0$ , and choose vector fields  $W_1, \dots, W_k$  as in Lemma 6.8. Consider the competitor function  $f_{\alpha,\beta}(x) = \tilde{N}(f(x) + \sum_i \alpha_i W_i + \beta W)$ . We define  $L : \mathbb{R}^{k+1} \rightarrow \mathbb{R}^k$  by

$$L(\alpha, \beta) = \mathbf{E}[f_{\alpha,\beta}].$$

Then

$$\frac{\partial L}{\partial \beta}(0, 0) = \frac{d}{dt}\Big|_{t=0} \mathbf{E}[\mathcal{V}_{t,W}f] = 0 \quad (27)$$

and

$$\frac{\partial L}{\partial \alpha_i}(0, 0) = \frac{d}{dt}\Big|_{t=0} \mathbf{E}[\mathcal{V}_{t,W_i}f], \quad (28)$$

which by our choice of  $W_i$  implies that  $DL$  (as a  $(k+1) \times k$  matrix) has rank  $k$ . By the implicit function theorem, there is some interval  $(-\epsilon, \epsilon)$  and a differentiable curve  $\eta : (-\epsilon, \epsilon) \rightarrow \mathbb{R}^{k+1}$  such that  $\eta(0) = 0$ ,  $\eta'(0) \neq 0$  and  $L(\eta(t)) = 0$  for all  $t \in (-\epsilon, \epsilon)$ . We'll write  $\alpha(t)$  for the first  $k$  coordinates of  $\eta$ , and  $\beta(t)$  for the last coordinate.

Now, the fact that  $\eta'(0) \neq 0$  implies that at least one of  $\alpha'(0)$  or  $\beta'(0)$  is non-zero. But the chain rule and the fact that  $L(\eta(t))$  is constant gives

$$0 = \frac{d}{dt}\Big|_{t=0} L(\eta(t)) = \sum_i \frac{\partial L}{\partial \alpha_i}(0, 0) \alpha'_i(0) + \frac{\partial L}{\partial \beta}(0, 0) \beta'(0);$$

the term involving  $\beta$  vanishes because of (27), while (28) implies that the vectors  $\frac{\partial}{\partial \alpha_i} L(0, 0)$  are linearly independent. It follows that  $\alpha'(0) = 0$ , and so then we must have  $\beta'(0) \neq 0$ .

Finally, we consider the objective value

$$J(\alpha, \beta) = \mathbf{E}_{\mathbf{x} \sim_{\rho} \mathbf{y}} [\langle f_{\alpha,\beta}(\mathbf{x}), f_{\alpha,\beta}(\mathbf{y}) \rangle].$$

Since  $f$  is optimally stable and  $f_{\alpha(t),\beta(t)}$  satisfies the constraints,  $\frac{d}{dt}\Big|_{t=0} J(\alpha(t), \beta(t)) = 0$ . On the other hand, the chain rule gives

$$\begin{aligned} 0 = \frac{d}{dt}\Big|_{t=0} J(\alpha(t), \beta(t)) &= \sum_i \alpha'_i(0) \frac{d}{dt}\Big|_{t=0} \mathbf{E}_{\mathbf{x} \sim_{\rho} \mathbf{y}} [\langle (\mathcal{V}_{t,W_i}f)(\mathbf{x}), (\mathcal{V}_{t,W_i}f)(\mathbf{y}) \rangle] \\ &\quad + \beta'(0) \frac{d}{dt}\Big|_{t=0} \mathbf{E}_{\mathbf{x} \sim_{\rho} \mathbf{y}} [\langle (\mathcal{V}_{t,W}f)(\mathbf{x}), (\mathcal{V}_{t,W}f)(\mathbf{y}) \rangle]. \end{aligned}$$

We showed that  $\alpha'_i(0) = 0$  for all  $i$ , and  $\beta'(0) \neq 0$ , we conclude that

$$\frac{d}{dt}\Big|_{t=0} \mathbf{E}_{\mathbf{x} \sim_{\rho} \mathbf{y}} [\langle (\mathcal{V}_{t,W}f)(\mathbf{x}), (\mathcal{V}_{t,W}f)(\mathbf{y}) \rangle] = 0. \quad \square$$

There is also a Lagrangian interpretation of Lemma 6.9:

**Lemma 6.10.** *If  $\rho \in (0, 1)$  and  $f : \mathbb{R}^n \rightarrow B^k$  is optimally stable then there exists some  $\lambda \in \mathbb{R}^k$  such that for every bounded, measurable vector field  $W$ ,*

$$\left. \frac{d}{dt} \right|_{t=0} \mathbf{E}_{\mathbf{x} \sim_{\rho} \mathbf{y}} [(\mathcal{V}_{t,W} f)(\mathbf{x}), (\mathcal{V}_{t,W} f)(\mathbf{y})] = \left\langle \lambda, \left. \frac{d}{dt} \right|_{t=0} \mathbf{E}[\mathcal{V}_{t,W} f] \right\rangle.$$

*Proof.* For a bounded, measurable vector field  $W$ , let

$$\phi(W) = \left. \frac{d}{dt} \right|_{t=0} \mathbf{E}[\mathcal{V}_{t,W} f] \in \mathbb{R}^k$$

and

$$\psi(W) = \left. \frac{d}{dt} \right|_{t=0} \mathbf{E}_{\mathbf{x} \sim_{\rho} \mathbf{y}} [(\mathcal{V}_{t,W} f)(\mathbf{x}), (\mathcal{V}_{t,W} f)(\mathbf{y})] \in \mathbb{R}$$

noting that both  $\phi(W)$  and  $\psi(W)$  are linear functions of  $W$ . Let  $\mathcal{X}$  be any finite-dimensional subspace of bounded measurable vector fields for which  $\{\phi(W) : W \in \mathcal{X}\}$  spans  $\mathbb{R}^k$ , and consider the linear map  $L : \mathcal{X} \rightarrow \mathbb{R}^{k+1}$  given by  $L(W) = (\phi(W), \psi(W))$ . By Lemma 6.9,  $(0, \dots, 0, 1)$  does not belong to the range of  $L$ ; it follows that there exists  $\lambda = \lambda(\mathcal{X})$  such that  $(-\lambda, 1)$  is orthogonal to the range of  $L$  (for example,  $\lambda$  can be found by rescaling the residual of the orthogonal projection of  $(0, \dots, 0, 1)$  onto the range of  $L$ ). For this  $\lambda$ , we have  $\langle \lambda, \phi(W) \rangle = \psi(W)$  for all  $W \in \mathcal{X}$ . Note that  $\lambda(\mathcal{X})$  is unique, because the range of  $L$  has dimension at least  $k$ .

Now, if  $\mathcal{X} \subset \mathcal{X}'$  are two vector spaces satisfying the spanning property above then  $\lambda(\mathcal{X}') = \lambda(\mathcal{X})$  (because  $(-\lambda(\mathcal{X}'), 1)$  is orthogonal to  $L(\mathcal{X}')$  and hence also  $L(\mathcal{X})$ , and  $\lambda(\mathcal{X})$  is the unique vector with that property). It follows then that there is a  $\lambda$  satisfying  $\langle \lambda, \phi(W) \rangle = \psi(W)$  for all bounded, measurable  $W$ : take any  $\mathcal{X}$  for which  $\{\phi(W) : W \in \mathcal{X}\}$  spans  $\mathbb{R}^k$ , and take  $\lambda = \lambda(\mathcal{X})$ . Then for any bounded, measurable  $W$ , consider  $\mathcal{X}' = \text{span}(W \cup \mathcal{X})$ ; since  $\lambda(\mathcal{X}') = \lambda$ , it follows that  $\langle \lambda, \phi(W) \rangle = \psi(W)$ .  $\square$

To make Lemma 6.10 more useful, we test it on “local” vector fields  $W$  to extract valuable *pointwise* information about optimally stable functions. First, note that the Taylor expansion (26) implies that

$$\begin{aligned} \left. \frac{d}{dt} \right|_{t=0} \mathbf{E}_{\mathbf{x} \sim_{\rho} \mathbf{y}} [(\mathcal{V}_{t,W} f)(\mathbf{x}), (\mathcal{V}_{t,W} f)(\mathbf{y})] &= 2 \mathbf{E}_{\mathbf{x} \sim_{\rho} \mathbf{y}} [(f(\mathbf{x}), W(\mathbf{y}) - \langle W(\mathbf{y}), f(\mathbf{y}) \rangle + f(\mathbf{y}))] \\ &= 2 \mathbf{E}[\langle W - \langle W, f \rangle + f, U_{\rho} f \rangle]. \end{aligned}$$

Therefore, Lemma 6.10 implies that there exists  $\lambda \in \mathbb{R}^k$  such that

$$2 \mathbf{E}[\langle W - \langle W, f \rangle + f, U_{\rho} f - \lambda/2 \rangle] = 0 \tag{29}$$

for every bounded, measurable  $W$ .

**Lemma 6.11.** *If  $\rho \in (0, 1)$  and  $f : \mathbb{R}^k \rightarrow B^k$  is optimally stable then for the  $\lambda$  of Lemma 6.10, we have*

$$|U_{\rho} f - \lambda/2| f = U_{\rho} f - \lambda/2 \text{ a.e.}$$

*Proof.* Suppose not, and choose some small  $\epsilon > 0$  such that the set

$$A := \{x : |U_{\rho} f - \lambda/2| > \epsilon \text{ and } |f - N(U_{\rho} f - \lambda/2)| > \epsilon\}$$

has positive measure. Then find some  $v \in B^k$  such that

$$C := \{x : x \in A \text{ and } |f(x) - v| < \epsilon^3\}$$

has positive measure. Since  $|f - N(U_\rho f - \lambda/2)| > \epsilon$  on  $C$ , it follows that  $|v - N(U_\rho f - \lambda/2)| > \epsilon - \epsilon^3$  on  $C$ , and so (for small enough  $\epsilon > 0$ ) we can find some unit vector  $w \in v^\perp$  such that  $\langle N(U_\rho f - \lambda/2), w \rangle \geq \epsilon/2$  on  $C$ . Now set  $W = w1_C$ .

On the set  $C$ ,  $w \in v^\perp$  and  $|f(x) - v| \leq \epsilon^3$  imply that  $|\langle W, f \rangle + f| \leq \epsilon^3$ . On the other hand, we also have (still on the set  $C$ )

$$\langle W, U_\rho f - \lambda/2 \rangle = \langle W, N(U_\rho f - \lambda/2) \rangle |U_\rho f - \lambda/2| \geq \frac{\epsilon}{2} |U_\rho f - \lambda/2| \geq \frac{\epsilon^2}{2},$$

and if  $\epsilon > 0$  is small enough then this contradicts (29).  $\square$

## 6.4 Spatial perturbations

Next, we compute the first and second derivatives of the objectives and constraints for the spatial perturbations  $\mathcal{S}_{t,W}f$  which, recall, is defined by letting  $\{F_t : t \in \mathbb{R}\}$  be the flow along  $W$ , and setting

$$(\mathcal{S}_{t,W}f)(x) = f(F_t^{-1}(x)).$$

**Lemma 6.12.** *For any bounded function  $f : \mathbb{R}^n \rightarrow \mathbb{R}^k$  and any tame vector field  $W$ ,  $\mathbf{E}[\mathcal{S}_{t,W}f]$  is differentiable in  $t$  and satisfies*

$$\left. \frac{d}{dt} \right|_{t=0} \mathbf{E}[\mathcal{S}_{t,W}f] = \mathbf{E}[f \operatorname{div}_\gamma W].$$

*Proof.* For  $f \in \mathcal{C}_c^1$ , this follows by writing out the definition of  $\mathcal{S}_{t,W}f$ , differentiating inside the integral, and integrating by parts using (23):

$$\left. \frac{d}{dt} \right|_{t=0} \mathbf{E}[\mathcal{S}_{t,W}f] = \left. \frac{d}{dt} \right|_{t=0} \mathbf{E}[f \circ F_{-t}] = -\mathbf{E}[D_W f] = \mathbf{E}[f \operatorname{div}_\gamma W].$$

Because  $\mathcal{S}_{s,W}\mathcal{S}_{t,W}f = \mathcal{S}_{s+t,W}f$ , this also implies that

$$\frac{d}{dt} \mathbf{E}[\mathcal{S}_{t,W}f] = \mathbf{E}[\mathcal{S}_{t,W}f \operatorname{div}_\gamma W] \quad (30)$$

for  $f \in \mathcal{C}_c^1$ .

Next we handle the case of general bounded  $f$ . Take an approximating sequence  $f_n$  as in Lemma 6.3. Defining  $\phi(t) = \mathbf{E}[\mathcal{S}_{t,W}f]$  and  $\phi_n(t) = \mathbf{E}[\mathcal{S}_{t,W}f_n]$ , we see from (30) and the uniform boundedness of  $f_n$  that  $\phi'_n(t)$  is continuous in  $t$ , and bounded uniformly in  $n$  and  $t$ . Moreover, Lemma 6.3 ensures that  $\phi_n(t) \rightarrow \phi(t)$  uniformly for  $t \in [-1, 1]$ , and that  $\phi'_n(t)$  converges uniformly. It follows that  $\phi(t)$  is differentiable in  $t$  and satisfies

$$\phi'(0) = \lim_{n \rightarrow \infty} \phi'_n(0) = \mathbf{E}[f \operatorname{div}_\gamma W].$$

$\square$

Next, we do a similar computation for the objective function:

**Lemma 6.13.** *For any bounded, measurable function  $f : \mathbb{R}^n \rightarrow \mathbb{R}^k$  and any tame vector field  $W$ ,  $\mathbf{E}_{\mathbf{x} \sim_\rho \mathbf{y}}[\langle \mathcal{S}_{t,W}f(\mathbf{x}), \mathcal{S}_{t,W}f(\mathbf{y}) \rangle]$  is differentiable in  $t$  and satisfies*

$$\left. \frac{d}{dt} \right|_{t=0} \mathbf{E}_{\mathbf{x} \sim_\rho \mathbf{y}}[\langle \mathcal{S}_{t,W}f(\mathbf{x}), \mathcal{S}_{t,W}f(\mathbf{y}) \rangle] = 2 \mathbf{E}[\langle f, U_\rho f \rangle \operatorname{div}_\gamma W + \langle f, D_W U_\rho f \rangle].$$

*Proof.* If  $f$  is  $\mathcal{C}_c^\infty$ , we compute by calculus that

$$\begin{aligned} \frac{d}{dt} \Big|_{t=0} \mathbf{E}_{\mathbf{x} \sim_\rho \mathbf{y}} [\langle \mathcal{S}_{t,W} f(\mathbf{x}), \mathcal{S}_{t,W} f(\mathbf{y}) \rangle] &= -2 \mathbf{E}_{\mathbf{x} \sim_\rho \mathbf{y}} [\langle D_W f(\mathbf{x}), f(\mathbf{y}) \rangle] \\ &= -2 \mathbf{E}[\langle D_W f, U_\rho f \rangle] \end{aligned}$$

For each coordinate, we integrate by parts using (23):

$$-2 \mathbf{E}[D_W f_i \cdot U_\rho f_i] = 2 \mathbf{E}[f_i \operatorname{div}_\gamma((U_\rho f_i)W)] = 2 \mathbf{E}[f_i U_\rho f_i \operatorname{div}_\gamma(W) + f_i D_W U_\rho f_i].$$

Summing over  $i$  completes the proof for  $f \in \mathcal{C}_c^\infty$ ; note that because  $\mathcal{S}_{s,W} \mathcal{S}_{t,W} f = \mathcal{S}_{s+t,W} f$ , we also have the derivative at  $t \neq 0$ :

$$\frac{d}{dt} \mathbf{E}_{\mathbf{x} \sim_\rho \mathbf{y}} [\langle \mathcal{S}_{t,W} f(\mathbf{x}), \mathcal{S}_{t,W} f(\mathbf{y}) \rangle] = 2 \mathbf{E}[\langle \mathcal{S}_{t,W} f, U_\rho \mathcal{S}_{t,W} f \rangle \operatorname{div}_\gamma W + \langle \mathcal{S}_{t,W} f, D_W U_\rho \mathcal{S}_{t,W} f \rangle].$$

Now consider a bounded function  $f$  and choose an approximating sequence  $f_n$  as in Lemma 6.3. Letting  $\phi(t) = \mathbf{E}_{\mathbf{x} \sim_\rho \mathbf{y}} [\langle \mathcal{S}_{t,W} f(\mathbf{x}), \mathcal{S}_{t,W} f(\mathbf{y}) \rangle]$  and  $\phi_n(t) = \mathbf{E}_{\mathbf{x} \sim_\rho \mathbf{y}} [\langle \mathcal{S}_{t,W} f_n(\mathbf{x}), \mathcal{S}_{t,W} f_n(\mathbf{y}) \rangle]$ , we note from the formula above that  $\phi'_n(t)$  is uniformly bounded and converging uniformly (for  $t \in [-1, 1]$ ), where the boundedness of the term involving  $D_W U_\rho \mathcal{S}_{t,W} f_n$  follows from Lemma 6.5). Since  $\phi_n(t) \rightarrow \phi(t)$  uniformly for  $t \in [-1, 1]$ , it follows that  $\phi(t)$  is differentiable and  $\phi'(0) = \lim_{n \rightarrow \infty} \phi'_n(0)$ .  $\square$

## 6.5 Second variation

Here, we establish the second-order optimality condition for spatial perturbations. First, let us observe that everything is twice differentiable in  $t$ :

**Lemma 6.14.** *For any measurable  $f : \mathbb{R}^n \rightarrow B^k$  and any tame vector field  $W$ , both  $\mathbf{E}[\mathcal{S}_{t,W} f]$  and  $\mathbf{E}_{\mathbf{x} \sim_\rho \mathbf{y}} [\langle \mathcal{S}_{t,W} f(\mathbf{x}), \mathcal{S}_{t,W} f(\mathbf{y}) \rangle]$  are twice differentiable in  $t$ .*

*Proof.* This can be seen simply by writing out the definitions and changing variables so that the derivatives fall only on the Gaussian kernel, much as in the proof of Lemma 6.5.  $\square$

**Lemma 6.15.** *Suppose  $f$  is optimally stable. If  $\rho > 0$  and  $W$  is a tame vector field such that*

$$\frac{d}{dt} \Big|_{t=0} \mathbf{E}[\mathcal{S}_{t,W} f] = 0,$$

then

$$\frac{d^2}{dt^2} \Big|_{t=0} \mathbf{E}_{\mathbf{x} \sim_\rho \mathbf{y}} [\langle \mathcal{S}_{t,W} f(\mathbf{x}), \mathcal{S}_{t,W} f(\mathbf{y}) \rangle] - \left\langle \lambda, \frac{d^2}{dt^2} \Big|_{t=0} \mathbf{E}[\mathcal{S}_{t,W} f] \right\rangle \leq 0.$$

If  $\rho < 0$  then under the same assumptions, the left hand side above is at least zero.

This motivates the definition:

**Definition 6.16.** For a tame vector field  $W$ , define the *index form*

$$Q(W) = \frac{d^2}{dt^2} \Big|_{t=0} \mathbf{E}_{\mathbf{x} \sim_\rho \mathbf{y}} [\langle \mathcal{S}_{t,W} f(\mathbf{x}), \mathcal{S}_{t,W} f(\mathbf{y}) \rangle] - \left\langle \lambda, \frac{d^2}{dt^2} \Big|_{t=0} \mathbf{E}[\mathcal{S}_{t,W} f] \right\rangle$$

*Proof.* Take vector fields  $W_1, \dots, W_m$  as in the proof of Lemma 6.9, and define

$$f_{\alpha, \beta}(x) = N(f(F_\beta(x))) + \sum_i \alpha_i W_i(x),$$

where  $F_\beta$  is the flow along the vector field  $W$ . As in the proof of Lemma 6.9, defining  $L(\alpha, \beta) = \mathbf{E}[f_{\alpha, \beta}]$  implies that  $\frac{\partial L}{\partial \beta}(0, 0) = 0$ , while  $DL(0, 0)$  has rank  $k$ . Therefore (as in the proof of Lemma 6.9) we can find smooth curves  $\alpha(t) \in \mathbb{R}^k$  and  $\beta(t) \in \mathbb{R}$  such that  $\alpha'(0) = 0$ ,  $\beta'(0) \neq 0$ , and  $L(\alpha(t), \beta(t)) \equiv 0$  for  $t$  in some interval  $(-\epsilon, \epsilon)$ . Taking second derivatives with respect to  $t$ , we have

$$0 = \frac{d^2}{dt^2} \Big|_{t=0} L(\alpha(t), \beta(t)) = \sum_i \alpha_i''(0) \frac{\partial L}{\partial \alpha_i}(0, 0) + (\beta'(0))^2 \frac{\partial^2 L}{\partial \beta^2}(0, 0) \quad (31)$$

Define  $J(\alpha, \beta) = \mathbf{E}_{\mathbf{x} \sim_\rho \mathbf{y}} \langle f_{\alpha, \beta}(\mathbf{x}), f_{\alpha, \beta}(\mathbf{y}) \rangle$  and let  $K(t) = J(\alpha(t), \beta(t))$ . First, note that (because  $\alpha'(0) = 0$ )  $K'(0) = \beta'(0) \frac{\partial J}{\partial \beta}(0, 0)$ . Since  $\beta'(0) \neq 0$ , we must have  $\frac{\partial J}{\partial \beta}(0, 0) = 0$  – otherwise, there would be some small  $t$  (either positive or negative) giving a contradiction to the optimality of  $f$ . Taking another derivative, the optimality of  $f$  implies that

$$0 \geq K''(0) = \sum_i \alpha_i''(0) \frac{\partial J}{\partial \alpha_i}(0, 0) + (\beta'(0))^2 \frac{\partial^2 J}{\partial \beta^2}(0, 0).$$

Finally, recall from Lemma 6.10 that  $\frac{\partial J}{\partial \alpha_i} = \langle \lambda, \frac{\partial L}{\partial \alpha_i} \rangle$ ; going back to (31), we obtain

$$0 \geq K''(0) = -(\beta'(0))^2 \left\langle \lambda, \frac{\partial^2 L}{\partial \beta^2}(0, 0) \right\rangle + (\beta'(0))^2 \frac{\partial^2 J}{\partial \beta^2}(0, 0);$$

dropping the (positive)  $(\beta'(0))^2$  terms and untangling the notation, this is equivalent to the claim.  $\square$

## 6.6 The index form for translations

Here, we compute the index form  $Q$  for constant vector fields  $W \equiv w$ , giving the rigorous, integrated-by-parts analogue of (22) (at least, for  $W \equiv w$ , which is all that we will need).

**Lemma 6.17.** *For any measurable  $f : \mathbb{R}^n \rightarrow B^k$  and any  $w \in \mathbb{R}^n$ ,*

$$Q(w) = 2 \sum_i \left( \mathbf{E}[f_i \operatorname{div}_\gamma(\operatorname{div}_\gamma((U_\rho f_i - \lambda/2)w))] - \frac{1}{\rho} \mathbf{E}[f_i \operatorname{div}_\gamma((D_w U_\rho f_i)w)] \right).$$

*Proof.* Note that  $\mathcal{S}_{t,w}f(x) = f(x - tw)$ . If  $f \in C^\infty$ , we simply compute

$$\frac{d^2}{dt^2} \Big|_{t=0} \mathbf{E}[\mathcal{S}_{t,w}f] = \mathbf{E}[D_w(D_w f)]$$

and

$$\begin{aligned} & \frac{d^2}{dt^2} \Big|_{t=0} \mathbf{E}_{\mathbf{x} \sim_\rho \mathbf{y}} [\langle \mathcal{S}_{t,w}f(\mathbf{x}), \mathcal{S}_{t,w}f(\mathbf{y}) \rangle] \\ &= 2 \mathbf{E}[\langle D_{w,w}^2 f(\mathbf{x}), f(\mathbf{y}) \rangle + \langle D_w f(\mathbf{x}), D_w f(\mathbf{y}) \rangle] \\ &= 2 \mathbf{E}[\langle D_{w,w}^2 f, U_\rho f \rangle + \langle D_w f, U_\rho D_w f \rangle] \\ &= 2 \mathbf{E}[\langle D_{w,w}^2 f, U_\rho f \rangle] + \frac{2}{\rho} \mathbf{E}[\langle D_w f, D_w U_\rho f \rangle], \end{aligned}$$

and hence

$$Q(w) = 2 \mathbf{E}[\langle D_{w,w}^2 f, U_\rho f - \lambda/2 \rangle] + \frac{2}{\rho} \mathbf{E}[\langle D_w f, D_w U_\rho f \rangle]. \quad (32)$$

Integrating the first term by parts twice gives

$$\begin{aligned} \mathbf{E}[D_{w,w}^2 f_i \cdot (U_\rho f_i - \lambda_i/2)] &= - \mathbf{E}[D_w f_i \cdot \operatorname{div}_\gamma((U_\rho f_i - \lambda_i/2)w)] \\ &= \mathbf{E}[f_i \operatorname{div}_\gamma(\operatorname{div}_\gamma((U_\rho f_i - \lambda_i/2)w)w)]; \end{aligned}$$

integrating the second term of (32) by parts gives

$$\mathbf{E}[D_w f_i \cdot D_w U_\rho f_i] = - \mathbf{E}[f_i \operatorname{div}_\gamma((D_w U_\rho f_i)w)].$$

Overall, we obtain

$$Q(w) = 2 \sum_i \mathbf{E} \left[ f_i \operatorname{div}_\gamma(\operatorname{div}_\gamma((U_\rho f_i - \lambda_i/2)w)w) - \frac{1}{\rho} f_i \operatorname{div}_\gamma((D_w U_\rho f_i)w) \right].$$

By the familiar approximation argument (noting that the terms involving second derivatives of  $U_\rho f_i$  are controlled by Lemma 6.5), the same formula applies for all bounded, measurable functions  $f$ .  $\square$

Our formula for  $Q(w)$  in Lemma 6.17 doesn't require  $f$  to be optimally stable. However, if  $f$  is optimally stable, the first-order conditions allow us to find a simpler formula.

**Lemma 6.18.** *If  $f : \mathbb{R}^n \rightarrow B^k$  is optimally stable then for any  $w \in \mathbb{R}^n$ ,*

$$Q(w) = 2 \frac{\rho - 1}{\rho} \sum_i \mathbf{E}[f_i \operatorname{div}_\gamma((D_w U_\rho f_i)w)].$$

*Proof.* The point is to show that

$$\sum_i \mathbf{E}[f_i \operatorname{div}_\gamma(\operatorname{div}_\gamma((U_\rho f_i - \lambda_i/2)w)w)] = \sum_i \mathbf{E}[f_i \operatorname{div}_\gamma((D_w U_\rho f_i)w)];$$

then the claim follows immediately. To show the identity above, note that (by the product rule)

$$\operatorname{div}_\gamma((U_\rho f_i - \lambda_i/2)w) = (U_\rho f_i - \lambda_i/2) \operatorname{div}_\gamma w + D_w U_\rho f_i;$$

plugging this in above, it suffices to show that

$$\sum_i \mathbf{E}[f_i \operatorname{div}_\gamma((U_\rho f_i - \lambda_i/2) \operatorname{div}_\gamma w \cdot w)] = 0 \quad (33)$$

But Lemma 6.10 implies that  $U_\rho f - \lambda/2 = |U_\rho f - \lambda/2|f$ ; (33) then follows from Lemma 6.4 with  $\psi = |U_\rho f - \lambda/2|$  and  $W = (\operatorname{div}_\gamma w)w$ .  $\square$

*Proof of Theorem 6.1.* Suppose  $f$  is optimally stable and suppose that  $\rho \in (0, 1)$ . By Lemma 6.15, for any  $w$  with

$$\left. \frac{d}{dt} \right|_{t=0} \mathbf{E}[\mathcal{S}_{t,w} f] = 0,$$

we have  $Q(w) \leq 0$ . Using the formula for  $Q$  in Lemma 6.18, for such  $w$  we have

$$\sum_i \mathbf{E}[f_i \operatorname{div}_\gamma((D_w U_\rho f_i)w)] \geq 0.$$

But then Lemma 6.6 implies that  $f(x)$  can be written as a function of  $\Pi_{w^\perp}x$ .

Note that the map

$$L(w) = \left. \frac{d}{dt} \right|_{t=0} \mathbf{E}[\mathcal{S}_{t,w}f]$$

is a linear map  $\mathbb{R}^n \rightarrow \mathbb{R}^k$ . Then  $\ker L$  has dimension at least  $n - k$ . After applying a change of coordinates in  $\mathbb{R}^n$ , we may assume that  $\ker L$  contains the span of  $e_{k+1}, \dots, e_n$ ; then the previous paragraph implies that  $f(x)$  depends only on  $x_1, \dots, x_m$ .  $\square$

## 6.7 The case of negative $\rho$

Many of the technical results we developed above apply to the case of negative  $\rho$ , with a few sign changes. Notably, the sign in Lemma 6.11 changes to

$$|U_\rho f - \lambda/2|f = \lambda/2 - U_\rho f;$$

and because the negative- $\rho$  case is a minimization problem instead of a maximization problem, the sign of the second-order conditions flips also: if  $\rho < 0$  then the final inequality of Lemma 6.15 is reversed.

Because of these sign changes, the constant vector fields  $w$  turn out not to contradict any stability. In fact, with  $\rho < 0$  then  $Q(w) \leq 0$  for every  $w$ , whether or not  $\mathcal{S}_{t,w}f$  preserves expectations to first-order. It remains plausible that there are some other vector fields that will imply Theorem 6.1 in the case of negative  $\rho$ , but we were unable to find them.

## Part III

# Hardness of QUANTUM MAX-CUT

The goal of this part is to use the results from [Part II](#) to derive hardness of approximation results for QUANTUM MAX-CUT assuming the vector-valued Borell’s inequality. This part is organized as follows: we begin with preliminaries in [Section 7](#). Then we show our integrality gaps in [Section 8](#) and our algorithmic gap in [Section 9](#). Finally, in [Section 10](#) we develop a *dictator test* for product states, which is a crucial ingredient in our hardness proof, contained in [Section 11](#).

## 7 Preliminaries

### 7.1 Pauli matrices

**Definition 7.1** (Pauli matrices). The *Pauli matrices* are the Hermitian matrices

$$X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad Y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

**Notation 7.2.** We will generally use  $P$  or  $Q$  for a variable in  $\{I, X, Y, Z\}$ .

**Proposition 7.3** (Properties of the Pauli matrices). *The Pauli matrices have the following properties.*

1.  $X^2 = Y^2 = Z^2 = I$ .
2.  $XY = iZ$ ,  $YZ = iX$ , and  $ZX = iY$ .
3.  $XY = -YX$ ,  $YZ = -ZY$ , and  $ZX = -XZ$ .
4.  $\text{tr}[X] = \text{tr}[Y] = \text{tr}[Z] = 0$ .

### 7.2 On QUANTUM MAX-CUT and the Heisenberg model

The *quantum Heisenberg model* is a family of Hamiltonians first studied by Heisenberg in [\[Hei28\]](#). Given an unweighted graph  $G = (V, E)$ , a Hamiltonian from this model is written as

$$H = - \sum_{(u,v) \in E} (J_X \cdot X_u X_v + J_Y \cdot Y_u Y_v + J_Z \cdot Z_u Z_v) - m \sum_{u \in V} Z_u,$$

where  $P_u$  for  $P \in \{X, Y, Z\}$  refers to the Pauli matrix  $P$  applied to the  $u$ -th qubit,  $J_X, J_Y, J_Z$  are real-valued coefficients known as *coupling constants*, and  $m$  is a real-valued coefficient known as the *external magnetic field*. As is typical in Hamiltonian complexity, and unlike in QUANTUM MAX-CUT, the ground state energy of this Hamiltonian is defined to be its *minimum* (rather than maximum) eigenvalue, and the ground state is defined to be the corresponding eigenvector. The *ferromagnetic case* refers to the case when  $J_X, J_Y, J_Z \geq 0$ , in which case neighboring qubits tend to have the same values in the  $X, Y, Z$  bases, and the *anti-ferromagnetic case* is when  $J_X, J_Y, J_Z \leq 0$ , in which case they have opposing values.

The *anti-ferromagnetic Heisenberg XYZ model* (which we will henceforth simply refer to as the “Heisenberg model”), is the case when  $J_X = J_Y = J_Z = -1$  and  $m = 0$ . It is natural to allow for

the graph  $G = (V, E, w)$  to be weighted, in which case we can write a Hamiltonian from this model as

$$H_G^{\text{HEIS}} = \mathbf{E}_{(\mathbf{u}, \mathbf{v}) \sim E} [X_{\mathbf{u}} X_{\mathbf{v}} + Y_{\mathbf{u}} Y_{\mathbf{v}} + Z_{\mathbf{u}} Z_{\mathbf{v}}].$$

As we have mentioned before, QUANTUM MAX-CUT was defined to be a natural maximization version of the Heisenberg model. Indeed, if  $H_G$  is the QUANTUM MAX-CUT instance corresponding to  $G$ , then  $H_G = (I - H_G^{\text{HEIS}})/4$ . This means that if  $|\psi\rangle$  is the minimum energy state of  $H_G^{\text{HEIS}}$  and has energy  $\nu$ , then  $|\psi\rangle$  is also the *maximum* energy state of  $H_G$  and has energy  $(1 - \nu)/4$ .

Where the two variants differ is in their approximability. As we have seen throughout this work, one can achieve a constant-factor approximation to the QUANTUM MAX-CUT objective in polynomial time. On the other hand, the best known approximation algorithm for the Heisenberg model objective is due to [BGKT19] and achieves a  $1/O(\log(n))$  approximation. In particular, if the minimum energy of  $H_G^{\text{HEIS}}$  is  $\nu$ , this algorithm finds a product state with energy no greater than  $\nu/O(\log(n))$ . The source of this difference is the identity term  $I \otimes I$  in the QUANTUM MAX-CUT objective, which “inflates” the energy of a state relative to its energy in the Heisenberg model. For example, a tensor product of maximally mixed qubits always has objective value  $1/4$  in QUANTUM MAX-CUT due to these identity terms (which, in turn, implies one can always trivially achieve an approximation ratio of  $1/4$ ). In the Heisenberg model, however, its objective value is 0, and so it gives no approximation to the optimum value. An analogous situation occurs in the classical world, where the MAX-CUT objective

$$\max_{f: V \rightarrow \{-1, 1\}} \mathbf{E}_{(\mathbf{u}, \mathbf{v}) \sim E} \left[ \frac{1}{2} - \frac{1}{2} f(\mathbf{u}) f(\mathbf{v}) \right]$$

has a constant-factor 0.878567-approximation [GW95], but the shifted and rescaled objective

$$\min_{f: V \rightarrow \{-1, 1\}} \mathbf{E}_{(\mathbf{u}, \mathbf{v}) \sim E} [f(\mathbf{u}) f(\mathbf{v})]$$

gives us the (anti-ferromagnetic) *Ising model* problem, for which the best-known algorithm is due to Charikar and Wirth [CW04] and achieves an approximation ratio of  $1/O(\log(n))$ .

The Heisenberg model is “notoriously difficult to solve even on bipartite graphs, in contrast to MAX-CUT” [GP19]. Only a few explicit solutions have been found, several of which are well-known results in the physics literature. These include the Heisenberg model on the cycle graph, whose solution due to Bethe is known as the “Bethe ansatz” [Bet31], and on the complete bipartite graph, known as the “Lieb-Mattis model” [LM62]. To our knowledge, [CM16, Section 5.2] contains a complete list of known explicit solutions. This difficulty of the finding solutions for the Heisenberg model was explained by the works of [CM16, PM17], who showed that it is a QMA-complete problem. This implies that QUANTUM MAX-CUT is also QMA-complete.

### 7.3 Alternative expressions for the QUANTUM MAX-CUT interaction

There are several alternative ways of writing the QUANTUM MAX-CUT interaction

$$h = \frac{1}{4}(I \otimes I - X \otimes X - Y \otimes Y - Z \otimes Z)$$

which are common in the literature. The first involves the singlet state.

**Definition 7.4** (Singlet state). The *two-qubit singlet state* is

$$|s\rangle = \frac{1}{\sqrt{2}} |01\rangle - \frac{1}{\sqrt{2}} |10\rangle.$$

It is also known as the *two-qubit anti-symmetric state*, and as the element  $|\Psi^-\rangle$  of the *Bell basis* of two-qubit states

$$|\Phi^\pm\rangle = \frac{1}{\sqrt{2}}|00\rangle \pm \frac{1}{\sqrt{2}}|11\rangle, \quad |\Psi^\pm\rangle = \frac{1}{\sqrt{2}}|01\rangle \pm \frac{1}{\sqrt{2}}|10\rangle.$$

The following proposition gives a convenient expression for the QUANTUM MAX-CUT interaction as the projector on the singlet state.

**Proposition 7.5** (Rewriting the QUANTUM MAX-CUT interaction).

$$h = \frac{1}{4} \cdot (I \otimes I - X \otimes X) \cdot (I \otimes I - Z \otimes Z) = |s\rangle\langle s|.$$

*Proof.* The first equality follows from  $(X \otimes X)(Z \otimes Z) = -Y \otimes Y$ . We verify the second equality by checking that both sides have the same eigendecomposition. Let

$$\frac{1}{\sqrt{2}}|0, a\rangle + \frac{1}{\sqrt{2}}(-1)^b|1, 1+a\rangle \tag{34}$$

be a member of the Bell basis, for  $a, b \in \{0, 1\}$ . This is a 1-eigenvector of  $|s\rangle\langle s|$  if  $a, b = 1$  and a 0-eigenvector otherwise. Now we verify this holds for the LHS:

$$\begin{aligned} & \frac{1}{4} \cdot (I \otimes I - X \otimes X) \cdot (I \otimes I - Z \otimes Z) \cdot \left( \frac{1}{\sqrt{2}}|0, a\rangle + \frac{1}{\sqrt{2}}(-1)^b|1, 1+a\rangle \right) \\ &= \frac{1}{4\sqrt{2}} \cdot (I \otimes I - X \otimes X) \cdot (|0, a\rangle - (-1)^a|0, a\rangle + (-1)^b|1, 1+a\rangle - (-1)^b \cdot (-1)^{2+a}|1, 1+a\rangle) \\ &= \frac{1}{4\sqrt{2}} \cdot (1 - (-1)^a) \cdot (I \otimes I - X \otimes X) \cdot (|0, a\rangle + (-1)^b|1, 1+a\rangle) \\ &= \frac{1}{4\sqrt{2}} \cdot (1 - (-1)^a) \cdot (|0, a\rangle - |1, 1+a\rangle + (-1)^b|1, 1+a\rangle - (-1)^b|0, a\rangle) \\ &= \frac{1}{4\sqrt{2}} \cdot (1 - (-1)^a) \cdot (1 - (-1)^b) \cdot (|0, a\rangle - |1, 1+a\rangle). \end{aligned}$$

If  $a, b = 1$ , then this is equal to (34), showing that it is a 1-eigenvector. Otherwise, this is zero, which completes the proof.  $\square$

Though we will not need it, we note the following additional way of rewriting the QUANTUM MAX-CUT interaction for didactic purposes. The proof is left to the reader.

**Proposition 7.6.**

$$h = \frac{1}{2} \cdot (I \otimes I - \text{SWAP}),$$

where SWAP is the two-qubit swap gate.

## 7.4 Product states

The following definition gives a convenient decomposition for single-qubit quantum states. It can be derived using the properties in Proposition 7.3.

**Definition 7.7** (Bloch spheres and Bloch vectors). Let  $\rho$  be a one qubit density matrix. Then there exists a coefficient vector  $c = (c_X, c_Y, c_Z) \in B^3$  such that

$$\rho = \frac{1}{2} \cdot (I + c_X X + c_Y Y + c_Z Z).$$

In addition,  $\rho$  is a pure state if and only if  $\|c\| = 1$ ; equivalently, if  $c \in S^2$ . We'll refer to the vector  $c$  as the *Bloch vector* for  $\rho$ .

Using this, we can now prove the alternate form for the product state value.

**Proposition 7.8** (Rewriting the product state value; [Proposition 2.9](#) restated).

$$\text{PROD}(G) = \max_{f:V \rightarrow S^2} \mathbf{E}_{(\mathbf{u}, \mathbf{v}) \sim E} \left[ \frac{1}{4} - \frac{1}{4} \langle f(\mathbf{u}), f(\mathbf{v}) \rangle \right].$$

*Proof.* Let  $|\psi_G\rangle = \otimes_{v \in V} |\psi_v\rangle$  be a product state. For each  $v \in V$ , let  $f(v) = (v_X, v_Y, v_Z) \in S^2$  be its Bloch sphere coefficient vector. In other words,

$$|\psi_v\rangle \langle \psi_v| = \frac{1}{2} \cdot (I + v_X X + v_Y Y + v_Z Z).$$

Then the energy of  $|\psi_G\rangle$  is given by

$$\text{tr}[H_G \cdot |\psi_G\rangle \langle \psi_G|] = \mathbf{E}_{(\mathbf{u}, \mathbf{v}) \sim E} \text{tr}[h_{\mathbf{u}, \mathbf{v}} \cdot |\psi_G\rangle \langle \psi_G|] = \mathbf{E}_{(\mathbf{u}, \mathbf{v}) \sim E} \text{tr}[h_{\mathbf{u}, \mathbf{v}} \cdot |\psi_{\mathbf{u}}\rangle \langle \psi_{\mathbf{u}}| \otimes |\psi_{\mathbf{v}}\rangle \langle \psi_{\mathbf{v}}|]. \quad (35)$$

For any edge  $(u, v) \in E$ , the energy of the  $(u, v)$ -interaction is

$$\begin{aligned} & \text{tr}[h_{u, v} \cdot |\psi_u\rangle \langle \psi_u| \otimes |\psi_v\rangle \langle \psi_v|] \\ &= \text{tr} \left[ \frac{1}{4} \cdot (I \otimes I - X \otimes X - Y \otimes Y - Z \otimes Z) \right. \\ & \quad \left. \cdot \frac{1}{2} \cdot (I + u_X X + u_Y Y + u_Z Z) \otimes \frac{1}{2} \cdot (I + v_X X + v_Y Y + v_Z Z) \right] \\ &= \frac{1}{4} \cdot (1 - u_X v_X - u_Y v_Y - u_Z v_Z) \quad (\text{by } \text{Proposition 7.3}) \\ &= \frac{1}{4} \cdot (1 - \langle f(u), f(v) \rangle). \end{aligned}$$

Substituting into [Equation \(35\)](#), the energy of  $|\psi_G\rangle$  is

$$\mathbf{E}_{(\mathbf{u}, \mathbf{v}) \sim E} \left[ \frac{1}{4} - \frac{1}{4} \langle f(\mathbf{u}), f(\mathbf{v}) \rangle \right].$$

This concludes the proof.  $\square$

**Remark 7.9.** One consequence of [Proposition 7.8](#) is that  $\text{PROD}(G) \leq 1/2$  always, even though  $\text{QMAX-CUT}(G)$  can be as large as 1. Indeed,  $\langle \psi_1, \psi_2 | h | \psi_1, \psi_2 \rangle \leq 1/2$  for any qubit states  $|\psi_1\rangle, |\psi_2\rangle \in \mathbb{C}^2$ , even though  $\langle s | h | s \rangle = 1$  by [Proposition 7.5](#).

Although [Remark 7.9](#) shows that the product states can in general give a poor approximation to the energy, there are interesting special cases in which they still give good approximations. The following result shows that this holds provided that the degree of  $G$  is large.

**Theorem 7.10** (Corollary 4 of [\[BH16\]](#)). *Let  $G = (V, E, w)$  be a  $D$ -regular graph with uniform edge weights. Then*

$$\text{PROD}(G) \geq \text{QMAX-CUT}(G) - O\left(\frac{1}{D^{1/3}}\right).$$

Unfortunately, we will not be able to apply this theorem directly because our graphs will not be precisely unweighted,  $D$ -regular graphs, but instead high-degree graphs with different weights on different edges. In this case, Brandão and Harrow provide the following bound.

**Theorem 7.11** (Theorem 8 of [\[BH16\]](#)). *Let  $G = (V, E, w)$  be a weighted graph. Define*

1. the probability distribution  $(p_u)_{u \in V}$  such that  $p_u = \frac{1}{2} \Pr_{e \sim E}[e \text{ contains } u]$ ,

2. the  $|V| \times |V|$  matrix  $A$  such that  $A_{u,v} = \Pr_{(\mathbf{u}', \mathbf{v}') \sim E}[\mathbf{u}' = u \mid \mathbf{v}' = v]$ .

Then the following inequality holds.

$$\text{QMAX-CUT}(G) \leq \text{PROD}(G) + 20 \cdot (\text{tr}[A^2] \|p\|_2^2)^{1/8} + \|p\|_2^2.$$

We note that in the case of a  $D$ -regular graph,  $\text{tr}[A^2] = n/D$  and  $\|p\|_2^2 = 1/n$ . We will use the following corollary, which we will find easier to apply.

**Corollary 7.12.** *In the setting of [Theorem 7.11](#), the following inequality holds.*

$$\text{QMAX-CUT}(G) \leq \text{PROD}(G) + 20 \cdot (n \cdot \max_{u,v} \{A_{u,v}\} \cdot \max_u \{p_u\})^{1/8} + \max_u \{p_u\},$$

where  $n$  is the number of vertices in  $G$ .

*Proof.* First, we bound the  $\text{tr}[A^2]$  term:

$$\text{tr}[A^2] = \sum_{u,v} A_{u,v} \cdot A_{v,u} \leq \max_{v,u} \{A_{v,u}\} \cdot \sum_{u,v} A_{u,v} = \max_{v,u} \{A_{v,u}\} \cdot \sum_v 1 = \max_{v,u} \{A_{v,u}\} \cdot n.$$

Next, we bound the  $\|p\|_2^2$  term:

$$\|p\|_2^2 = \sum_u p_u^2 \leq \max_u \{p_u\} \cdot \sum_u p_u = \max_u \{p_u\}.$$

Substituting these bounds into [Theorem 7.11](#) completes the proof.  $\square$

## 7.5 Deriving the basic SDPs

Now we will show to derive the basic SDP for QUANTUM MAX-CUT. As a warm-up, we will recall a standard method for deriving the MAX-CUT SDP

$$\text{SDP}_{\text{MC}}(G) = \max_{f: V \rightarrow S^{n-1}} \mathbf{E}_{(\mathbf{u}, \mathbf{v}) \sim E} \left[ \frac{1}{2} - \frac{1}{2} \langle f(\mathbf{u}), f(\mathbf{v}) \rangle \right]. \quad (36)$$

One way of deriving this SDP is as follows: let  $f: V \rightarrow \{-1, 1\}$  be an assignment to the vertices. Consider the  $n \times n$  matrix  $M$  defined as  $M(u, v) = f(u) \cdot f(v)$ . Then  $M$  is a real, PSD matrix such that  $M(v, v) = 1$  for all  $v \in V$ . Furthermore, we can write the value of  $f$  in terms of  $M$  as

$$\mathbf{E}_{(\mathbf{u}, \mathbf{v}) \sim E} \left[ \frac{1}{2} - \frac{1}{2} M(\mathbf{u}, \mathbf{v}) \right]. \quad (37)$$

Now, we relax our problem and consider optimizing [Equation \(37\)](#) over all real, PSD matrices  $M$  such that  $M(v, v) = 1$ . Such an  $M$  can be written as the Gram matrix of a set of real vectors of dimension  $n$ ; i.e. there is a function  $f: V \rightarrow \mathbb{R}^n$  such that  $M(u, v) = \langle f(u), f(v) \rangle$ . This yields the SDP in [Equation \(36\)](#).

**The basic SDP for QUANTUM MAX-CUT.** Now we show how to derive the basic SDP for QUANTUM MAX-CUT. To begin, let  $G = (V, E, w)$  be an  $n$ -vertex graph. Let  $|\psi\rangle \in (\mathbb{C}^2)^V$  be a quantum state. Consider the set of  $3n$  vectors  $X_u |\psi\rangle, Y_u |\psi\rangle, Z_u |\psi\rangle$ , where  $P_u$  denotes the Pauli matrix  $P$  acting on qubit  $u$ . The Gram matrix of these vectors, denoted  $M(\cdot, \cdot)$ , is the  $3n \times 3n$  matrix whose rows and columns are indexed by Pauli matrices  $P_u$  such that

$$M(P_u, Q_v) = \langle \psi | P_u Q_v | \psi \rangle.$$

Using  $M$ , we can express the energy of  $|\psi\rangle$  as follows:

$$\begin{aligned} \langle \psi | H_G | \psi \rangle &= \mathbf{E}_{(\mathbf{u}, \mathbf{v}) \sim E} \langle \psi | h_{\mathbf{u}, \mathbf{v}} | \psi \rangle = \mathbf{E}_{(\mathbf{u}, \mathbf{v}) \sim E} \frac{1}{4} \cdot \langle \psi | (I_{\mathbf{u}} \otimes I_{\mathbf{v}} - X_{\mathbf{u}} \otimes X_{\mathbf{v}} - Y_{\mathbf{u}} \otimes Y_{\mathbf{v}} - Z_{\mathbf{u}} \otimes Z_{\mathbf{v}}) | \psi \rangle \\ &= \frac{1}{4} \cdot \mathbf{E}_{(\mathbf{u}, \mathbf{v}) \sim E} [1 - M(X_{\mathbf{u}}, X_{\mathbf{v}}) - M(Y_{\mathbf{u}}, Y_{\mathbf{v}}) - M(Z_{\mathbf{u}}, Z_{\mathbf{v}})]. \end{aligned} \quad (38)$$

Let us derive some constraints on this matrix:

1. **PSD:**  $M$  is Hermitian and PSD.
2. **Unit length:** For each  $P_u$ ,  $M(P_u, P_u) = 1$ .
3. **Commuting Paulis:** For each  $P_u, Q_v$  such that  $u \neq v$ ,  $P_u$  commutes with  $Q_v$ . This implies that  $M(P_u, Q_v) = M(Q_v, P_u)$  and is therefore real because  $M$  is Hermitian.
4. **Anti-commuting Paulis:** For each  $P_u, Q_u$  such that  $P \neq Q$ ,  $P_u$  anti-commutes with  $Q_u$ . This implies that  $M(P_u, Q_u) = -M(Q_u, P_u)$  and therefore has no real part because  $M$  is Hermitian.

Now we relax our problem and consider optimizing [Equation \(38\)](#) over all matrices  $M$  that satisfy these four conditions. This is a relaxation because not all matrices  $M$  correspond to Gram matrices of vectors of the form  $P_u |\psi\rangle$ .

Prior to stating the SDP, we perform one final simplification. Given such an  $M$ , consider the matrix  $M' = \frac{1}{2}(M + M^T)$ . This satisfies all four conditions, has the same energy as  $M$ , and moreover satisfies  $M'(P_u, Q_u) = 0$  for  $P \neq Q$ . We can therefore replace [Item 4](#) with this stronger condition, which implies that  $M'$  is real. Thus,  $M'$  is a real, symmetric  $3n \times 3n$  PSD matrix, so we can write it as the Gram matrix of a set of real vectors of dimension  $3n$ . In other words, there are functions

$$f_X, f_Y, f_Z : V \rightarrow \mathbb{R}^{3n}$$

such that  $M'(P_u, Q_v) = \langle f_P(u), f_Q(v) \rangle$ . Putting everything together, we have the following SDP.

**Proposition 7.13** (QUANTUM MAX-CUT SDP). *Let  $G = (V, E, w)$  be an  $n$ -vertex graph. The value of the SDP for QUANTUM MAX-CUT can be written as*

$$\begin{aligned} \text{SDP}_{\text{QMC}}(G) &= \max \frac{1}{4} \cdot \mathbf{E}_{(\mathbf{u}, \mathbf{v}) \sim E} [1 - \langle f_X(\mathbf{u}), f_X(\mathbf{v}) \rangle - \langle f_Y(\mathbf{u}), f_Y(\mathbf{v}) \rangle - \langle f_Z(\mathbf{u}), f_Z(\mathbf{v}) \rangle], \quad (39) \\ \text{s.t. } &\langle f_P(v), f_Q(v) \rangle = 0, \quad \forall v \in V, P \neq Q \in \{X, Y, Z\}, \\ &f_X, f_Y, f_Z : V \rightarrow S^{3n-1}. \end{aligned}$$

This SDP can also be viewed as the degree-2 relaxation for QUANTUM MAX-CUT in the *non-commutative Sum of Squares (ncSoS)* hierarchy. We give a didactic treatment of this perspective in [Appendix A](#) (which is not necessary to understand the rest of this paper). We now further simplify the SDP relaxation for QUANTUM MAX-CUT and derive the expression from [Definition 2.13](#).

**Proposition 7.14** (QUANTUM MAX-CUT SDP, simplified version; [Definition 2.13](#)).

$$\text{SDP}_{\text{QMC}}(G) = \max_{f: V \rightarrow S^{3n-1}} \mathbf{E}_{(\mathbf{u}, \mathbf{v}) \sim E} \left[ \frac{1}{4} - \frac{3}{4} \langle f(\mathbf{u}), f(\mathbf{v}) \rangle \right]. \quad (40)$$

*Proof.* We first show how to convert a solution for (39) into a solution for (40) without decreasing the value. To begin, we can rewrite (39) as

$$\frac{1}{3} \cdot \mathbf{E}_{(\mathbf{u}, \mathbf{v}) \sim E} \left[ \frac{1}{4} - \frac{3}{4} \langle f_X(\mathbf{u}), f_X(\mathbf{v}) \rangle \right] + \frac{1}{3} \cdot \mathbf{E}_{(\mathbf{u}, \mathbf{v}) \sim E} \left[ \frac{1}{4} - \frac{3}{4} \langle f_Y(\mathbf{u}), f_Y(\mathbf{v}) \rangle \right] + \frac{1}{3} \cdot \mathbf{E}_{(\mathbf{u}, \mathbf{v}) \sim E} \left[ \frac{1}{4} - \frac{3}{4} \langle f_Z(\mathbf{u}), f_Z(\mathbf{v}) \rangle \right]. \quad (41)$$

Pick the term  $P \in \{X, Y, Z\}$  with the largest value, and set  $f = f_P$ . Then  $f$  has value in (40) at least the value of  $f_X, f_Y, f_Z$  in (41). The only caveat is that  $f$  maps into  $S^{3n-1}$  rather than  $S^{n-1}$ . However,  $f$  only outputs  $n$  different vectors, so these can be represented in  $n$ -dimensional space while preserving inner products.

Next, we reverse. Let  $f : V \rightarrow S^{n-1}$  be a solution to (40). We define

$$f_X(v) = e_1 \otimes f(v), \quad f_Y(v) = e_2 \otimes f(v), \quad f_Z(v) = e_3 \otimes f(v),$$

where  $e_1, e_2, e_3$  are standard basis vectors in  $\mathbb{R}^3$ . Then  $\langle f_P(v), f_Q(v) \rangle = 0$  for  $P \neq Q$  because  $\langle e_i, e_j \rangle = 0$  for  $i \neq j$ . In addition,  $\langle f_P(u), f_P(v) \rangle = \langle f(u), f(v) \rangle$ . Thus, the value of this assignment

$$\begin{aligned} (39) &= \frac{1}{4} \cdot \mathbf{E}_{(\mathbf{u}, \mathbf{v}) \sim E} [1 - \langle f_X(\mathbf{u}), f_X(\mathbf{v}) \rangle - \langle f_Y(\mathbf{u}), f_Y(\mathbf{v}) \rangle - \langle f_Z(\mathbf{u}), f_Z(\mathbf{v}) \rangle] \\ &= \frac{1}{4} \cdot \mathbf{E}_{(\mathbf{u}, \mathbf{v}) \sim E} [1 - \langle f(\mathbf{u}), f(\mathbf{v}) \rangle - \langle f(\mathbf{u}), f(\mathbf{v}) \rangle - \langle f(\mathbf{u}), f(\mathbf{v}) \rangle] \\ &= \mathbf{E}_{(\mathbf{u}, \mathbf{v}) \sim E} \left[ \frac{1}{4} - \frac{3}{4} \langle f(\mathbf{u}), f(\mathbf{v}) \rangle \right]. \end{aligned}$$

As a result, the value remains unchanged. This completes the proof.  $\square$

## 7.6 Projection rounding

We recall the performance of projection rounding, first stated in Equation (8).

**Theorem 7.15** ([BdOFV10, Lemma 2.1]). *Let  $-1 \leq \rho \leq 1$ , and let  $u$  and  $v$  be two  $n$ -dimensional unit vectors such that  $\langle u, v \rangle = \rho$ . Let  $\mathbf{Z}$  be a random  $k \times n$  matrix consisting of  $kn$  i.i.d. standard Gaussians. Then*

$$F^*(k, \rho) := \mathbf{E}_{\mathbf{Z}} \left\langle \frac{\mathbf{Z}u}{\|\mathbf{Z}u\|}, \frac{\mathbf{Z}v}{\|\mathbf{Z}v\|} \right\rangle = \frac{2}{k} \left( \frac{\Gamma((k+1)/2)}{\Gamma(k/2)} \right)^2 \langle u, v \rangle {}_2F_1(1/2, 1/2; k/2 + 1; \langle u, v \rangle^2),$$

where  ${}_2F_1(\cdot, \cdot; \cdot; \cdot)$  is the Gaussian hypergeometric function.

In the QUANTUM MAX-CUT SDP, if an edge  $(u, v)$  has value  $\frac{1}{4} - \frac{3}{4}\rho$ , then projection rounding will produce a solution whose value on this edge is  $\frac{1}{4} - \frac{1}{4}F^*(3, \rho)$  in expectation. Similarly, in the product state SDP, an edge with value  $\frac{1}{4} - \frac{1}{4}\rho$  will be rounded into a solution with value  $\frac{1}{4} - \frac{1}{4}F^*(3, \rho)$  on this edge. We can then define our approximation ratios as the worst case rounding over all values of  $\rho$ .

**Definition 7.16** (Approximation ratios). The constant  $\alpha_{\text{GP}}$  is defined as the solution to the minimization problem

$$\alpha_{\text{GP}} = \min_{-1 \leq \rho < 1/3} \frac{\frac{1}{4} - \frac{1}{4}F^*(3, \rho)}{\frac{1}{4} - \frac{3}{4}\rho},$$

and the constant  $\rho_{\text{GP}}$  is defined as the minimizing value of  $\rho$ . In addition, the constant  $\alpha_{\text{BOV}}$  is defined as the solution to the minimization problem

$$\alpha_{\text{BOV}} = \min_{-1 \leq \rho \leq 1} \frac{\frac{1}{4} - \frac{1}{4}F^*(3, \rho)}{\frac{1}{4} - \frac{1}{4}\rho},$$

and the constant  $\rho_{\text{BOV}}$  is defined as the minimizing value of  $\rho$ .

We note that the minimization for  $\alpha_{\text{GP}}$  is only over  $\rho \leq 1/3$ , because when  $\rho \geq 1/3$  the denominator is  $\leq 0$ .

**Proposition 7.17** (Formula for the optimum value). *Let  $f_{\text{opt}} : \mathbb{R}^n \rightarrow S^{k-1}$  be defined by  $f_{\text{opt}}(x) = x_{\leq k} / \|x_{\leq k}\|$ , where  $x_{\leq k} = (x_1, \dots, x_k)$ . Then*

$$\mathbf{E}_{\mathbf{x} \sim_{\rho} \mathbf{y}} \langle f_{\text{opt}}(\mathbf{x}), f_{\text{opt}}(\mathbf{y}) \rangle = F^*(k, \rho).$$

*Proof.* Let  $u, v \in \mathbb{R}^n$  be any unit vectors with  $\langle u, v \rangle = \rho$ . Let  $\mathbf{Z}$  be an  $n \times n$  matrix consisting of  $n^2$  i.i.d. standard Gaussians. Then  $\frac{1}{\sqrt{n}}\mathbf{Z}u$  and  $\frac{1}{\sqrt{n}}\mathbf{Z}v$  are distributed as  $\rho$ -correlated Gaussians. If  $\Pi_{\leq k}$  is the  $k \times n$  matrix which projects a vector down to its first  $k$  coordinates, we have

$$\begin{aligned} \mathbf{E}_{\mathbf{x} \sim_{\rho} \mathbf{y}} \langle f_{\text{opt}}(\mathbf{x}), f_{\text{opt}}(\mathbf{y}) \rangle &= \mathbf{E}_{\mathbf{Z}} \langle f_{\text{opt}}(\frac{1}{\sqrt{n}}\mathbf{Z}u), f_{\text{opt}}(\frac{1}{\sqrt{n}}\mathbf{Z}v) \rangle \\ &= \mathbf{E}_{\mathbf{Z}} \langle f_{\text{opt}}(\mathbf{Z}u), f_{\text{opt}}(\mathbf{Z}v) \rangle = \mathbf{E}_{\mathbf{Z}} \left\langle \frac{\Pi_{\leq k}\mathbf{Z}u}{\|\Pi_{\leq k}\mathbf{Z}u\|}, \frac{\Pi_{\leq k}\mathbf{Z}v}{\|\Pi_{\leq k}\mathbf{Z}v\|} \right\rangle. \end{aligned}$$

Now note that  $\Pi_{\leq k}\mathbf{Z}$  is distributed as a random  $k \times n$  matrix consisting of  $kn$  i.i.d. standard Gaussians. Applying [Theorem 7.15](#), this is exactly equal to  $F^*(k, \rho)$ .  $\square$

## 7.7 Fourier analysis on the hypercube

We will review basic concepts in the Fourier analysis of Boolean functions. See [\[O'D14\]](#) for further details.

**Definition 7.18** (Fourier transform). Let  $f : \{-1, 1\}^n \rightarrow \mathbb{R}$  be a function. Then it has a unique representation as a multilinear polynomial known as the *Fourier transform*, given by

$$f(x) = \sum_{S \subseteq [n]} \widehat{f}(S) \chi_S(x),$$

where  $\widehat{f}(S)$  is a real coefficient called the *Fourier coefficient*, and  $\chi_S(x)$  is the monomial  $\prod_{i \in S} x_i$ .

We extend this definition to functions  $f : \{-1, 1\}^n \rightarrow \mathbb{R}^k$  as follows: let  $f = (f_1, \dots, f_k)$ . Then

$$f(x) = (f_1(x), \dots, f_k(x)) = \sum_{S \subseteq [n]} (\widehat{f}_1(S), \dots, \widehat{f}_k(S)) \chi_S(x) = \sum_{S \subseteq [n]} \widehat{f}(S) \chi_S(x),$$

where  $\widehat{f}(S) = (\widehat{f}_1(S), \dots, \widehat{f}_k(S))$  is a vector-valued Fourier coefficient.

**Definition 7.19** (Variance and influences). Let  $f : \{-1, 1\}^n \rightarrow \mathbb{R}^k$ . Then

$$\mathbf{E}_{\mathbf{x} \sim \{-1, 1\}^n} \|f(\mathbf{x})\|_2^2 = \sum_{S \subseteq [n]} \|\widehat{f}(S)\|_2^2.$$

Its *variance* is the quantity

$$\mathbf{Var}[f] = \mathbf{E}_{\mathbf{x} \sim \{-1, 1\}^n} \|f(\mathbf{x}) - \mathbf{E}[f]\|_2^2 = \sum_{S \subseteq [n], S \neq \emptyset} \|\widehat{f}(S)\|_2^2.$$

Given a coordinate  $i$ , we define its *influence* as

$$\mathbf{Inf}_i[f] = \sum_{S \subseteq [n], S \ni i} \|\widehat{f}(S)\|_2^2 = \sum_{S \subseteq [n], S \ni i} \sum_j \widehat{f}_j(S)^2 = \sum_j \mathbf{Inf}_i[f_j].$$

We will also need truncated versions of these two measures:

$$\mathbf{Inf}_i^{\leq m}[f] = \sum_{|S| \leq m, S \ni i} \|\widehat{f}(S)\|_2^2, \quad \mathbf{Var}[f]^{> m} = \sum_{|S| > m} \|\widehat{f}(S)\|_2^2.$$

**Proposition 7.20** (Only few noticeable coordinates). *Let  $f : \{-1, 1\}^n \rightarrow B^k$ . Then there are at most  $m/\delta$  coordinates  $i$  such that  $\mathbf{Inf}_i^{\leq m}[f] \geq \delta$ .*

*Proof.* Let  $N$  be the set of all such coordinates. Then

$$\begin{aligned} |N| \cdot \delta &\leq \sum_{i \in N} \mathbf{Inf}_i^{\leq m}[f] = \sum_{i \in N} \sum_{|S| \leq m, S \ni i} \|\widehat{f}(S)\|_2^2 = \sum_{|S| \leq m} |S \cap N| \cdot \|\widehat{f}(S)\|_2^2 \\ &\leq \sum_{|S| \leq m} m \cdot \|\widehat{f}(S)\|_2^2 \leq m \cdot \sum_{S \subseteq [n]} \|\widehat{f}(S)\|_2^2 = m \cdot \mathbf{E}_{\mathbf{x}} \|f(\mathbf{x})\|_2^2 \leq m. \end{aligned}$$

Rearranging this gives  $|N| \leq m/\delta$ .  $\square$

**Definition 7.21** (Correlated Boolean variables [O'D14, Definition 2.40]). Given a fixed  $x \in \{-1, 1\}^n$ , we say that  $\mathbf{y} \in \{-1, 1\}^n$  is  $\rho$ -correlated to  $x$  if each coordinate  $\mathbf{y}_i$  is sampled independently according to the following distribution:

$$\mathbf{y}_i = \begin{cases} x_i & \text{with probability } \frac{1}{2} + \frac{1}{2}\rho \\ -x_i & \text{with probability } \frac{1}{2} - \frac{1}{2}\rho. \end{cases}$$

In addition, we say that  $\mathbf{x}$  and  $\mathbf{y}$  are  $\rho$ -correlated  $n$ -dimensional Boolean strings if  $\mathbf{x}$  is sampled from  $\{-1, 1\}^n$  uniformly at random and  $\mathbf{y}$  is  $\rho$ -correlated to  $\mathbf{x}$ . Note that for each  $i$ ,  $\mathbf{E}[\mathbf{x}_i \mathbf{y}_i] = \rho$ .

**Definition 7.22** (Noise stability). Let  $f : \{-1, 1\}^n \rightarrow \mathbb{R}^k$ , and let  $-1 \leq \rho \leq 1$ . Given an input  $x \in \{-1, 1\}^n$ , we write

$$\mathbf{T}_\rho f(x) = \mathbf{E}_{\substack{\mathbf{y} \text{ which is} \\ \rho\text{-correlated to } x}} [f(\mathbf{y})] = \sum_{S \subseteq [n]} \rho^{|S|} \widehat{f}(S) \chi_S(x).$$

Then the *Boolean noise stability* of  $f$  at  $\rho$  is

$$\mathbf{Stab}_\rho[f] = \mathbf{E}_{\substack{(\mathbf{x}, \mathbf{y}) \text{ } \rho\text{-correlated} \\ n\text{-dim Boolean strings}}} \langle f(\mathbf{x}), f(\mathbf{y}) \rangle = \mathbf{E}_{\mathbf{x} \sim \{-1, 1\}^n} \langle f(\mathbf{x}), \mathbf{T}_\rho f(\mathbf{x}) \rangle = \sum_{S \subseteq [n]} \rho^{|S|} \widehat{f}(S)^2.$$

This coincides with the Gaussian noise sensitivity of  $f$ . To see this, note that the Fourier expansion allows us to extend  $f$ 's domain to all of  $\mathbb{R}^n$ . Then

$$\mathbf{E}_{\mathbf{x} \sim_\rho \mathbf{y}} \langle f(\mathbf{x}), f(\mathbf{y}) \rangle = \sum_{S \subseteq [n]} \rho^{|S|} \widehat{f}(S)^2.$$

Hence, we use  $\mathbf{Stab}_\rho[f]$  as for both notions.

## 8 Integrality gaps

In this section, we prove [Theorems 2.18](#) and [2.19](#), which we restate here.

**Theorem 8.1** (Integrality gap for the QUANTUM MAX-CUT SDP; [Theorem 2.18](#) restated). *Assuming [Conjecture 1.1](#), the QUANTUM MAX-CUT semidefinite program  $\text{SDP}_{\text{QMC}}(G)$  has integrality gap  $\alpha_{\text{GP}}$ .*

**Theorem 8.2** (Integrality gap for product state SDP; [Theorem 2.19](#) restated). *Assuming [Conjecture 1.1](#), the product state semidefinite program  $\text{SDP}_{\text{PROD}}(G)$  has integrality gap  $\alpha_{\text{BOV}}$ .*

Recalling [Definition 2.16](#), our goal is to compute

$$\inf_{\text{instances } \mathcal{I} \text{ of } \mathcal{P}} \left\{ \frac{\text{OPT}(\mathcal{I})}{\text{SDP}(\mathcal{I})} \right\},$$

where  $\mathcal{P}$  is the problem of either computing the value or the product state value of a QUANTUM MAX-CUT instance. To upper bound this quantity, we construct a specific instance  $\mathcal{I}$  and give an upper-bound for  $\text{OPT}(\mathcal{I})$  and a lower-bound for  $\text{SDP}(\mathcal{I})$ . Note that for the product state case, we only optimize over product states, whereas for the general QUANTUM MAX-CUT we consider all quantum states. However, the specific instance  $\mathcal{I}$  we consider will correspond to a graph of high degree, and for such graphs Brandao and Harrow [[BH16](#)] show it suffices to consider product states.

The instance we use for both integrality gaps is the  $\rho$ -correlated Gaussian graph. This was also used as an integrality gap for the MAX-CUT problem in the work of [[OW08](#)], and we will follow their proof closely. As in their proof, we will have to deal with the technicality that the Gaussian graph is actually an infinite graph, and so our final integrality gap instance will involve discretizing the Gaussian graph to produce a finite graph.

## 8.1 The Gaussian graph as an integrality gap

To begin, we provide a lower-bound for the QUANTUM MAX-CUT SDP.

**Lemma 8.3** (QUANTUM MAX-CUT SDP Lower Bound).

$$\text{SDP}_{\text{QMC}}(\mathcal{G}_\rho^n) \geq \frac{1}{4} - \frac{3}{4}\rho - O(\sqrt{\log n/n}).$$

*Proof.* Consider the feasible solution  $f_{\text{ident}}(x) = x/\|x\|$  to the QUANTUM MAX-CUT SDP. It has value

$$\mathbf{E}_{\mathbf{x} \sim_\rho \mathbf{y}} \left[ \frac{1}{4} - \frac{3}{4} \left\langle \frac{\mathbf{x}}{\|\mathbf{x}\|}, \frac{\mathbf{y}}{\|\mathbf{y}\|} \right\rangle \right] = \frac{1}{4} - \frac{3}{4} \mathbf{E}_{\mathbf{x} \sim_\rho \mathbf{y}} \left\langle \frac{\mathbf{x}}{\|\mathbf{x}\|}, \frac{\mathbf{y}}{\|\mathbf{y}\|} \right\rangle. \quad (42)$$

Intuitively, we expect  $\langle \mathbf{x}/\|\mathbf{x}\|, \mathbf{y}/\|\mathbf{y}\| \rangle$  to roughly be equal to  $\rho$ , at least when  $n$  is large. Formally, we will use the inequality

$$\mathbf{E}_{\mathbf{x} \sim_\rho \mathbf{y}} \left\langle \frac{\mathbf{x}}{\|\mathbf{x}\|}, \frac{\mathbf{y}}{\|\mathbf{y}\|} \right\rangle \leq \rho + O(\sqrt{\log n/n}),$$

which was shown in the proof of [[OW08](#), Theorem 4.3]. This implies that

$$(42) \geq \frac{1}{4} - \frac{3}{4}\rho - O(\sqrt{\log n/n}).$$

As this lower-bounds the value of  $f_{\text{ident}}$ , it also lower-bounds the value of the SDP.  $\square$

An essentially identical proof also yields the following lemma, which gives a lower-bound for the product state SDP.

**Lemma 8.4** (Product State SDP Lower Bound).

$$\text{SDP}_{\text{PROD}}(\mathcal{G}_\rho^n) \geq \frac{1}{4} - \frac{1}{4}\rho - O(\sqrt{\log n/n}).$$

On the other hand, assuming [Conjecture 1.1](#), the optimal product state assignment is given by  $f_{\text{opt}}(x) = x_{\leq 3}/\|x_{\leq 3}\|$ , and so by [Proposition 7.17](#) the product state value can be computed exactly as

$$\text{PROD}(\mathcal{G}_\rho^n) = \frac{1}{4} - \frac{1}{4}F^*(3, \rho).$$

Not only that,  $\mathcal{G}_\rho^n$  is a weighted, regular graph of infinite degree. Thus, we should now be able to apply [Theorem 7.10](#) (or one of its nonuniform analogues) to show that its maximum energy is exactly equal to its product value. Strictly speaking, the maximum energy  $\text{QMAX-CUT}(\mathcal{G}_\rho^n)$  is not well-defined because  $\mathcal{G}_\rho^n$  is an infinite graph. However, we will define  $\text{QMAX-CUT}(\mathcal{G}_\rho^n)$  to be  $\text{PROD}(\mathcal{G}_\rho^n)$  and show in the following section that these quantities are indeed approximately equal in the discretized graph.

## 8.2 Discretizing the Gaussian graph

The following lemma shows  $\mathcal{G}_\rho^n$  can be discretized with a negligible loss in value. The proof is given in [Appendix B.1](#).

**Lemma 8.5** (Graph Discretization). *Let  $G = \mathcal{G}_\rho^n$  be the  $\rho$ -correlated Gaussian graph. Then for every  $\epsilon > 0$ , there exists a finite, weighted graph  $G'$  such that*

$$\begin{aligned} \text{SDP}_{\text{QMC}}(G') &\geq \text{SDP}_{\text{QMC}}(G) - \epsilon, & \text{QMAX-CUT}(G') &\leq \text{QMAX-CUT}(G) + \epsilon, \\ \text{SDP}_{\text{PROD}}(G') &\geq \text{SDP}_{\text{PROD}}(G) - \epsilon, & \text{PROD}(G') &\leq \text{PROD}(G) + \epsilon. \end{aligned}$$

With the instance in hand, we now prove that it yields our desired integrality gaps.

*Proof of [Theorems 8.1 and 8.2](#).* We start with the QUANTUM MAX-CUT SDP. Assume [Conjecture 1.1](#). Then combining [Lemmas 8.3 and 8.5](#) and taking the dimension  $n$  suitably large, there exists a graph  $G$  such that

$$\frac{\text{QMAX-CUT}(G)}{\text{SDP}_{\text{QMC}}(G)} \leq \frac{\frac{1}{4} - \frac{1}{4}F^*(3, \rho)}{\frac{1}{4} - \frac{3}{4}\rho} + \epsilon,$$

for each  $\epsilon > 0$ . Taking the infimum over  $\epsilon$ , the integrality gap of  $\text{SDP}_{\text{QMC}}$  is at most

$$\frac{\frac{1}{4} - \frac{1}{4}F^*(3, \rho)}{\frac{1}{4} - \frac{3}{4}\rho}.$$

This is minimized by  $\rho = \rho_{\text{GP}}$ , in which case it is equal to  $\alpha_{\text{GP}}$ . Hence, the integrality gap of the QUANTUM MAX-CUT SDP is at most  $\alpha_{\text{GP}}$ . On the other hand, the integrality gap is at least  $\alpha_{\text{GP}}$  because the GP algorithm shows there always exists a solution of value at least  $\alpha_{\text{GP}} \cdot \text{SDP}_{\text{QMC}}(G)$ . As a result, the integrality gap is exactly  $\alpha_{\text{GP}}$ , concluding the proof.

The case of the product state SDP follows by a similar argument. □

## 9 Algorithmic gap for the BOV algorithm

The main goal of this section is to prove [Theorem 2.20](#), which we restate here. We note that it does not require assuming [Conjecture 1.1](#).

**Theorem 9.1** (Algorithmic gap for product state SDP; [Theorem 2.20](#) restated). *The Briët-Oliveira-Vallentin algorithm has algorithmic gap  $\alpha_{\text{BOV}}$ .*

Recall that the algorithmic gap is the quantity

$$\inf_{\text{graphs } G} \left\{ \frac{A_{\text{BOV}}(G)}{\text{PROD}(G)} \right\},$$

where  $A_{\text{BOV}}(G)$  is the average value of the product state output by the BOV algorithm on graph  $G$ . This is at least  $\alpha_{\text{BOV}}$  by [Theorem 2.14](#), and so we need to show that it is also at most  $\alpha_{\text{BOV}}$ , which entails finding a graph  $G$  in which the BOV algorithm outputs a solution of value  $\alpha_{\text{BOV}} \cdot \text{PROD}(G)$ . Our construction is based on a classic algorithmic gap instance for the Goemans-Williamson SDP called the *noisy hypercube graph*, essentially due to Karloff [[Kar99](#)] (cf. the exposition in [[O'D08](#)]).

The BOV algorithm solves the product state SDP and rounds its solution using projection rounding. The SDP is only guaranteed to return *some* optimal (or near-optimal) solution, but we are free to choose which of these optimal solutions to provide to the BOV algorithm (see [[O'D08](#)] for more details). As in the construction of the integrality gaps in [Section 8](#), the algorithmic gap instance and the optimal SDP solution are motivated by the fact that the BOV algorithm performs worst on edges  $(u, v)$  where the SDP vectors have inner product  $\langle f_{\text{SDP}}(u), f_{\text{SDP}}(v) \rangle = \rho_{\text{BOV}}$ . The graph is an analogue of the  $\rho$ -correlated sphere graph on the Boolean hypercube, known as the noisy hypercube.

**Definition 9.2** (Noisy hypercube graph). Let  $n$  be a positive integer and  $-1 \leq \rho \leq 1$ . We define the  $\rho$ -noisy hypercube to be the graph  $\mathcal{H}_\rho^n$  with vertex set  $\{-1, 1\}^n$  in which a random edge  $(\mathbf{x}, \mathbf{y})$  is distributed as two  $\rho$ -correlated Boolean strings.

As defined, the noisy hypercube does not correspond to a legitimate QUANTUM MAX-CUT instance, as it contains self-loops. For now, we will analyze the noisy hypercube as if this is not an issue, and we will remove the self-loops at the end of the section.

The SDP solution we will consider is the “identity solution”, i.e. the function  $f_{\text{ident}} : \{-1, 1\}^n \rightarrow S^{n-1}$  defined by  $f_{\text{ident}}(x) = \frac{1}{\sqrt{n}}x$  for each  $x \in \{-1, 1\}^n$ . The following three lemmas ([9.3](#), [9.4](#), [9.5](#)) establish an upper bound on  $A_{\text{BOV}}(\mathcal{H}_\rho^n)$ .

**Lemma 9.3** (Value of the SDP solution). *The feasible SDP solution  $f_{\text{ident}}(x)$  has value  $\frac{1}{4} - \frac{1}{4}\rho$ .*

*Proof.* We compute

$$\mathbf{E}_{\substack{(\mathbf{x}, \mathbf{y}) \text{ } \rho\text{-correlated} \\ n\text{-dim Boolean strings}}} [\langle f_{\text{ident}}(\mathbf{x}), f_{\text{ident}}(\mathbf{y}) \rangle] = \frac{1}{n} \mathbf{E}_{\mathbf{x}, \mathbf{y}} \langle \mathbf{x}, \mathbf{y} \rangle = \frac{1}{n} \sum_{i=1}^n \mathbf{E}_{\mathbf{x}, \mathbf{y}} [x_i y_i] = \rho.$$

As a result, the value of  $f_{\text{ident}}$  is

$$\mathbf{E}_{\mathbf{x}, \mathbf{y}} \left[ \frac{1}{4} - \frac{1}{4} \langle f_{\text{ident}}(\mathbf{x}), f_{\text{ident}}(\mathbf{y}) \rangle \right] = \frac{1}{4} - \frac{1}{4}\rho. \quad \square$$

Our next lemma shows that  $f_{\text{ident}}$  achieves the optimal SDP value, and so it is fair for the BOV algorithm to receive it as a solution to the product state SDP.

**Lemma 9.4** (Value of the SDP). *The value of the SDP is  $\text{SDP}_{\text{PROD}}(\mathcal{H}_\rho^n) = \frac{1}{4} - \frac{1}{4}\rho$ . As a result,  $f_{\text{ident}}$  is an optimal SDP solution.*

*Proof.* The function  $f_{\text{ident}}$  is a feasible solution with value  $\frac{1}{4} - \frac{1}{4}\rho$ , and so it suffices to show that  $\text{SDP}_{\text{PROD}}(\mathcal{H}_\rho^n) \leq \frac{1}{4} - \frac{1}{4}\rho$ . This entails showing that

$$\mathbf{E}_{\substack{(\mathbf{x}, \mathbf{y}) \text{ } \rho\text{-correlated} \\ n\text{-dim Boolean strings}}} \left[ \frac{1}{4} - \frac{1}{4} \langle f(\mathbf{x}), f(\mathbf{y}) \rangle \right] \leq \frac{1}{4} - \frac{1}{4}\rho,$$

for all functions  $f : \{-1, 1\}^n \rightarrow S^{N-1}$ , where  $N = 2^n$  is the number of vertices in  $\mathcal{H}_\rho^n$ . Equivalently, we will show  $\mathbf{E}_{\mathbf{x}, \mathbf{y}} \langle f(\mathbf{x}), f(\mathbf{y}) \rangle \geq \rho$  for all such  $f$ .

Write  $f = (f_1, \dots, f_N)$  where  $f_i : \{-1, 1\}^n \rightarrow \mathbb{R}$  for each  $i$ . Then,

$$\mathbf{E}_{\substack{(\mathbf{x}, \mathbf{y}) \text{ } \rho\text{-correlated} \\ n\text{-dim Boolean strings}}} \langle f(\mathbf{x}), f(\mathbf{y}) \rangle = \sum_{i=1}^N \mathbf{E}_{\mathbf{x}, \mathbf{y}} [f_i(\mathbf{x})f_i(\mathbf{y})] \geq \sum_{i=1}^N \rho \cdot \mathbf{E}_{\mathbf{x}} [f_i(\mathbf{x})^2] = \rho \cdot \mathbf{E}_{\mathbf{x}} \left[ \sum_{i=1}^N f_i(\mathbf{x})^2 \right] = \rho.$$

The inequality here is due to [Proposition B.2](#) and we defer it to the appendix. This completes the proof.  $\square$

Given  $f_{\text{opt}} = f_{\text{ident}}$ , the BOV algorithm performs projection rounding and outputs the solution. The following lemma shows the value of the output.

**Lemma 9.5** (Value of the projection rounding). *Given the SDP solution  $f_{\text{ident}}$ , projection rounding will produce a random product state whose average value is at most*

$$\frac{1}{4} - \frac{1}{4}F^*(3, \rho) + O(\sqrt{\log(n)/n}).$$

*Proof.* By the Chernoff bound,

$$\mathbf{Pr}_{\substack{(\mathbf{x}, \mathbf{y}) \text{ } \rho\text{-correlated} \\ n\text{-dim Boolean strings}} [\langle \mathbf{x}, \mathbf{y} \rangle = \rho \cdot n \pm O(\sqrt{n \log n})] \leq O(1/\sqrt{n}). \quad (43)$$

Let  $(x, y)$  be an edge in  $\mathcal{H}_\rho^n$  such that  $\langle x, y \rangle = \rho \cdot n \pm O(\sqrt{n \log n})$ . Then

$$\rho_{x,y} := \langle f_{\text{ident}}(x), f_{\text{ident}}(y) \rangle = \rho \pm O(\sqrt{\log n/n}).$$

On this edge, the random product state produced by projection rounding has average value  $\frac{1}{4} - \frac{1}{4}F^*(3, \rho_{x,y})$  due to [Theorem 7.15](#). Since  $\rho_{x,y}$  is within  $O(\sqrt{\log n/n})$  of  $\rho$ , and  $F^*(3, \cdot)$  is Lipschitz by [Lemma B.6](#), we can bound the average value of this edge by

$$\frac{1}{4} - \frac{1}{4}F^*(3, \rho) + O(\sqrt{\log n/n}).$$

As for the remaining edges, we can trivially bound the average value of each by 1, and this contributes an extra  $O(1/\sqrt{n})$  to the total value of the product state due to [Equation \(43\)](#).  $\square$

Now, we compute the optimal product state value for  $\mathcal{H}_\rho^n$ . To do so, we consider a family of product state solutions  $f : \{-1, 1\}^n \rightarrow S^2$  we call *embedded dictators*, in which there exists an  $i \in [n]$  such that  $f(x) = (x_i, 0, 0)$  for all  $x$ .

**Lemma 9.6** (Value of embedded dictators). *Embedded dictators  $f(x) = (x_i, 0, 0)$  achieve value  $\frac{1}{4} - \frac{1}{4}\rho$ . Hence, the product state value of  $\mathcal{H}_\rho^n$  is  $\text{PROD}(\mathcal{H}_\rho^n) = \frac{1}{4} - \frac{1}{4}\rho$ .*

*Proof.* Consider the product state corresponding to the function  $f_{\text{dict}} : \{-1, 1\}^n \rightarrow S^2$  given by  $f_{\text{dict}}(x) = (x_i, 0, 0)$ . It has value

$$\frac{1}{4} - \frac{1}{4} \mathbf{E}_{\substack{(\mathbf{x}, \mathbf{y}) \text{ } \rho\text{-correlated} \\ n\text{-dim Boolean strings}}} [\langle f_{\text{dict}}(\mathbf{x}), f_{\text{dict}}(\mathbf{y}) \rangle] = \frac{1}{4} - \frac{1}{4} \mathbf{E}_{\mathbf{x} \sim_\rho \mathbf{y}} [x_i y_i] = \frac{1}{4} - \frac{1}{4}\rho.$$

This shows the product state value is at least  $\frac{1}{4} - \frac{1}{4}\rho$ . It is also at most  $\frac{1}{4} - \frac{1}{4}\rho$  because this is the value of the product state SDP, which is an upper bound on the product state value.  $\square$

*Proof of Theorem 9.1.* To begin, we will remove the self-loops from  $\mathcal{H}_\rho^n$ , which have total weight

$$w_{\text{loops}} = \Pr_{\substack{(\mathbf{x}, \mathbf{y}) \text{ } \rho\text{-correlated} \\ n\text{-dim Boolean strings}}} [\mathbf{x} = \mathbf{y}] = \left(\frac{1}{2} + \frac{1}{2}\rho\right)^n.$$

Let  $\mathcal{H}'$  be the graph with vertex set  $\{-1, 1\}^n$  in which a random edge  $(\mathbf{x}, \mathbf{y})$  is distributed as two  $\rho$ -correlated Boolean strings, conditioned on  $\mathbf{x} \neq \mathbf{y}$ . Then for each edge  $(x, y)$ , if  $w(x, y)$  is its weight in  $\mathcal{H}_\rho^n$  and  $w'(x, y)$  is its weight in  $\mathcal{H}'$ , we have that  $w'(x, x) = 0$ , and

$$w'(x, y) = \frac{1}{1 - w_{\text{loops}}} \cdot w(x, y)$$

for  $x \neq y$ . Consider an SDP solution  $f : \{-1, 1\}^n \rightarrow S^{N-1}$ . It has value 0 on each self-loop in  $\mathcal{H}_\rho^n$ . As a result, if it has value  $\nu$  in  $\mathcal{H}_\rho^n$ , then it has value  $\nu/(1 - w_{\text{loops}})$  in  $\mathcal{H}'$ . This argument applies to the value of product states as well.

In summary, all values are “scaled up” by a factor of  $1/(1 - w_{\text{loops}})$  in  $\mathcal{H}'$ . This implies that  $f_{\text{ident}}$  is still an optimal SDP solution, and by Lemma 9.5 and Lemma 9.6, the ratio of the average value of the resulting solution to  $\text{PROD}(\mathcal{H}')$  is

$$\frac{\frac{1}{4} - \frac{1}{4}F^*(3, \rho) + O(\sqrt{\log(n)/n})}{\frac{1}{4} - \frac{1}{4}\rho}.$$

Taking an infimum over  $n$ , we can upper bound the algorithmic gap by

$$\frac{\frac{1}{4} - \frac{1}{4}F^*(3, \rho)}{\frac{1}{4} - \frac{1}{4}\rho}.$$

This is minimized at  $\rho = \rho_{\text{BOV}}$ , in which case it is  $\alpha_{\text{BOV}}$ , matching the approximation ratio of the BOV algorithm.  $\square$

## 10 A dictator test for the product state value

Now we show that the noisy hypercube serves as a dictatorship test for functions of the form  $f : \{-1, 1\}^n \rightarrow B^k$ , assuming Conjecture 1.1. Informally, this means that if  $f$  is an embedded dictator, it should have high value, and if it is “far” from a dictator (in the sense that it has no “notable” input coordinates) then it should have low value. This will be an important ingredient in our Unique-Games hardness proof in Section 11 below.

We have already shown that embedded dictators achieve value  $\frac{1}{4} - \frac{1}{4}\rho$  on the noisy hypercube in Lemma 9.6. Now we will upper-bound the value that functions “far” from dictators achieve. We will show that their value, up to small error, is at most the optimum product state value on the Gaussian graph  $\mathcal{G}_\rho^n$ , which we have shown to be  $\frac{1}{4} - \frac{1}{4}F^*(k, \rho)$ . Throughout this section, we will make heavy use of the various Fourier analytic quantities defined in Section 7.7.

**Theorem 10.1** (Dictatorship test soundness). *Assume Conjecture 1.1. Let  $-1 < \rho \leq 0$ . Then for any  $\epsilon > 0$ , there exists a small enough  $\delta = \delta(\epsilon, \rho) > 0$  and large enough  $m = m(\epsilon, \rho) \geq 0$  such that the following is true. Let  $f : \{-1, 1\}^n \rightarrow B^k$  be any function satisfying*

$$\text{Inf}_i^{\leq m}[f] = \sum_{j=1}^k \text{Inf}_i^{\leq m}[f_j] \leq \delta, \quad \text{for all } i = 1, \dots, n.$$

Then

$$\mathbf{E}_{\substack{(\mathbf{x}, \mathbf{y}) \text{ } \rho\text{-correlated} \\ n\text{-dim Boolean strings}}} [f(\mathbf{x})f(\mathbf{y})] = \mathbf{Stab}_\rho[f] \geq F^*(k, \rho) - \epsilon.$$

In other words, the value of  $f$  on the noisy hypercube  $\mathcal{H}_\rho^n$  is at most  $\frac{1}{4} - \frac{1}{4}F^*(k, \rho) + \epsilon$ .

The  $k = 1$  case is the negative  $\rho$  case of the Majority is Stablest theorem of [MOO10], which serves as the soundness case for the MAX-CUT dictatorship test; our theorem generalizes the negative  $\rho$  case of Majority is Stablest to larger values of  $k$ . The proof follows the same outline as the proof of Majority is Stablest appearing in [O'D14, Chapter 11.7]: we apply an ‘‘invariance principle’’ to exchange  $f$ 's Boolean inputs with Gaussians of the same mean and variance. Then we use our [Conjecture 1.1](#) on the noise stability of functions in Gaussian space to upper-bound the value of  $f$ . The invariance principle we will use is the following one due to Isaksson and Mossel [IM12], which applies to vector-valued functions.

**Theorem 10.2** (Vector-valued invariance principle [IM12]). *Fix  $\delta, \gamma \in (0, 1)$  and set  $m = \frac{1}{18} \log \frac{1}{\delta}$ . Let  $f = (f_1, \dots, f_k)$  be a  $k$ -dimensional multilinear polynomial such that  $\mathbf{Var}[f_j] \leq 1$ ,  $\mathbf{Var}[f_j^{>m}] \leq (1 - \gamma)^{2m}$ , and  $\mathbf{Inf}_i^{\leq m}[f_j] \leq \delta$  for each  $j \in [k]$  and  $i \in [n]$ . Let  $\mathbf{x}$  be a uniformly random string over  $\{-1, 1\}^n$  and  $\mathbf{y}$  be an  $n$ -dimensional standard Gaussian random variable. Furthermore, let  $\Psi : \mathbb{R}^k \rightarrow \mathbb{R}$  be Lipschitz continuous with Lipschitz constant  $A$ . Then,*

$$|\mathbf{E}[\Psi(f(\mathbf{x}))] - \mathbf{E}[\Psi(f(\mathbf{y}))]| \leq C_k A \delta^{\gamma/(18 \ln 2)}$$

where  $C_k$  is a parameter depending only on  $k$ .

As a first step in proving [Theorem 10.1](#), we claim that we can assume  $f$  is odd (i.e.  $f(x) = -f(-x)$ ) without loss of generality.

**Lemma 10.3.** *Fix  $\rho \in (-1, 0]$  and  $f : \{-1, 1\}^n \rightarrow B^k$ . Define the function  $g(x) = \frac{1}{2}(f(x) - f(-x))$ . Then,  $g$  is odd, has range  $B^k$ , and satisfies*

$$\mathbf{Stab}_\rho[f] \geq \mathbf{Stab}_\rho[g]$$

and

$$\forall i \in [n], \mathbf{Inf}_i^{\leq m}[f] \geq \mathbf{Inf}_i^{\leq m}[g].$$

*Proof.* Oddness follows from  $g(-x) = \frac{1}{2}(f(-x) - f(x)) = -g(x)$ . In addition,  $g(x)$  maps into  $B^k$  because it is the average of two points in  $B^k$ . The Fourier transform of  $f$  is  $\sum_{S \subseteq [n]} \hat{f}(S) \chi_S(x)$ . Noting that  $\chi_S(x) = \chi_S(-x)$  for  $|S|$  even and  $-\chi_S(-x)$  for  $|S|$  odd, the Fourier transform of  $g$  is

$$g(x) = \sum_{|S| \text{ odd}} \hat{f}(S) \chi_S(x).$$

Thus, for any  $i \in [n]$ ,

$$\mathbf{Inf}_i^{\leq m}[g] = \sum_{\substack{|S| \leq m, S \ni i, \\ |S| \text{ odd}}} \|\hat{f}(S)\|_2^2 \leq \sum_{|S| \leq m, S \ni i} \|\hat{f}(S)\|_2^2 = \mathbf{Inf}_i^{\leq m}[f].$$

Furthermore, using that  $\rho^{|S|} \geq 0$  for  $|S|$  even,

$$\mathbf{Stab}_\rho[g] = \sum_{|S| \text{ odd}} \rho^{|S|} \|\hat{f}(S)\|_2^2 \leq \sum_{S \subseteq [n]} \rho^{|S|} \|\hat{f}(S)\|_2^2 = \mathbf{Stab}_\rho[f]. \quad \square$$

Therefore, to lower bound  $\mathbf{Stab}_\rho[f]$ , it suffices to lower bound the noise stability of the odd function  $g$ . Then, the proof of [Theorem 10.1](#) is essentially by reduction to case of *positive*  $\rho$ , i.e.  $\rho \in [0, 1]$ , but only for odd functions. This requires an analogue of [Conjecture 1.1](#) for the case of positive  $\rho$  and odd  $f$ , which we derive as follows.

**Corollary 10.4** (Vector-valued Borell’s inequality; positive  $\rho$  and odd  $f$  case). *Assume [Conjecture 1.1](#). Then it holds in the reverse direction when  $\rho \in [0, 1]$ , with the additional assumption that  $f$  is odd. In particular,*

$$\mathbf{Stab}_\rho[f] \leq \mathbf{Stab}_\rho[f_{\text{opt}}],$$

where  $f_{\text{opt}}(x) = x_{\leq k} / \|x_{\leq k}\|$  and  $x_{\leq k} = (x_1, \dots, x_k)$ .

*Proof.* Because  $f$  is odd,

$$\mathbf{Stab}_\rho[f] = \mathbf{E}_{\mathbf{x} \sim_\rho \mathbf{y}} [\langle f(\mathbf{x}), f(\mathbf{y}) \rangle] = - \mathbf{E}_{\mathbf{x} \sim_\rho \mathbf{y}} [\langle f(\mathbf{x}), f(-\mathbf{y}) \rangle] = - \mathbf{E}_{\mathbf{x} \sim_{-\rho} \mathbf{y}} [\langle f(\mathbf{x}), f(\mathbf{y}) \rangle] = -\mathbf{Stab}_{-\rho}[f], \quad (44)$$

where we have used the fact that if  $\mathbf{x}$  and  $\mathbf{y}$  are  $\rho$ -correlated, then  $\mathbf{x}$  and  $-\mathbf{y}$  are  $(-\rho)$ -correlated. Recalling that  $\rho \geq 0$ , we apply [Conjecture 1.1](#) and obtain  $\mathbf{Stab}_{-\rho}[f] \geq \mathbf{Stab}_{-\rho}[f_{\text{opt}}]$ . Observing that  $f_{\text{opt}}$  is an odd function, we further have

$$\mathbf{Stab}_{-\rho}[f] \geq \mathbf{Stab}_{-\rho}[f_{\text{opt}}] = -\mathbf{Stab}_\rho[f_{\text{opt}}].$$

Thus, negating the above inequality and combining with [Equation \(44\)](#) yields

$$\mathbf{Stab}_\rho[f] = -\mathbf{Stab}_{-\rho}[f] \leq \mathbf{Stab}_\rho[f_{\text{opt}}].$$

This concludes the proof. □

Now we prove the main theorem of this section.

*Proof of [Theorem 10.1](#).* To begin, we choose parameters

$$\begin{aligned} \gamma &= \frac{1}{6}(1 + \rho)\epsilon, && \text{(dictated by [Equation \(45\)](#))} \\ \delta &= \left( \frac{\epsilon}{12kC_k} \right)^{(18 \ln 2)/\gamma}, && \text{(so that the error in [Theorem 10.2](#) is at most } \epsilon/(6k)\text{)} \\ m &= \frac{1}{18} \log \frac{1}{\delta}. && \text{(dictated by [Theorem 10.2](#))} \end{aligned}$$

Using [Lemma 10.3](#), assume without loss of generality that  $f$  is odd. Throughout, we’ll use  $x$  to denote a string in  $\{-1, 1\}^n$  and  $y$  to denote a vector in  $\mathbb{R}^n$ . To prove the claim, we’ll perform a series of modifications to the initial function  $f$  so that we can apply [Theorem 10.2](#). At each step, we’ll show that each modified function has noise stability close to  $\mathbf{Stab}_\rho[f]$ . In particular, we’ll consider the following functions:

1.  $g(x) = \mathbf{T}_{1-\gamma}f(x)$ ,
2.  $g(y)$ , where  $y \in \mathbb{R}^n$ ,
3.  $\mathcal{R} \circ g(y)$ , which is  $g(y)$  rounded to  $B^k$ .

**Step 1.** Since the statement of the vector-valued invariance principle ([Theorem 10.2](#)) requires a function with low high-degree variance, we consider  $g = \mathbf{T}_{1-\gamma}f$ . Then for each  $j \in [k]$ ,

$$\mathbf{Var}[g_j^{>m}] = \sum_{|S|>m} (1-\gamma)^{2|S|} \widehat{f}_j(S)^2 \leq (1-\gamma)^{2m} \sum_{S \subseteq [n]} \widehat{f}_j(S)^2 \leq (1-\gamma)^{2m}.$$

Also,  $\mathbf{Inf}_i^{\leq m}[g_j] \leq \mathbf{Inf}_i^{\leq m}[f_j]$  for all  $i \in [n]$ . Furthermore, note that since  $f$  is odd,  $g$  is also odd. Next, we bound the error in the quantity  $\mathbf{Stab}_\rho[f]$  when we consider  $g$  in place  $f$ . Since  $g = \mathbf{T}_{1-\gamma}f$ , we see that  $\mathbf{Stab}_\rho[g] = \mathbf{Stab}_{\rho(1-\gamma)^2}[f]$ , and so it suffices to bound  $|\mathbf{Stab}_\rho[f] - \mathbf{Stab}_{\rho(1-\gamma)^2}[f]|$ . To do so, we use [Proposition B.1](#). However, note this proposition only applies for  $\rho > 0$ . But since  $f$  is odd, we have,

$$|\mathbf{Stab}_\rho[f] - \mathbf{Stab}_{\rho(1-\gamma)^2}[f]| = |-\mathbf{Stab}_{-\rho}[f] + \mathbf{Stab}_{-\rho(1-\gamma)^2}[f]| = |\mathbf{Stab}_{-\rho}[f] - \mathbf{Stab}_{-\rho(1-\gamma)^2}[f]|,$$

which allows us to apply [Proposition B.1](#) with  $\rho' = -\rho$ . This yields

$$|\mathbf{Stab}_\rho[f] - \mathbf{Stab}_{\rho(1-\gamma)^2}[f]| \leq \frac{2\gamma}{1+\rho} \mathbf{Var}[f] \leq \frac{2\gamma}{1+\rho} \mathbf{E}[\|f\|_2^2] \leq \frac{2\gamma}{1+\rho}, \quad (45)$$

where we used  $\mathbf{Var}[f] \leq \mathbf{E}[\|f\|_2^2]$  for all functions and the fact that  $f$ 's range is  $B^k$ . For our choice of  $\gamma$ , this is equal to  $\epsilon/3$ .

**Step 2.** Next, we bound the error accrued when we apply  $g$  on Gaussian inputs. Consider the function

$$\Psi(v) = \begin{cases} \|v\|_2^2 & \text{if } \|v\|_2 \leq 1, \\ 1 & \text{otherwise.} \end{cases}$$

By [Lemma B.3](#),  $\Psi$  is 2-Lipschitz. We'll rewrite  $\mathbf{Stab}_\rho[g]$  using  $\Psi$  in order to apply [Theorem 10.2](#).

$$\begin{aligned} \mathbf{Stab}_\rho[g(\mathbf{x})] &= \mathbf{E}_{\mathbf{x}}[\langle g(\mathbf{x}), \mathbf{T}_\rho g(\mathbf{x}) \rangle] \\ &= \mathbf{E}_{\mathbf{x}}[\langle \mathbf{T}_{\sqrt{-\rho}} g(\mathbf{x}), \mathbf{T}_{-\sqrt{-\rho}} g(\mathbf{x}) \rangle] \\ &= -\mathbf{E}_{\mathbf{x}}[\langle \mathbf{T}_{\sqrt{-\rho}} g(\mathbf{x}), \mathbf{T}_{\sqrt{-\rho}} g(\mathbf{x}) \rangle] \quad (\text{Since } g \text{ is odd}) \\ &= -\mathbf{E}_{\mathbf{x}}[\langle \|\mathbf{T}_{\sqrt{-\rho}} g(\mathbf{x})\|_2^2 \rangle] \\ &= -\mathbf{E}_{\mathbf{x}}[\Psi(\mathbf{T}_{\sqrt{-\rho}} g(\mathbf{x}))], \end{aligned}$$

where the last step used that  $\mathbf{T}_{\sqrt{-\rho}} g(\mathbf{x}) \in B^k$  and hence is unchanged by  $\Psi(\cdot)$ . Applying [Theorem 10.2](#), we get

$$\left| \mathbf{E}_{\mathbf{x}}[\Psi(\mathbf{T}_{\sqrt{-\rho}}[g(\mathbf{x}))]] - \mathbf{E}_{\mathbf{y}}[\Psi(\mathbf{T}_{\sqrt{-\rho}}[g(\mathbf{y}))]] \right| \leq \epsilon/(6k) \leq \epsilon/3.$$

Here, we chose the “ $f$ ” function in [Theorem 10.2](#) to be  $T_{\sqrt{-\rho}}g$ , which satisfies the low variance and influence properties because  $g$  does.

**Step 3.** The term  $-\mathbf{E}_{\mathbf{y}}[\Psi(\mathbf{T}_{\sqrt{-\rho}}[g(\mathbf{y}))]]$  is *almost* ready for application of [Conjecture 1.1](#) through [Corollary 10.4](#). However, although  $g(\mathbf{x})$  is bounded in  $B^k$ , the same might not hold for  $g(\mathbf{y})$ , where  $\mathbf{y} \in \mathbb{R}^n$ . To fix this, we consider  $\mathcal{R} \circ g$ , where

$$\mathcal{R}(v) = \begin{cases} v & \text{if } \|v\|_2 \leq 1, \\ \frac{v}{\|v\|} & \text{otherwise.} \end{cases}$$

In other words,  $\mathcal{R}(v)$  rounds a vector  $v$  to the unit ball  $B^m$ . Then

$$\begin{aligned}
& \left| \mathbf{E}_{\mathbf{y}} \left[ \Psi(\mathbf{T}_{\sqrt{-\rho}}[g(\mathbf{y})]) \right] - \mathbf{E}_{\mathbf{y}} \left[ \Psi(\mathbf{T}_{\sqrt{-\rho}}[\mathcal{R} \circ g(\mathbf{y})]) \right] \right| \\
& \leq \mathbf{E}_{\mathbf{y}} \left| \Psi(\mathbf{T}_{\sqrt{-\rho}}[g(\mathbf{y})]) - \Psi(\mathbf{T}_{\sqrt{-\rho}}[\mathcal{R} \circ g(\mathbf{y})]) \right| \\
& \leq 2 \mathbf{E}_{\mathbf{y}} \left\| \mathbf{T}_{\sqrt{-\rho}}[g(\mathbf{y})] - \mathbf{T}_{\sqrt{-\rho}}[\mathcal{R} \circ g(\mathbf{y})] \right\|_2 \quad (\Psi \text{ is 2-Lipschitz}) \\
& \leq 2 \sum_{i=1}^k \mathbf{E}_{\mathbf{y}} \left| \mathbf{T}_{\sqrt{-\rho}}[g_i(\mathbf{y})] - \mathbf{T}_{\sqrt{-\rho}}[(\mathcal{R} \circ g)_i(\mathbf{y})] \right| \quad (\|\cdot\|_2 \leq \|\cdot\|_1) \\
& \leq 2 \sum_{i=1}^k \mathbf{E}_{\mathbf{y}} |g_i(\mathbf{y}) - (\mathcal{R} \circ g)_i(\mathbf{y})| \quad (\mathbf{T}_{\sqrt{-\rho}} \text{ is a contraction}) \\
& = 2 \sum_{i=1}^k \mathbf{E}_{\mathbf{y}} |\Phi_i(g(\mathbf{y}))|, \tag{46}
\end{aligned}$$

where we define the map  $\Phi_i : \mathbb{R}^k \rightarrow \mathbb{R}$  to be  $\Phi_i(v) = v_i - \mathcal{R}(v)_i$ , which is 2-Lipschitz by [Corollary B.5](#). To evaluate this, we first recall that on *Boolean* inputs  $x$ ,  $g(x) \in B^k$ , and so we have that  $\Phi_i(g(x)) = 0$ . Therefore

$$\left| \mathbf{E}_x[\Phi_i(g(x))] - \mathbf{E}_{\mathbf{y}}[\Phi_i(g(\mathbf{y}))] \right| = \left| \mathbf{E}_{\mathbf{y}}[\Phi_i(g(\mathbf{y}))] \right|,$$

and applying [Theorem 10.2](#) one last time, we can upper bound this by  $\epsilon/6k$ . The sum in [Equation \(46\)](#) is in turn upper bounded by  $\epsilon/3$ . Finally, applying [Corollary 10.4](#) to

$$-\mathbf{E}_{\mathbf{y}} \left[ \Psi(\mathbf{T}_{\sqrt{-\rho}}[\mathcal{R} \circ g(\mathbf{y})]) \right] = -\mathbf{Stab}_{-\rho}[\mathcal{R} \circ g]$$

yields a lower bound of  $-\mathbf{Stab}_{-\rho}[f_{\text{opt}}] = \mathbf{Stab}_{\rho}[f_{\text{opt}}]$ , for which [Proposition 7.17](#) and [Theorem 7.15](#) give an explicit formula. Through the three transformations, we accrue an error of at most  $\epsilon$ , which proves the claim.  $\square$

## 11 Unique Games hardness of QUANTUM MAX-CUT

Now we prove hardness of QUANTUM MAX-CUT. Our starting point is the Unique Games problem.

**Definition 11.1** (Unique Games). The Unique Games problem is defined as follows. An instance is a tuple  $\mathcal{I}(U, V, E, [M], \{\pi_{u \rightarrow v}\}_{(u,v) \in E})$ , corresponding to a bipartite graph with left side vertices  $U$ , right side vertices  $V$ , and edges  $E$ , in addition to a bijection  $\pi_{u \rightarrow v} : [M] \rightarrow [M]$  for each  $(u, v) \in E$ . We will also write  $\pi_{v \rightarrow u}$  for  $\pi_{u \rightarrow v}^{-1}$ . A labeling of the vertices is a function  $L : U \cup V \rightarrow [M]$ , which satisfies the edge  $(u, v) \in E$  if  $\pi_{u \rightarrow v}(L(u)) = L(v)$ . The value of  $L$  is the fraction of edges it satisfies, and the value of the instance  $\mathcal{I}$  is the maximum value of any labeling.

**Conjecture 11.2** (Unique Games Conjecture [[Kho02](#)]). *For any  $\gamma > 0$ , there exists a constant  $M = M(\gamma)$  such that it is NP-hard to distinguish whether an instance of the Unique Games problem with label set size  $M$  has value at least  $1 - \gamma$  or at most  $\gamma$ . Furthermore, we may assume the constraint graph  $\mathcal{C} = (U \cup V, E)$  is biregular.*

The fact that the constraint graph may be taken to be biregular is a consequence of the result of Khot and Regev [[KR08](#)], as pointed out by Bansal and Khot [[BK10](#)]. Our hardness result is stated as follows.

**Theorem 11.3** (UG-Hardness of Approximating QUANTUM MAX-CUT). *Assume [Conjecture 1.1](#). For any  $\rho \in (-1, 0)$  and  $\epsilon > 0$ , given an instance of QUANTUM MAX-CUT, the Unique Games Conjecture implies that the following two tasks are NP-hard:*

1. *distinguishing if the product state value is greater than  $\frac{1}{4} - \frac{1}{4}\rho - \epsilon$  or less than  $\frac{1}{4} - \frac{1}{4}F^*(3, \rho) + \epsilon$ ,*
2. *distinguishing if the maximum energy is greater than  $\frac{1}{4} - \frac{1}{4}\rho - \epsilon$  or less than  $\frac{1}{4} - \frac{1}{4}F^*(3, \rho) + \epsilon$ .*

*Choosing  $\rho = \rho_{\text{BOV}}$ , these imply that approximating the product state value and maximum energy to a factor  $\alpha_{\text{BOV}} + \epsilon$  is NP-hard, assuming the Unique Games Conjecture.*

The proof mostly follows the standard outline for UG-hardness proofs introduced in [\[KKMO07\]](#). In particular, the graph produced by the reduction is exactly the same as the one produced in their MAX-CUT reduction, with the one exception that we will eventually eliminate all self-loops so that it is a well-defined QUANTUM MAX-CUT instance. The chief new difficulty is that in order to estimate the maximum energy of the graph produced by the reduction, we use [Corollary 7.12](#) to relate it to the product state value; however, the error term this theorem produces is a somewhat odd analogue of degree for weighted graphs, and bounding it is slightly tedious.

*Proof of [Theorem 11.3](#).* The proof is by reduction from the Unique Games problem. To begin, we choose parameters

$$\begin{aligned} \delta &= \delta(\epsilon/2, \rho), & (\delta(\cdot, \cdot) \text{ from } \a href="#">Theorem 10.1) \\ m &= m(\epsilon/2, \rho), & (m(\cdot, \cdot) \text{ from } \a href="#">Theorem 10.1) \\ M &= \max\{M(\gamma), \frac{8 \log(\epsilon/200)}{\log(1/2 - \rho/2)}\}. \\ & \quad (M(\cdot) \text{ from } \a href="#">Conjecture 11.2; \text{ dictated by } \a href="#">Equations (49) \text{ and } (50)) \\ \gamma &= \frac{\epsilon \delta^2}{16M}, & (\text{dictated by } \a href="#">Equations (47) \text{ and } (48)) \end{aligned}$$

Let  $\mathcal{I}(U, V, E, [M], \{\pi_{u \rightarrow v}\}_{(u,v) \in E})$  be a biregular instance of the Unique Games problem. The reduction produces a QUANTUM MAX-CUT instance with graph  $G$  whose vertex set is  $V \times \{-1, 1\}^M$ . A random edge in  $G$  is sampled as follows: pick  $\mathbf{u} \in U$  uniformly at random, and sample two uniformly random neighbors  $\mathbf{v}, \mathbf{w} \sim N(\mathbf{u})$  independently, where  $N(\mathbf{u})$  is the set of  $\mathbf{u}$ 's neighbors. Let  $\mathbf{x}$  and  $\mathbf{y}$  be  $\rho$ -correlated  $M$ -dimensional Boolean strings. Output the edge between  $(\mathbf{v}, \mathbf{x} \circ \pi_{\mathbf{v} \rightarrow \mathbf{u}})$  and  $(\mathbf{w}, \mathbf{y} \circ \pi_{\mathbf{w} \rightarrow \mathbf{u}})$ . Given  $w \in \{-1, 1\}^M$  and  $\sigma : [M] \rightarrow [M]$ , we write  $w \circ \sigma \in \{-1, 1\}^M$  for the string in which  $(w \circ \sigma)_i = w_{\sigma(i)}$ .

A product state assignment to  $G$  corresponds to a function  $f_v : \{-1, 1\}^M \rightarrow S^2$  for each  $v \in V$ . It has value

$$\mathbf{E}_{\mathbf{u} \sim U} \mathbf{E}_{\mathbf{v}, \mathbf{w} \sim N(\mathbf{u})} \mathbf{E}_{\substack{(\mathbf{x}, \mathbf{y}) \text{ } \rho\text{-correlated} \\ n\text{-dim Boolean strings}}} \left[ \frac{1}{4} - \frac{1}{4} \langle f_{\mathbf{v}}(\mathbf{x} \circ \pi_{\mathbf{v} \rightarrow \mathbf{u}}), f_{\mathbf{w}}(\mathbf{y} \circ \pi_{\mathbf{w} \rightarrow \mathbf{u}}) \rangle \right].$$

*Completeness.* Assume  $\mathcal{I}$  has a labeling  $L : U \cup V \rightarrow [M]$  satisfying more than  $(1 - \gamma)$ -fraction of the edges. For each  $v \in V$ , let  $f_v(x) = (x_{L(v)}, 0, \dots, 0)$ . To analyze the performance of  $f$ , let us first fix a vertex  $u \in U$  and two neighbors  $v, w \in N(u)$ , and condition on the case that  $L$  satisfies both edges  $(u, v)$  and  $(u, w)$ . This means that  $\pi_{v \rightarrow u}(L(v)) = L(u) = \pi_{w \rightarrow u}(L(w))$ . Thus, for each  $x \in \{-1, 1\}^M$ ,

$$f_v(x \circ \pi_{v \rightarrow u}) = ((x \circ \pi_{v \rightarrow u})_{L(v)}, 0, 0) = (x_{\pi_{v \rightarrow u}(L(v))}, 0, 0) = (x_{L(u)}, 0, 0),$$

and similarly  $f_w(y \circ \pi_{w \rightarrow u}) = (y_{L(u)}, 0, 0)$  for each  $y \in \{-1, 1\}^M$ . As a result, the value of  $f$  conditioned on  $u, v$ , and  $w$  is

$$\mathbf{E}_{\mathbf{x}, \mathbf{y}} \left[ \frac{1}{4} - \frac{1}{4} \langle f_v(\mathbf{x} \circ \pi_{v \rightarrow u}), f_w(\mathbf{y} \circ \pi_{w \rightarrow u}) \rangle \right] = \mathbf{E}_{\mathbf{x}, \mathbf{y}} \left[ \frac{1}{4} - \frac{1}{4} \langle (\mathbf{x}_{L(u)}, 0, 0), (\mathbf{y}_{L(u)}, 0, 0) \rangle \right],$$

which is just the value of the  $L(u)$ -th embedded dictator on the noisy hypercube, i.e.  $1/4 - 1/4\rho$ .

Now we average over  $\mathbf{u}, \mathbf{v}, \mathbf{w}$ . Because  $\mathcal{L}$  is a biregular Unique Games instance, it is in particular left-regular, and so picking a random vertex  $\mathbf{u} \in U$  and neighbor  $\mathbf{v} \in N(\mathbf{u})$  is equivalent to picking a uniformly random edge from  $E$ . Therefore, by the union bound, the probability that the assignment  $L$  satisfies both edges  $(\mathbf{u}, \mathbf{v})$  and  $(\mathbf{u}, \mathbf{w})$  is at least  $1 - 2\gamma$ . As we have seen, conditioned on this event, the assignment  $f$  has value at least  $1/4 - 1/4\rho$ . Due to our choice of  $\gamma$ , we can lower-bound the value of  $f$  by

$$\begin{aligned} (1 - 2\gamma) \cdot \left( \frac{1}{4} - \frac{1}{4}\rho \right) &\geq \frac{1}{4} - \frac{1}{4}\rho - \gamma \\ &\geq \frac{1}{4} - \frac{1}{4}\rho - \frac{1}{2}\epsilon. \end{aligned} \tag{47}$$

This completes the completeness case.

*Soundness.* We will show the contrapositive. Suppose there is a product state assignment  $\{f_v\}_{v \in V}$  to  $G$  with value at least  $\frac{1}{4} - \frac{1}{4}F^*(3, \rho) + \frac{1}{2}\epsilon$ . We will use this to construct a randomized assignment  $\mathbf{L} : U \cup V \rightarrow [M]$  whose average value is at least  $\gamma$ , which implies that the Unique Games instance has value at least  $\gamma$ .

For each  $u \in U$ , we define the function  $g_u : \{-1, 1\}^M \rightarrow B^3$  as

$$g_u(x) = \mathbf{E}_{v \sim N(u)} [f_v(x \circ \pi_{v \rightarrow u})].$$

Then we can rewrite the value of the assignment  $\{f_v\}$  as

$$\begin{aligned} \mathbf{E}_{\mathbf{u}} \mathbf{E}_{v, w \sim N(\mathbf{u})} \mathbf{E}_{\mathbf{x}, \mathbf{y}} \left[ \frac{1}{4} - \frac{1}{4} \langle f_v(\mathbf{x} \circ \pi_{v \rightarrow u}), f_w(\mathbf{y} \circ \pi_{w \rightarrow u}) \rangle \right] &= \mathbf{E}_{\mathbf{u}} \mathbf{E}_{\mathbf{x}, \mathbf{y}} \left[ \frac{1}{4} - \frac{1}{4} \langle g_u(\mathbf{x}), g_u(\mathbf{y}) \rangle \right] \\ &= \mathbf{E}_{\mathbf{u}} \left[ \frac{1}{4} - \frac{1}{4} \mathbf{Stab}_\rho[g_u] \right]. \end{aligned}$$

Since  $f$  has value at least  $\frac{1}{4} - \frac{1}{4}F^*(3, \rho) + \frac{1}{2}\epsilon$ , an averaging argument implies that at least an  $\epsilon/4$  fraction of  $u \in U$  satisfy  $\frac{1}{4} - \frac{1}{4}\mathbf{Stab}_\rho[g_u] \geq \frac{1}{4} - \frac{1}{4}F^*(3, \rho) + \epsilon/4$ . Rearranging, these  $u$ 's satisfy

$$\mathbf{Stab}_\rho[g_u] \leq F^*(3, \rho) - \epsilon.$$

We call any such  $u$  “good”. We now apply the soundness of our dictatorship test to the good  $u$ 's: by [Theorem 10.1](#), any such  $u$  has a “notable” coordinate, i.e. an  $i$  such that  $\mathbf{Inf}_i^{\leq m}[g_u] > \delta$ . Our random assignment will then use this  $i$  as its label for the vertex  $u$ :  $\mathbf{L}(u) = i$ . (If  $u$  has multiple notable coordinates, then we pick one of these arbitrarily as the label for  $u$ .)

Next we'll need to obtain labels for the neighbors of  $u$ . We will use the condition that  $u$  is good to derive that many of  $u$ 's neighbors  $v$  have notable coordinates. This requires relating the Fourier spectrum of  $g_u$  to the Fourier spectra of the neighboring  $f_v$ 's. To begin, for any subset  $S \subseteq [M]$ ,

$$\chi_S(x \circ \pi_{v \rightarrow u}) = \prod_{i \in S} (x \circ \pi_{v \rightarrow u})_i = \prod_{i \in S} x_{\pi_{v \rightarrow u}(i)} = \prod_{j \in T} x_j = \chi_T(x),$$

where  $T = \pi_{v \rightarrow u}(S) = \{\pi_{v \rightarrow u}(i) : i \in S\}$ . As a result,

$$f_v(x \circ \pi_{v \rightarrow u}) = \sum_{S \subseteq [n]} \widehat{f}_v(S) \chi_S(x \circ \pi_{v \rightarrow u}) = \sum_{T \subseteq [n]} \widehat{f}_v(\pi_{u \rightarrow v}(T)) \chi_T(x).$$

Averaging over all  $\mathbf{v} \in N(u)$ ,

$$g_u(x) = \mathbf{E}_{\mathbf{v} \sim N(u)} [f_{\mathbf{v}}(x \circ \pi_{\mathbf{v} \rightarrow u})] = \sum_{T \subseteq [n]} \mathbf{E}_{\mathbf{v} \sim N(u)} [\widehat{f}_{\mathbf{v}}(\pi_{u \rightarrow \mathbf{v}}(T))] \chi_T(x) = \sum_{T \subseteq [n]} \widehat{g}(T) \chi_T(x).$$

Hence,

$$\begin{aligned} \delta &< \mathbf{Inf}_i^{\leq m} [g_u] \\ &= \sum_{|T| \leq m: T \ni i} \|\widehat{g}(T)\|_2^2 \\ &= \sum_{|T| \leq m: T \ni i} \left\| \mathbf{E}_{\mathbf{v} \sim N(u)} [\widehat{f}_{\mathbf{v}}(\pi_{u \rightarrow \mathbf{v}}(T))] \right\|_2^2 \\ &\leq \sum_{|T| \leq m: T \ni i} \mathbf{E}_{\mathbf{v} \sim N(u)} \left\| \widehat{f}_{\mathbf{v}}(\pi_{u \rightarrow \mathbf{v}}(T)) \right\|_2^2 && \text{(because } \|\cdot\|_2^2 \text{ is convex)} \\ &= \mathbf{E}_{\mathbf{v} \sim N(u)} \left[ \sum_{|T| \leq m: T \ni i} \left\| \widehat{f}_{\mathbf{v}}(\pi_{u \rightarrow \mathbf{v}}(T)) \right\|_2^2 \right] \\ &= \mathbf{E}_{\mathbf{v} \sim N(u)} [\mathbf{Inf}_{\pi_{u \rightarrow \mathbf{v}}(i)}^{\leq m} [f_{\mathbf{v}}]]. \end{aligned}$$

By another averaging argument, a  $\delta/2$ -fraction of  $u$ 's neighbors  $v$  satisfy  $\mathbf{Inf}_{\pi_{u \rightarrow v}(i)}^{\leq m} [f_v] \geq \delta/2$ . We call these the ‘‘good neighbors’’. For each good neighbor  $v$ , the set of possible labels

$$S_v = \{j : \mathbf{Inf}_j^{\leq m} [f_v] \geq \delta/2\}$$

is non-empty. In addition, one of these labels  $j$  satisfies  $j = \pi_{u \rightarrow v}(i)$ . On the other hand, by [Proposition 7.20](#),  $|S_v| \leq 2m/\delta$  and so this set is not too large either. For each good neighbor, we assign the label of  $\mathbf{L}(v)$  by picking a uniformly random  $j \in S_v$ . For all other vertices (i.e. those which are not good or good neighbors), we assign  $\mathbf{L}$  a random label.

Now we consider the expected number of edges in  $\mathcal{I}$  satisfied by  $\mathbf{L}$ . Given a random edge  $(\mathbf{u}, \mathbf{v})$ , the probability that  $\mathbf{u}$  is good is at least  $\epsilon/4$ ; conditioned on this, the probability that  $\mathbf{v}$  is a good neighbor is at least  $\delta/2$ . Assuming both hold, since  $S_v$  is of size at most  $2M/\delta$  and contains one label equal to  $\pi_{\mathbf{u} \rightarrow \mathbf{v}}(\mathbf{L}(\mathbf{u}))$ , then  $\mathbf{L}$  satisfies the edge  $(\mathbf{u}, \mathbf{v})$  with probability at least  $\delta/2M$ . In total,  $\mathbf{L}$  satisfies at least an

$$\frac{\epsilon \delta^2}{16M} = \gamma \tag{48}$$

fraction of the edges. This concludes the proof.

*Moving from the product state value to the maximum energy.* First, we modify the graph  $G$  to remove any self-loops. To do this, we modify the distribution on edges  $(\mathbf{v}, \mathbf{x} \circ \pi_{\mathbf{v} \rightarrow \mathbf{u}})$  and  $(\mathbf{w}, \mathbf{y} \circ \pi_{\mathbf{w} \rightarrow \mathbf{u}})$  so that  $\mathbf{x}$  and  $\mathbf{y}$  are distributed as  $\rho$ -correlated Boolean strings *conditioned on them not being equal*. This removes all self-loops, as any self-loop in the graph must have  $\mathbf{v} = \mathbf{w}$  and  $\mathbf{x} \circ \pi_{\mathbf{v} \rightarrow \mathbf{u}} = \mathbf{y} \circ \pi_{\mathbf{w} \rightarrow \mathbf{u}}$ , which implies that  $\mathbf{x} = \mathbf{y}$ . (Note that this also removes some edges which are *not* self-loops, namely those for which  $\mathbf{v} \neq \mathbf{w}$ .) Given that  $\rho$ -correlated  $\mathbf{x}$  and  $\mathbf{y}$  are equal with probability

$$\left(\frac{1}{2} + \frac{1}{2}\rho\right)^M \leq \epsilon/4, \tag{49}$$

removing this event can only change the product state value of the graph by at most  $\epsilon/4$ .

Next, we apply [Corollary 7.12](#) to bound the value of  $H_G$  over general states in the soundness case. To do so, let us write  $E$  for the edges of  $G$ , and define  $p_{v,x} = \frac{1}{2} \Pr_{e \sim E}[e \text{ contains } (v, x)]$ . Note that the distribution of a random edge  $(v, \mathbf{x} \circ \pi_{v \rightarrow u})$  and  $(w, \mathbf{y} \circ \pi_{w \rightarrow u})$  is symmetric and never contains self-loops, and so  $p_{v,x} = \Pr[(w, \mathbf{y} \circ \pi_{w \rightarrow u}) = (v, x)]$ . But the UG instance  $\mathcal{I}$  is biregular, and so  $w$  is just a uniformly random element of  $V$ , and  $\mathbf{y}$  is just a uniformly random string in  $\{-1, 1\}^M$ . Hence,  $p_{v,x} = |V|^{-1} 2^{-M}$  for each  $v, x$ , and so  $\max_{v,x} \{p_{v,x}\} = |V|^{-1} 2^{-M}$ .

The next thing we have to bound to apply [Corollary 7.12](#) is the maximum of

$$\Pr[(v, \mathbf{x} \circ \pi_{v \rightarrow u}) = (v', x') \mid (w, \mathbf{y} \circ \pi_{w \rightarrow u}) = (w', y')]$$

over all  $v', w' \in V$  and  $x', y' \in \{-1, 1\}^M$ . Note that if we condition on a fixed value for  $u$  and on the event that  $v = v'$ , then this is just the maximum probability that  $\mathbf{x} \circ \pi_{v \rightarrow u}$  equals a fixed string, given that  $\mathbf{x}$  is  $\rho$ -correlated but not equal to  $y'$ . Given that  $\mathbf{x}$  is most likely to be  $-y'$  since  $\rho$  is negative, this probability is

$$\frac{1}{1 - (\frac{1}{2} + \frac{1}{2}\rho)^M} \cdot (\frac{1}{2} - \frac{1}{2}\rho)^M \leq 2 \cdot (\frac{1}{2} - \frac{1}{2}\rho)^M,$$

where we normalized by  $1 - (\frac{1}{2} + \frac{1}{2}\rho)^M$  due to the condition that  $\mathbf{x} \neq \mathbf{y}$ . Then averaging over  $u$  and  $v$  can only decrease this bound.

Now we can apply [Corollary 7.12](#), which states that

$$\text{QMAX-CUT}(G) \leq \text{PROD}(G) + 20 \cdot (2 \cdot (\frac{1}{2} - \frac{1}{2}\rho)^M)^{1/8} + \frac{1}{|V|^{1/2M}},$$

which is at most  $\epsilon/4$  by our choice of  $M$ . Hence,

$$\text{QMAX-CUT}(G) \leq \frac{1}{4} - \frac{1}{4}F^*(3, \rho) + \epsilon, \tag{50}$$

which completes the proof. □

# Part IV

## Appendix

### A The non-commutative Sum of Squares hierarchy

The Sum of Squares (SoS) hierarchy gives a canonical method for strengthening the basic SDP to achieve better approximation ratios. It features a tunable parameter  $d$ ; as  $d$  is increased, the quality of the approximation improves, but the runtime needed to compute the optimum increases as well. We will give a didactic overview of the SoS hierarchy in order to explain how our basic SDP for QUANTUM MAX-CUT arises naturally as the level-2 SoS relaxation. For a more extensive treatment of Sum of Squares, consult the excellent notes at [BS16].

#### A.1 Sum-of-squares relaxations for MAX-CUT

We begin with the sum-of-squares relaxation for the MAX-CUT problem, which generalizes the basic SDP from Definition 2.11. In fact, we will state the SoS hierarchy in terms of a general polynomial optimization problem over Boolean (i.e.  $\pm 1$ ) variables. Let  $\mathcal{I}$  be a finite set and  $x = \{x_i\}_{i \in \mathcal{I}}$  be a set of indeterminates indexed by  $\mathcal{I}$ . Let  $p(x)$  be a polynomial in the  $x_i$ 's, and consider the optimization problem

$$\begin{aligned} \max \quad & p(x) \\ \text{s.t.} \quad & x_i^2 = 1, \forall i \in \mathcal{I}. \end{aligned}$$

For example, MAX-CUT is the case when  $\mathcal{I} = V$  and  $p(x) = \mathbf{E}_{(\mathbf{u}, \mathbf{v}) \sim E}[\frac{1}{2} - \frac{1}{2}x_{\mathbf{u}}x_{\mathbf{v}}]$ . An alternative way to write this maximization is over probability distributions  $\mu$  on  $\pm 1$  assignments, i.e. functions  $\mu : \{-1, 1\}^{\mathcal{I}} \rightarrow \mathbb{R}^{\geq 0}$  such that  $\sum_x \mu(x) = 1$ . Then the optimum value is equal to

$$\begin{aligned} \max_{\mu} \quad & \mathbf{E}_{\mu}[p(x)] = \sum_{x \in \{-1, 1\}^{\mathcal{I}}} \mu(x) \cdot p(x) \\ \text{s.t.} \quad & \mu \text{ is a probability distribution,} \end{aligned}$$

because we can take  $\mu$  to have support only on the optimizing  $x$ 's. Note that because  $\mu$  is a probability distribution, it satisfies two properties: (i)  $\mathbf{E}_{\mu}[1] = 1$ , (ii) for each polynomial  $q(x)$ ,  $\mathbf{E}_{\mu}[q(x)^2] \geq 0$ . Indeed, both of these properties hold pointwise, for all  $x$ . The SoS hierarchy replaces this optimization over probability distributions with an optimization over ‘pseudo-distributions’ while partially maintaining these two properties.

**Definition A.1** (The SoS hierarchy for Boolean optimization problems). Let  $\mu : \{-1, 1\}^{\mathcal{I}} \rightarrow \mathbb{R}$  be a function. Given a polynomial  $q(x)$ , we write

$$\tilde{\mathbf{E}}_{\mu}[q(x)] = \sum_{x \in \{-1, 1\}^{\mathcal{I}}} \mu(x) \cdot q(x).$$

We say that  $\mu$  is a *degree- $d$  pseudo-distribution* if  $\tilde{\mathbf{E}}_{\mu}[1] = 1$  and  $\tilde{\mathbf{E}}_{\mu}[q(x)^2] \geq 0$  for all polynomials  $q$  of degree at most  $d/2$ . In this case, we say that  $\tilde{\mathbf{E}}_{\mu}[\cdot]$  is a *degree- $d$  pseudo-expectation*. The value of the degree- $d$  SoS relaxation is simply the maximum of  $\tilde{\mathbf{E}}_{\mu}[p(x)]$  over all pseudo-distributions  $\mu$ .

It can be shown that the value of the degree-2 SoS is equal to the basic SDP.

## A.2 Sum-of-squares relaxations for QUANTUM MAX-CUT

Now, we extend the Sum of Squares hierarchy to optimization problems over quantum states. We will consider states consisting of qubits indexed by a finite set  $\mathcal{I}$ , i.e. unit vectors in  $(\mathbb{C}^2)^{\otimes \mathcal{I}}$ . Let  $H$  be a square matrix acting on  $(\mathbb{C}^2)^{\otimes \mathcal{I}}$  and consider the optimization problem

$$\begin{aligned} \max \quad & \text{tr}[\rho \cdot H] \\ \text{s.t.} \quad & \rho \text{ is a density matrix.} \end{aligned}$$

For example, QUANTUM MAX-CUT is the case when  $\mathcal{I} = V$  and  $H = \mathbf{E}_{(u,v) \sim E}[h_{u,v}]$ . Because  $\rho$  is a density matrix, it satisfies three properties: (1)  $\rho$  is Hermitian, (2)  $\text{tr}[\rho \cdot I] = 1$ , and (3) for any matrix  $M$ ,  $\text{tr}[\rho \cdot M^\dagger M] \geq 0$ . The last of these is because  $A = M^\dagger M$  is PSD, and a matrix  $\rho$  is PSD if and only if  $\text{tr}[\rho \cdot A] \geq 0$  for all PSD matrices  $A$ . The SoS hierarchy will instead optimize over “pseudo-density matrices” while partially maintaining these three properties. To begin, we first define a matrix analogue of degree.

**Definition A.2** (The basis of Pauli matrices). The Pauli matrices  $\{I, X, Y, Z\}^{\otimes n}$  form an orthogonal basis for the set of  $2^n \times 2^n$  matrices. Given  $P, Q \in \{I, X, Y, Z\}^{\otimes n}$ , they satisfy

$$\text{tr}[PQ] = \begin{cases} 2^n & \text{if } P = Q, \\ 0 & \text{otherwise.} \end{cases}$$

Given a  $2^n \times 2^n$  matrix  $M$ , we write  $\widehat{M}(P)$  for the coefficient of  $M$  on  $P$  in this basis. In other words,

$$M = \sum_{P \in \{I, X, Y, Z\}^{\otimes n}} \widehat{M}(P) \cdot P.$$

In addition,  $M$  is Hermitian if and only if  $\widehat{M}(P)$  is real, for all  $P$ .

**Definition A.3** (Degree of a matrix). Given  $P \in \{I, X, Y, Z\}^{\otimes n}$ , the *degree* of  $P$ , denoted  $|P|$ , is the number of qubits on which  $P$  is not the  $2 \times 2$  identity matrix. More generally, we say that a  $2^n \times 2^n$  matrix  $M$  has degree- $d$  if  $\widehat{M}(P) = 0$  for all  $|P| > d$ .

Now we describe the analogue of the sum-of squares hierarchy for quantum states, which is known as the *NPA* or *non-commutative Sum of Squares (ncSoS) hierarchy*.

**Definition A.4** (The ncSoS hierarchy for quantum optimization problems). Let  $\rho$  and  $M$  be square matrices acting on  $(\mathbb{C}^2)^{\mathcal{I}}$ . We write

$$\widetilde{\mathbf{E}}_\rho[M] = \text{tr}[\rho \cdot M].$$

We say that  $\rho$  is a *degree- $d$  pseudo-density matrix* if  $\rho$  is Hermitian,  $\widetilde{\mathbf{E}}_\rho[I] = 1$ , and  $\widetilde{\mathbf{E}}_\rho[M^\dagger M] \geq 0$  for all matrices  $M$  of degree at most  $d/2$  (cf. [Definition A.3](#)). In this case, we say that  $\widetilde{\mathbf{E}}_\rho[\cdot]$  is a *degree- $d$  pseudo-expectation*. The value of the degree- $d$  ncSoS relaxation is simply the maximum of  $\widetilde{\mathbf{E}}_\rho[H]$  over all pseudo-distributions  $\rho$ .

**Remark A.5** (Convergence of the SoS relaxation). When  $d = 2n$ , where  $n = |\mathcal{I}|$ , the SoS relaxation solves the optimization problem exactly. This is because every square matrix  $M$  acting on  $(\mathbb{C}^2)^{\mathcal{I}}$  is degree- $n$ ; thus  $\text{tr}[\rho \cdot M^\dagger M] = \widetilde{\mathbf{E}}_\rho[M^\dagger M] \geq 0$ , and so  $\rho$  must be positive semidefinite.

### A.3 Degree-two non-commutative Sum of Squares

Now we analyze the degree-2 ncSoS relaxation for QUANTUM MAX-CUT and show that it coincides with the basic SDP we considered in [Section 7.5](#). We begin with a definition.

**Definition A.6** (Degree- $d$  slice of a matrix). Given a  $2^n \times 2^n$  matrix  $M$ , we write  $M^{=d}$  for its degree- $d$  component, i.e.

$$M^{=d} = \sum_{P:|P|=d} \widehat{M}(P) \cdot P.$$

Let  $\rho$  be a feasible solution to the degree-2 ncSoS relaxation. We will begin by showing that we may assume without loss of generality that  $\rho$  only has degree 0 and 2 components, i.e. that  $\rho = \rho^{=0} + \rho^{=2}$ . Prior to showing this, we will need a technical lemma.

**Lemma A.7.** *Let  $P, Q, R \in \{I, X, Y, Z\}^{\otimes n}$ . Then  $\text{tr}(PQR) = (-1)^{|P|+|Q|+|R|} \cdot \text{tr}(PRQ)$ .*

*Proof.* First, we prove this for  $n = 1$ . When  $n = 1$ , both sides are zero unless  $QR$  is the same Pauli matrix as  $P$ , up to a multiplicative constant. Suppose this is so. If one of  $P, Q$ , or  $R$  is the identity matrix, then the other two are equal to each other, and so  $\text{tr}(PQR) = \text{tr}(PRQ)$ . This satisfies the equality because  $|P| + |Q| + |R|$  is either 0 or 2 in this case. Otherwise, none of  $P, Q$ , or  $R$  is the identity matrix, and so they are distinct Pauli matrices, which means  $Q$  and  $R$  anticommute. So  $\text{tr}(PQR) = -\text{tr}(PRQ)$ , satisfying the equality because  $|P| + |Q| + |R| = 3$  in this case.

Now, the general  $n$  case follows from the  $n = 1$  case because

$$\text{tr}(PQR) = \prod_{i=1}^n \text{tr}(P_1 Q_1 R_1) = \prod_{i=1}^n \left( (-1)^{|P_1|+|Q_1|+|R_1|} \cdot \text{tr}(P_1 R_1 Q_1) \right) = (-1)^{|P|+|Q|+|R|} \cdot \text{tr}(PRQ).$$

This completes the proof.  $\square$

**Proposition A.8** (Restricting to the degree-0 and degree-2 slices). *Let  $\rho$  be a feasible solution. Then  $\rho^{=0} + \rho^{=2}$  is a feasible solution with the same value as  $\rho$ .*

*Proof.* To begin, we claim that  $\rho' = \rho^{=0} + \rho^{=1} + \rho^{=2}$  is a feasible solution with the same value as  $\rho$ . This is because the constraints  $\widetilde{\mathbf{E}}_{\rho}[I] = 1$  and  $\widetilde{\mathbf{E}}_{\rho}[M^\dagger M] \geq 0$  and the objective  $\widetilde{\mathbf{E}}_{\rho}[H_G]$  feature matrices of degree at most 2, and so these values are unchanged if we replace  $\rho$  with  $\rho^{=0} + \rho^{=1} + \rho^{=2}$ .

Next, we claim that  $\rho'' = \rho^{=0} - \rho^{=1} + \rho^{=2}$  is a feasible solution with the same value as  $\rho'$ . The constraint  $\widetilde{\mathbf{E}}_{\rho''}[I] = 1$  and the value  $\widetilde{\mathbf{E}}_{\rho''}[H_G]$  feature matrices which have no degree-1 terms, so negating  $\rho^{=1}$  doesn't affect these expressions. As for the remaining constraint, for each degree-1 matrix  $M$ ,

$$\begin{aligned} \widetilde{\mathbf{E}}_{\rho''}[M^\dagger M] &= \text{tr}((\rho^{=0} - \rho^{=1} + \rho^{=2}) \cdot M^\dagger M) \\ &= \sum_{|P| \leq 2} \sum_{|Q|, |R| \leq 1} (-1)^{|P|} \cdot \widehat{\rho}(P) \widehat{M}^\dagger(Q) \widehat{M}(R) \cdot \text{tr}(PQR) \\ &= \sum_{|P| \leq 2} \sum_{|Q|, |R| \leq 1} (-1)^{|R|+|Q|} \cdot \widehat{\rho}(P) \widehat{M}^\dagger(Q) \widehat{M}(R) \cdot \text{tr}(PRQ) && \text{(by Lemma A.7)} \\ &= \text{tr}((\rho^{=0} + \rho^{=1} + \rho^{=2}) \cdot (M^{=0} - M^{=1})(M^{=0} - M^{=1})^\dagger) \\ &= \widetilde{\mathbf{E}}_{\rho'}[(M^{=0} - M^{=1})(M^{=0} - M^{=1})^\dagger] \geq 0. \end{aligned}$$

Hence,  $\rho''$  satisfies this constraint as well.

We conclude by noting that  $\frac{1}{2}(\rho' + \rho'') = \rho^{=0} + \rho^{=2}$  is a feasible solution with the same value as  $\rho$ , because the constraints and the objective are linear functions of  $\rho$ .  $\square$

Henceforth, we assume  $\rho = \rho^{=0} + \rho^{=2}$ . Using this, we note that the constraint  $\widetilde{\mathbf{E}}_\rho[M^\dagger M] \geq 0$  holding for all  $M$  which are degree-1 is equivalent to it holding only for  $M$  which are homogeneous degree-1 (i.e. with no degree-0 term). This is because if we write  $M = M^{=0} + M^{=1} = \widehat{M}(I) \cdot I + M^{=1}$ , then

$$\begin{aligned} \text{tr}(\rho \cdot M^\dagger M) &= |\widehat{M}(I)|^2 \cdot \text{tr}(\rho) + \widehat{M}(I)^\dagger \cdot \text{tr}(\rho \cdot M^{=1}) + \widehat{M}(I) \cdot \text{tr}(\rho \cdot (M^{=1})^\dagger) + \text{tr}(\rho \cdot (M^{=1})^\dagger M^{=1}) \\ &= |\widehat{M}(I)|^2 + \text{tr}(\rho \cdot (M^{=1})^\dagger M^{=1}) \quad (\text{because } \rho \text{ has no degree-1 component}) \\ &\geq \text{tr}(\rho \cdot (M^{=1})^\dagger M^{=1}), \end{aligned}$$

which is  $\geq 0$  because  $\widetilde{\mathbf{E}}_\rho[(M^{=1})^\dagger M^{=1}] \geq 0$ .

Now, we let  $R(\cdot, \cdot)$  be the  $3n \times 3n$  matrix whose rows and columns are indexed by degree-1 Pauli matrices such that

$$R(P_i, Q_j) = \text{tr}(\rho \cdot P_i Q_j).$$

for all  $P, Q \in \{X, Y, Z\}$  and  $i, j \in \{1, \dots, n\}$ . When  $i \neq j$  or  $i = j$  and  $P = Q$ , then  $P_i Q_j$  is a Pauli matrix in  $\{I, X, Y, Z\}^{\otimes n}$ , and so

$$R(P_i, Q_j) = \text{tr}(\rho \cdot P_i Q_j) = 2^n \cdot \widehat{\rho}(P_i Q_j),$$

which is a real number. On the other hand, when  $i = j$  but  $P \neq Q$ , then  $P_i Q_j = P_i Q_i$  is a degree-1 Pauli matrix times a phase of  $i$  or  $-i$ . In this case,

$$R(P_i, Q_i) = \text{tr}(\rho \cdot P_i Q_i) = 0,$$

because  $\rho$  has no degree-1 component. Put together, these imply that  $R$  is a real-valued matrix.

We can now rewrite our constraints and objective function in terms of this matrix. First, the constraint  $\text{tr}(\rho) = 1$  corresponds to

$$R(P_i, P_i) = \text{tr}(\rho \cdot P_i P_i) = \text{tr}(\rho) = 1$$

for any  $P_i$ . Next, the objective function is

$$\begin{aligned} \text{tr}(\rho \cdot H_G) &= \mathbf{E}_{(i,j) \sim E} \text{tr}(\rho \cdot h_{i,j}) \\ &= \mathbf{E}_{(i,j) \sim E} \text{tr}(\rho \cdot \frac{1}{4} \cdot (I_i \otimes I_j - X_i \otimes X_j - Y_i \otimes Y_j - Z_i \otimes Z_j)) \\ &= \frac{1}{4} - \frac{1}{4} \mathbf{E}_{(i,j) \sim E} \sum_{P \in \{X, Y, Z\}} \text{tr}(\rho \cdot P_i \otimes P_j) \\ &= \frac{1}{4} - \frac{1}{4} \mathbf{E}_{(i,j) \sim E} \sum_{P \in \{X, Y, Z\}} R(P_i, P_j). \end{aligned}$$

For the last constraint, let  $M = \sum_{P_i} \widehat{M}(P_i) \cdot P_i$  be any homogeneous degree-1 matrix. Then

$$\begin{aligned} 0 \leq \text{tr}(\rho \cdot M^\dagger M) &= \sum_{P_i, Q_j} \widehat{M}(P_i)^\dagger \widehat{M}(Q_j) \cdot \text{tr}(\rho \cdot P_i Q_j) \\ &= \sum_{P_i, Q_j} \widehat{M}(P_i)^\dagger \widehat{M}(Q_j) \cdot R(P_i, Q_j) = \text{vec}(M)^\dagger \cdot R \cdot \text{vec}(M), \end{aligned}$$

where  $\text{vec}(M)$  is the height- $3n$  vector with  $\text{vec}(M)(P_i) = \widehat{M}(P_i)$ . As the  $\widehat{M}(P_i)$ 's are allowed to be arbitrary complex numbers, this condition is equivalent to  $R$  being positive semidefinite. As a result, this matrix has the exact same form and objective as the matrix  $M'(\cdot, \cdot)$  from [Section 7.5](#); following the steps in that proof, one can then convert  $R$  into a solution to the basic SDP. This completes the proof.

## B Other Lemmas

### B.1 Cardinality Reduction

In this section, we prove [Lemma 8.5](#), using an argument which closely follows [[OW08](#), Appendix B]. All of these transformations are standard.

*Proof.* First, we give a series of transformations to yield a well-behaved finite graph. For each transformation, we argue that SDP value and product state value are within  $\epsilon$  of the original graph. Finally, we show that in our final graph  $G'$ , using [Corollary 7.12](#),  $\text{PROD}(G') \geq \text{QMAX-CUT}(G') - \epsilon$ , concluding the proof.

Let  $G_0 = G$ . We will first construct  $G_1$ , which restricts the Gaussian graph  $\mathcal{G}_\rho^n$  to the sphere graph  $S^{n-1}$ . Next,  $G_2$  will be a graph on a finite vertex set. Following that, our final graph  $G' = G_3$  will remove self-loops, leaving us with a weighted, simple graph with finite vertex set. In this proof, we will identify a graph by the distribution on its edges. For  $u, v \subseteq S^{n-1}$ , we will write  $G(u, v)$  for the probability weight  $G$  puts on edges  $(u, v)$ .

We start with the construction for  $G_1$ . For  $G_0$ , let  $f : \mathbb{R}^n \rightarrow S^{n-1}$  be an SDP assignment obtaining  $\text{SDP}_{\text{QMC}}(G_0)$ . (Note that this is also an optimal SDP assignment for  $\text{SDP}_{\text{PROD}}(G_0)$ .) Let  $G_1$  be the graph in which  $G_1(u, v) = G_0(f^{-1}(u), f^{-1}(v))$ . Then if we take the identity map as the SDP embedding, we see that

$$\text{SDP}_{\text{QMC}}(G_1) \geq \mathbf{E}_{(u,v) \sim G_1} \left[ \frac{1}{4} - \frac{3}{4} \langle \mathbf{u}, \mathbf{v} \rangle \right] = \mathbf{E}_{\mathbf{u} \sim_\rho \mathbf{v}} \left[ \frac{1}{4} - \frac{3}{4} \langle f(\mathbf{u}), f(\mathbf{v}) \rangle \right] = \text{SDP}_{\text{QMC}}(G_0) =: c_{\text{H}}.$$

A similar argument shows that  $\text{SDP}_{\text{PROD}}(G_1) \geq \text{SDP}_{\text{PROD}}(G_0) =: c_{\text{PROD}}$ . Furthermore, for any assignment  $h : S^{n-1} \rightarrow S^2$  on  $G_1$ , the assignment  $h \circ f$  yields an assignment for  $G_0$  and thus  $\text{PROD}(G_1) \leq \text{PROD}(G_0)$ .

To construct  $G_2$ , we use an argument originally from [[FS02](#)]. Pick some  $\epsilon$ -net  $\mathcal{N}$  over  $S^{n-1}$ , so that every point in  $S^{n-1}$  is within distance  $\epsilon$  to some point in  $\mathcal{N}$ ; it is known that constructions exist with  $|\mathcal{N}| \leq 1/\epsilon^{O(d)}$ . Then partition  $S^{n-1}$  using Voronoi cells  $\{C_v\}_{v \in \mathcal{N}}$  based on  $\mathcal{N}$ . For each  $v \in \mathcal{N}$ , the corresponding cell  $C_v \subseteq S^{n-1}$  consists of all points in  $S^{n-1}$  which are closer to  $v$  than any other  $u \in \mathcal{N}$ . Then  $G_2$  is the finite graph on vertex set  $\mathcal{N}$  in which  $G_2(u, v) = G_1(C_u, C_v)$ . We first observe that

$$\text{PROD}(G_2) \leq \text{PROD}(G_1) = s$$

since any assignment  $f$  on  $G_2$  can be extended to an assignment of equal value on  $G_1$ . Furthermore, we claim

$$\text{SDP}_{\text{QMC}}(G_2) \geq c_{\text{H}} - 3\epsilon.$$

To see this, consider the SDP assignment  $f : \mathcal{N} \rightarrow S^{n-1}$  which maps each  $v \in \mathcal{N}$  to itself. We can extend this to a function with domain all of  $S^{n-1}$  by setting  $f(u) = v$  for each  $u \in C_v$ . Then

$$\text{SDP}_{\text{QMC}}(G_2) \geq \mathbf{E}_{(u,v) \sim G_2} \left[ \frac{1}{4} - \frac{3}{4} \langle f(\mathbf{u}), f(\mathbf{v}) \rangle \right] = \mathbf{E}_{(x,y) \sim G_1} \left[ \frac{1}{4} - \frac{3}{4} \langle f(\mathbf{x}), f(\mathbf{y}) \rangle \right]. \quad (51)$$

Let  $C_{\mathbf{u}}$  be the Voronoi cell  $\mathbf{x}$  falls inside and  $C_{\mathbf{v}}$  be the Voronoi cell  $\mathbf{y}$  falls inside. Then because  $\mathcal{N}$  is an  $\epsilon$ -net, we can write  $\mathbf{x} = (\mathbf{u} + \boldsymbol{\eta}_1)$  and  $\mathbf{y} = (\mathbf{v} + \boldsymbol{\eta}_2)$ , where  $\boldsymbol{\eta}_1$  and  $\boldsymbol{\eta}_2$  have length at most  $\epsilon$ . Thus,

$$\langle \mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{u} + \boldsymbol{\eta}_1, \mathbf{v} + \boldsymbol{\eta}_2 \rangle \geq \langle \mathbf{u}, \mathbf{v} \rangle - 3\epsilon = \langle f(\mathbf{x}), f(\mathbf{y}) \rangle - 3\epsilon.$$

As a result,

$$(51) \geq \mathbf{E}_{(x,y) \sim G_1} \left[ \frac{1}{4} - \frac{3}{4} \langle \mathbf{x}, \mathbf{y} \rangle \right] - 3\epsilon = c_{\text{H}} - 3\epsilon.$$

A similar argument shows that  $\text{SDP}_{\text{PROD}}(G_2) \geq c_{\text{PROD}} - 3\epsilon$ .

Finally, we use a simple construction appearing in [KO09] (and originally due to [ABH+05]) in order to remove self-loops. Conveniently, this construction will also make it easy to show that the product state value and maximum energy are close. Our graph  $G_3$  will be parameterized by an integer  $M$  which we will select later but which is at least  $1/\epsilon$ . For each vertex  $v \in G_2$ , we will create  $M$  many vertices  $\{(v, j)\}_{j \in [M]}$ , each with weight  $\frac{1}{M}$  of the original. To sample a random edge in  $G_3$ , we simply sample  $(\mathbf{u}, \mathbf{v})$  from  $G_2$ , let  $\mathbf{i}, \mathbf{j} \in [M]$  be independent, uniformly random, and output the edge between  $(\mathbf{u}, \mathbf{i})$  and  $(\mathbf{v}, \mathbf{j})$ .

It is clear that  $\text{PROD}(G_3) \geq \text{PROD}(G_2)$  because any assignment  $f : \mathcal{V} \rightarrow S^2$  can be converted into an assignment  $f'$  for  $G_3$  of equal value by setting  $f'(u, i) = f(u)$ . On the other hand,  $\text{PROD}(G_2) \geq \text{PROD}(G_3)$  as well. To see this, consider a product state assignment  $f : \mathcal{V} \times [M] \rightarrow S^2$  for  $G_3$ . It has value

$$\mathbf{E}_{(\mathbf{u}, \mathbf{v}) \sim G_2} \mathbf{E}_{\mathbf{i}, \mathbf{j} \sim [M]} \left[ \frac{1}{4} - \frac{1}{4} \langle f(\mathbf{u}, \mathbf{i}), f(\mathbf{v}, \mathbf{j}) \rangle \right] = \mathbf{E}_{\mathbf{u}, \mathbf{v} \sim G_2} \left[ \frac{1}{4} - \frac{1}{4} \langle \mathbf{E}_{\mathbf{i}} f(\mathbf{u}, \mathbf{i}), \mathbf{E}_{\mathbf{j}} f(\mathbf{v}, \mathbf{j}) \rangle \right].$$

This is the value that the assignment  $f' : \mathcal{V} \rightarrow B^3$  defined as  $f'(u) = \mathbf{E}_{\mathbf{i}} f(u, \mathbf{i})$  achieves on the graph  $G_2$ . As  $f'$  has range  $B^3$ , there exists a function with range  $S^2$  whose value is at least as high. As a result,  $\text{PROD}(G_3) = \text{PROD}(G_2)$ . A similar argument shows that the two SDP values remain the same as well. Now, the total weight of self-loops in this graph is at most  $\frac{1}{M}$ , and so by removing these edges and scaling the remaining weights to sum to one we produce a graph  $G'$  with no self-loops in which the SDP and product state values have increased by at most  $\frac{1}{M} \leq \epsilon$ .

Next, we use [Corollary 7.12](#) to relate the product state value to the optimal state value. Recall that we need to bound the quantity

$$20 \cdot \left( n \cdot \max_{(u, i), (v, j)} \{A_{(u, i), (v, j)}\} \cdot \max_{(u, i)} \{p_{(u, i)}\} \right)^{1/8} + \max_{(u, i)} \{p_{(u, i)}\} \quad (52)$$

for  $G'$ . Here,  $n$  is the number of vertices in  $G_3$ ,  $A_{(u, i), (v, j)}$  is the probability of an edge ending in  $(u, i)$  conditioned on starting from  $(v, j)$ , and  $p_{(u, i)}$  is one half the total weight of edges on  $(u, i)$ . We'll give a bound for the above quantity *before* removing self loops (i.e. for the graph  $G_3$ ). However, observe that self loops consist of at most  $1/M$  of the total edge weight and thus,

$$p'_{u, i} \leq p_{u, i} / \left(1 - \frac{1}{M}\right),$$

$$A'_{(u, i), (v, j)} \leq A_{(u, i), (v, j)} / \left(1 - \frac{1}{M}\right),$$

where  $p'_{u, i}$  and  $A'_{(u, i), (v, j)}$  are the quantities after removing self-loops (i.e. for the graph  $G'$ ). Choosing  $M$  sufficiently large makes this difference negligible.

Now, observe that if  $p_u^{(2)}$  is the weight function associated with  $G_2$ , then  $p_{u, i} = p_u^{(2)}/M$ , since each vertex  $(u, i)$  in  $G_3$  inherits  $1/M$  of the total edge weight of vertex  $u$ . Thus, by choosing  $M$  sufficiently large, we can bound the last additive term by  $\epsilon/2$ . Next, since whenever  $(v, j)$  is connected to a vertex  $(u, i)$ , it is in fact connected to all vertices  $\{(u, k)\}_{k \in [M]}$  with an equal weight, we have an easy upper bound of  $1/M$  on  $A_{(u, i), (v, j)}$ .

Finally, using that  $n = M|\mathcal{V}|$ , where  $\mathcal{V}$  is  $G_2$ 's vertex set, we can rewrite [Equation \(52\)](#) as,

$$\begin{aligned} 20 \cdot \left( M|\mathcal{V}| \cdot \max_{(u, i), (v, j)} \{A_{(u, i), (v, j)}\} \cdot \frac{1}{M} \max_u \{p_u\} \right)^{1/8} + \epsilon/2 \\ = 20 \cdot \left( |\mathcal{V}| \cdot \max_u \{p_u\} \cdot \frac{1}{M} \right)^{1/8} + \epsilon/2 \end{aligned}$$

Thus, noting that  $|\mathcal{V}| \cdot \max_u \{p_u\}$  is just a constant  $C$  which is independent of  $M$ , we can bound this equation by  $\epsilon$  by taking  $M$  sufficiently large.  $\square$

## B.2 Proofs of Various Fourier Properties

**Proposition B.1.** *Let  $f : \{-1, 1\}^n \rightarrow \mathbb{R}^k$  be a function and let  $\rho, \gamma \in [0, 1)$ . Then*

$$|\mathbf{Stab}_\rho[f] - \mathbf{Stab}_{\rho(1-\gamma)^2}[f]| \leq \frac{2\gamma}{1-\rho} \mathbf{Var}[f].$$

*Proof.* Let  $\epsilon = \rho - (1-\gamma)^2\rho \in [0, \rho)$ , so that  $\rho(1-\gamma)^2 = \rho - \epsilon$ . Then

$$|\mathbf{Stab}_\rho[f] - \mathbf{Stab}_{\rho-\epsilon}[f]| = \sum_{S \subseteq [n]} (\rho^{|S|} - (\rho - \epsilon)^{|S|}) \|\widehat{f}(S)\|_2^2.$$

When  $\epsilon = 0$ , the proposition clearly holds. Assume  $\epsilon \in (0, \rho)$ . For  $S \neq \emptyset$ , we'll bound the term  $(\rho^{|S|} - (\rho - \epsilon)^{|S|})$ . When  $S = \emptyset$  this quantity is just 0. Let  $k = |S|$ , and define the function  $g(\delta) = (\rho - \epsilon + \delta)^k$ . This function is continuous in  $\delta$ , and applying the Mean Value Theorem on  $\delta \in [0, \epsilon]$  yields a  $\delta'$  such that

$$\frac{dg}{d\delta}(\delta') = \frac{g(\epsilon) - g(0)}{\epsilon - 0} = \frac{\rho^k - (\rho - \epsilon)^k}{\epsilon}$$

Furthermore,  $g'(\delta') = k(\rho - \epsilon + \delta')^{k-1}$ . For any  $x \in (0, 1)$  and  $k \in \mathbb{N}^+$ , we have that  $(1-x)^{k-1}k \leq 1/x$ . Letting  $1-x = \rho - \epsilon + \delta'$ , we get

$$\frac{\rho^k - (\rho - \epsilon)^k}{\epsilon} = k(\rho - \epsilon + \delta')^{k-1} \leq \frac{1}{1 - \rho + \epsilon - \delta'} \leq \frac{1}{1 - \rho} \implies \rho^k - (\rho - \epsilon)^k \leq \frac{\epsilon}{1 - \rho}.$$

Thus,

$$|\mathbf{Stab}_\rho[f] - \mathbf{Stab}_{\rho-\epsilon}[f]| \leq \frac{\epsilon}{1-\rho} \sum_{S \subseteq [n]: S \neq \emptyset} \|\widehat{f}(S)\|_2^2 = \frac{\epsilon}{1-\rho} \mathbf{Var}[f].$$

Substituting in  $\epsilon = \rho - \rho(1-\gamma)^2 \leq 2\gamma$  concludes the proof.  $\square$

**Proposition B.2** (Stability Bound). *Let  $f : \{-1, 1\}^n \rightarrow \mathbb{R}$  and  $\rho \in [-1, 0]$ . Then,*

$$\mathbf{Stab}_\rho[f] \geq \rho \cdot \mathbf{E}[f(\mathbf{x})^2].$$

*Proof.* First, note that  $\mathbf{Stab}_\rho[f] = \sum_S \rho^{|S|} \widehat{f}(S)^2$ . Since  $\rho \in [-1, 0]$ , we can lower bound this by  $\rho \cdot \sum_S \widehat{f}(S)^2$ . Finally, using Parseval's theorem, this is exactly equal to  $\rho \cdot \mathbf{E}[f(\mathbf{x})^2]$ .  $\square$

## B.3 Proofs of Lipschitz Properties

**Lemma B.3** (Lipschitz Property of  $\Psi$ ). *Let  $\Psi : \mathbb{R}^n \rightarrow \mathbb{R}$  be defined as,*

$$\Psi(v) = \begin{cases} \|v\|_2^2 & \text{if } \|v\|_2 \leq 1, \\ 1 & \text{otherwise.} \end{cases}$$

*Then,  $\Psi$  is Lipschitz continuous with constant 2. In particular, for any  $u, v \in \mathbb{R}^n$ ,*

$$|\Psi(u) - \Psi(v)| \leq 2\|u - v\|_2$$

*Proof.* The proof is by reduction to the function  $\text{Sq} : \mathbb{R} \rightarrow \mathbb{R}$ , defined as,

$$\text{Sq}(x) = \begin{cases} 0 & \text{if } x < 0, \\ x^2 & \text{if } x \in [0, 1], \\ 1 & \text{if } x > 1. \end{cases}$$

As in the proof of the Majority is Stablest theorem in [O'D14], we see that  $\text{Sq}$  is 2-Lipschitz. Fix any  $u, v \in \mathbb{R}^n$ . By applying the Lipschitz property of  $\text{Sq}$ , we can write

$$|\Psi(u) - \Psi(v)| = |\text{Sq}(\|u\|_2) - \text{Sq}(\|v\|_2)| \leq 2|\|u\|_2 - \|v\|_2|.$$

Finally, applying the reverse triangle inequality, we obtain an upper bound of  $2\|u - v\|_2$ , which concludes the proof.  $\square$

**Lemma B.4** (Lipschitz Property of  $\Phi$ ). *Let  $\Phi : \mathbb{R}^n \rightarrow \mathbb{R}$  be defined as  $\Phi(v) = v - \mathcal{R}(v)$ , where  $\mathcal{R}$  rounds vectors to the unit ball  $B^n$  and is defined as*

$$\mathcal{R}(v) = \begin{cases} v & \text{if } \|v\|_2 < 1, \\ \frac{v}{\|v\|_2} & \text{otherwise.} \end{cases}$$

*Then  $\Phi$  is Lipschitz continuous with constant 2. In particular, for any  $u, v \in \mathbb{R}^n$ ,*

$$\|\Phi(u) - \Phi(v)\|_2 \leq 2\|u - v\|_2.$$

*Proof.* First, we show that  $\mathcal{R}(v)$  is in fact 1-Lipschitz. Without loss of generality, we can assume vectors  $u, v \in \mathbb{R}^2$  and take  $u = r \cdot (1, 0)$  and  $v = s \cdot (v_1, v_2)$  where  $\|(v_1, v_2)\|_2 = 1$ . We want to show

$$\|\mathcal{R}(u) - \mathcal{R}(v)\|_2 \leq \|u - v\|_2.$$

Certainly, this holds when  $r, s \leq 1$ . Consider the case when  $r, s \geq 1$ . Then

$$\begin{aligned} \|\mathcal{R}(u) - \mathcal{R}(v)\|_2 &= \|(1, 0) - (v_1, v_2)\|_2 \\ &= \sqrt{(1 - v_1)^2 + v_2^2} \\ &= \sqrt{(1 - v_1)^2 + 1 - v_1^2} \\ &= \sqrt{2 - 2v_1}. \end{aligned}$$

On the other hand,

$$\begin{aligned} \|u - v\|_2 &= \sqrt{(r - sv_1)^2 + s^2v_2^2} \\ &= \sqrt{r^2 - 2rsv_1 + s^2v_1^2 + s^2(1 - v_1^2)} \\ &= \sqrt{r^2 - 2rsv_1 + s^2}. \end{aligned}$$

Thus, it suffices to show  $2 - 2v_1 \leq r^2 - 2rsv_1 + s^2$ . Rearranging so that all terms including  $v_1$  are on the LHS, we get

$$v_1 \cdot 2(rs - 1) \stackrel{?}{\leq} r^2 + s^2 - 2.$$

Since  $s, r \geq 1$ , the LHS is maximized for  $v_1 = 1$  and thus this is true if and only if

$$0 \stackrel{?}{\leq} r^2 + s^2 - 2rs.$$

Factoring the RHS yields  $(r - s)^2$ , which is indeed at least 0.

Finally, we consider the case when  $r \leq 1$  and  $s \geq 1$  (the case of  $r \geq 1$  and  $s \leq 1$  is symmetric and we omit it). Then we again have  $\|u - v\|_2 = \sqrt{r^2 - 2rv_1 + s^2}$ . However, we now have

$$\|\mathcal{R}(u) - \mathcal{R}(v)\|_2 = \sqrt{(r - v_1)^2 + v_2} = \sqrt{(r - v_1)^2 + 1 - v_2^2} = \sqrt{r^2 - 2rv_1 + 1}.$$

As a result, we want to show

$$\begin{aligned} r^2 - 2rv_1 + 1 &\stackrel{?}{\leq} r^2 - 2rsv_1 + s^2 \\ v_1 \cdot 2(rs - r) + 1 &\stackrel{?}{\leq} s^2 \\ 2(rs - r) + 1 &\stackrel{?}{\leq} s^2 && \text{(LHS maximized when } v_1 = 1) \\ 2s - 2 + 1 &\stackrel{?}{\leq} s^2 && \text{(LHS maximized when } r = 1) \\ 0 &\stackrel{?}{\leq} s^2 - 2s + 1 = (s - 1)^2. \end{aligned}$$

We conclude that  $\mathcal{R}(\cdot)$  is 1-Lipschitz. Now we show that  $\Phi(u) = u - \mathcal{R}(u)$  is 2-Lipschitz.

$$\begin{aligned} \|\Phi(u) - \Phi(v)\|_2 &= \|u - \mathcal{R}(u) - v + \mathcal{R}(v)\|_2 \leq \|u - v\|_2 + \|\mathcal{R}(u) - \mathcal{R}(v)\|_2 \\ &\hspace{15em} \text{(by the triangle inequality)} \\ &\leq 2\|u - v\|_2. \quad \text{(by the Lipschitz property of } \mathcal{R}(\cdot)) \end{aligned}$$

This concludes the proof.  $\square$

**Corollary B.5.** *The function  $\Phi_i(v) : \mathbb{R}^n \rightarrow \mathbb{R}$ , defined as  $\Phi_i(v) = \Phi(v)_i$ , is 2-Lipschitz.*

*Proof.* Take any  $u, v \in \mathbb{R}^n$ . Then

$$|\Phi_i(u) - \Phi_i(v)| \leq \sqrt{\sum_i^n (\Phi_i(u) - \Phi_i(v))^2} = \|\Phi(u) - \Phi(v)\|_2 \leq 2\|u - v\|_2. \quad \square$$

**Lemma B.6.** *Let  $\mathbf{x} \sim_\rho \mathbf{y}$  be  $\rho$ -correlated random variables in  $k$  dimensions. The function:*

$$\mathbf{E}_{\mathbf{x} \sim_\rho \mathbf{y}} \left\langle \frac{\mathbf{x}}{\|\mathbf{x}\|}, \frac{\mathbf{y}}{\|\mathbf{y}\|} \right\rangle = \frac{2}{k} \left( \frac{\Gamma((k+1)/2)}{\Gamma(k/2)} \right)^2 \rho {}_2F_1[1/2, 1/2, k/2 + 1, \rho^2]$$

in  $\rho$  is:

1. non-negative for  $\rho \in [0, 1]$ ,
2. an odd function,
3.  $C$ -Lipschitz for  $\rho \in [-1, 1]$  for some constant  $C$  which is a function of  $k$  if  $k \geq 3$ ,
4.  $C$ -Lipschitz for  $\rho \in [-1 + \epsilon, 1 - \epsilon]$  for some constant  $C(k, \epsilon)$  for any  $\epsilon > 0$  if  $k = 1, 2$ , and
5. convex for  $\rho \in [0, 1]$ .

*Proof.* From the definition of  ${}_2F_1$ :

$${}_2F_1[a, b, c, z] = \sum_{n=0}^{\infty} \frac{(a)_n (b)_n}{(c)_n} \frac{z^n}{n!}, \quad (53)$$

we see that  ${}_2F_1[1/2, 1/2, k/2 + 1, \rho^2] \geq 0$ , establishing the first two properties.

For the third and fourth property, since the function  $f(\rho) = \rho {}_2F_1[1/2, 1/2, k/2 + 1, \rho^2]$  is differentiable in  $\rho$  so it is Lipschitz if we can upper bound the absolute value of the derivative in the interval containing  $\rho$ . By 15.2.1 in [AS72],

$$\frac{d}{dz} {}_2F_1[a, b, c, z] = \frac{ab}{c} {}_2F_1[a + 1, b + 1, c + 1, z]$$

Applying the product rule, taking the derivative of  $f$  yields:

$$\frac{d}{d\rho} f(\rho) = \frac{\rho^2 {}_2F_1\left(\frac{3}{2}, \frac{3}{2}; \frac{k}{2} + 2; \rho^2\right)}{2\left(\frac{k}{2} + 1\right)} + {}_2F_1\left(\frac{1}{2}, \frac{1}{2}; \frac{k}{2} + 1; \rho^2\right). \quad (54)$$

It is clear the derivative is non-negative where defined, from the definition of  ${}_2F_1$ , Equation (53), since it is a convergent sum of non-negative numbers as in the first property. For  $k \geq 3$  the derivative is defined at all  $\rho \in [-1, 1]$ , whereas for  $k = 1, 2$  the derivative is defined for all  $\rho \in (-1, 1)$ . Hence, we will consider an interval  $\rho \in [-1 + \delta, 1 - \delta]$  where  $\delta = \epsilon$  for  $k = 1, 2$  and  $\delta = 0$  for  $k \geq 3$ . The derivative is increasing in  $z$  for fixed  $a, b$  and  $c$  by Equation (53) so we may upper bound the derivative in the interval  $\rho \in [-1 + \delta, 1 - \delta]$  as:

$$\left| \frac{d}{d\rho} f(\rho) \right| \leq \left| \frac{d}{d\rho} f(\rho) \right|_{\rho=1-\delta} = \frac{(1-\delta)^2 {}_2F_1\left(\frac{3}{2}, \frac{3}{2}; \frac{k}{2} + 2; (1-\delta)^2\right)}{2\left(\frac{k}{2} + 1\right)} + {}_2F_1\left(\frac{1}{2}, \frac{1}{2}; \frac{k}{2} + 1; (1-\delta)^2\right),$$

establishing the third and fourth property.

For the last property, we compute the second derivative of  $f$  using Equation (54):

$$\frac{d^2}{d\rho^2} f(\rho) = \frac{9\rho^3 {}_2F_1\left(\frac{5}{2}, \frac{5}{2}; \frac{k}{2} + 3; \rho^2\right)}{4\left(\frac{k}{2} + 1\right)\left(\frac{k}{2} + 2\right)} + \frac{(\rho^2 + 2\rho) {}_2F_1\left(\frac{3}{2}, \frac{3}{2}; \frac{k}{2} + 2; \rho^2\right)}{2\left(\frac{k}{2} + 1\right)} + {}_2F_1\left(\frac{1}{2}, \frac{1}{2}; \frac{k}{2} + 1; \rho^2\right),$$

which is non-negative for  $\rho \in [0, 1)$  by the definition Equation (53). Note that  $\frac{d^2}{d\rho^2} f(\rho)$  fails to be defined at  $\rho = 1$  in general since  ${}_2F_1[a, b, c; 1]$  fails to be absolutely convergent when  $a + b > c$  and  $5/2 + 5/2$  may be larger than  $k/2 + 3$  depending on  $k$ . However, convexity in  $[0, 1)$  and continuity in  $[0, 1]$  of the function itself imply convexity in  $[0, 1]$ .  $\square$

## C Rank-constrained MAX-CUT

We now show how our results extend to the rank-constrained MAX-CUT problem. Recall from Definition 2.10 that rank- $k$  MAX-CUT is the problem of computing the value

$$\text{MAX-CUT}_k(G) = \max_{f: V \rightarrow S^{k-1}} \mathbf{E}_{(\mathbf{u}, \mathbf{v}) \sim E} \left[ \frac{1}{2} - \frac{1}{2} \langle f(\mathbf{u}), f(\mathbf{v}) \rangle \right].$$

This was introduced in the work of Briët, Oliveira, and Vallentin [BdOFV10, Section 6] as the Laplacian special case of a more general problem known as the *rank-constrained Grothendieck problem*. Their work lists numerous applications of the general rank-constrained Grothendieck problem, though to our knowledge there are no applications of the Laplacian special case (i.e. the rank-constrained MAX-CUT problem) aside from the  $k = 3$  case, which corresponds to the product state value of QUANTUM MAX-CUT, as we have seen.

The BOV algorithm that we have already seen for rank-3 MAX-CUT (equivalently, for the product state value of QUANTUM MAX-CUT) is actually the  $k = 3$  special case of an algorithm

for rank- $k$  MAX-CUT for general  $k$ , which we will also refer to as the “BOV algorithm” in this section. For general  $k$ , the BOV algorithm first solves the standard MAX-CUT SDP to produce a vector solution  $f_{\text{SDP}} : V \rightarrow S^{n-1}$ , which it then rounds into a random function  $\mathbf{f} : V \rightarrow S^{k-1}$  using projection rounding. To compute the algorithm’s approximation ratio, they go edge-by-edge: for each edge  $(u, v) \in E$ , if we set  $\rho_{u,v} = \langle f_{\text{SDP}}(u), f_{\text{SDP}}(v) \rangle$ , then the  $f_{\text{SDP}}$ ’s value for that edge is  $\frac{1}{2} - \frac{1}{2}\rho_{u,v}$ , whereas the expectation of  $\mathbf{f}$ ’s value for that edge is  $\frac{1}{2} - \frac{1}{2}F^*(k, \rho)$  (see [Theorem 7.15](#) for a definition of  $F^*(k, \rho)$ ). This motivates studying the following quantity.

**Definition C.1** (Approximation ratio for rank- $k$  MAX-CUT). Let  $k \geq 1$ . The constant  $\alpha_{k\text{MC}}$  is defined as the solution to the minimization problem

$$\alpha_{k\text{MC}} = \min_{-1 \leq \rho \leq 1} \frac{\frac{1}{2} - \frac{1}{2}F^*(k, \rho)}{\frac{1}{2} - \frac{1}{2}\rho},$$

and the constant  $\rho_{k\text{MC}}$  is defined as the minimizing value of  $\rho$ .

For  $k = 1$ ,  $\alpha_{1\text{MC}} = 0.8785\dots$  and  $\rho_{1\text{MC}} = -0.689\dots$ , corresponding to the Goemans-Williamson algorithm. For  $k = 3$ ,  $\alpha_{3\text{MC}} = 0.9563\dots$  and  $\rho_{3\text{MC}} = -0.584\dots$ , corresponding to the rank-3 BOV algorithm. [\[BdOFV10\]](#) also compute the  $k = 2$  values numerically and find  $\alpha_{2\text{MC}} = 0.9349\dots$  and  $\rho_{2\text{MC}} = -0.617\dots$ . Having defined these quantities, the expected value of  $\mathbf{f}$  is

$$\mathbf{E}_{\mathbf{f}} \mathbf{E}_{(u,v) \sim E} \left[ \frac{1}{2} - \frac{1}{2} \langle \mathbf{f}(u), \mathbf{f}(v) \rangle \right] = \mathbf{E}_{(u,v) \sim E} \left[ \frac{1}{2} - \frac{1}{2} F^*(k, \rho_{u,v}) \right] \geq \alpha_{k\text{MC}} \cdot \mathbf{E}_{(u,v) \sim E} \left[ \frac{1}{2} - \frac{1}{2} \rho_{u,v} \right] = \text{SDP}_{\text{MC}}(G),$$

and so the BOV algorithm has approximation ratio at least  $\alpha_{k\text{MC}}$ . This gives the following theorem.

**Theorem C.2** (Performance of the BOV algorithm for rank- $k$  MAX-CUT [\[BdOFV10\]](#)). *The BOV algorithm for rank- $k$  MAX-CUT achieves approximation ratio  $\alpha_{k\text{MC}}$ .*

Our results on the product state value of QUANTUM MAX-CUT imply that the BOV algorithm is optimal for rank-3 MAX-CUT. In fact, our proofs extend in a straightforward manner to show that the BOV algorithm is optimal for all constant values of  $k$ . One slight technicality is that our proofs require the worst-case  $\rho$  to be negative, as this is the only regime for which our vector-valued Borell’s inequality applies. The following proposition establishes this for rank- $k$  MAX-CUT.

**Proposition C.3** (Negative  $\rho$  is the worst case). *For all  $k \geq 1$ ,  $-1 \leq \rho_{k\text{MC}} \leq 0$ .*

*Proof.* The proposition follows from two claims about  $F^*(k, \rho)$ : (i) that  $F^*(k, \rho)$  always has the same sign as  $\rho$ , and (ii) that  $|F^*(k, \rho)| \leq |\rho|$ . Together, these imply that  $\frac{1}{2} - \frac{1}{2}F^*(k, \rho) \geq \frac{1}{2} - \frac{1}{2}\rho$  whenever  $0 \leq \rho \leq 1$ , and so their ratio is always at least 1 for  $\rho$  in this range. But  $\alpha_{k\text{MC}}$  is an approximation ratio and so is always between 0 and 1, and thus the minimizing value of  $\rho$  must be in the interval  $[-1, 0]$ .

Now we prove the claims. We recall from [Theorem 7.15](#) that

$$F^*(k, \rho) = \frac{2}{k} \left( \frac{\Gamma((k+1)/2)}{\Gamma(k/2)} \right)^2 \rho {}_2F_1[1/2, 1/2, k/2 + 1, \rho^2].$$

Property (i) follows because, by [Lemma B.6](#),  $F^*(k, \rho) \geq 0$  for  $\rho \in [0, 1]$ , and  $F^*$  is an odd function in  $\rho$ . [Lemma B.6](#) also establishes that  $F^*$  is convex for  $\rho \in [0, 1]$ . One may directly evaluate the expression for  $F^*(k, \rho)$  above to see that  $F^*(k, 0) = 0$  and  $F^*(k, 1) = 1$ . Thus  $F^*(k, \rho) \leq \rho$  for  $\rho \in [0, 1]$ ; because  $F^*$  is odd in  $\rho$ ,  $F^*(k, \rho) \geq \rho$  for  $\rho \in [-1, 0]$ , establishing (ii).  $\square$

Having established this, it is straightforward to extend our proofs to the case of rank- $k$  MAX-CUT, and we omit the details. (One very minor difference in the  $k = 1, 2$  case for the algorithmic gap is that the Lipschitz guarantee from [Lemma B.6](#) only holds when bounded away from  $-1$  and  $1$ . However, inspecting the proof of the algorithmic gap shows that this suffices.) Our results for rank- $k$  MAX-CUT are stated as follows.

**Theorem C.4** (Hardness for rank- $k$  MAX-CUT). *Let  $k \geq 1$  be fixed. Then the following three statements hold.*

1. *The MAX-CUT semidefinite program  $\text{SDP}_{\text{MC}}(G)$ , when viewed as a relaxation of  $\text{MAX-CUT}_k(G)$ , has integrality gap  $\alpha_{k\text{MC}}$ .*
2. *The BOV algorithm for rank- $k$  MAX-CUT has algorithmic gap  $\alpha_{k\text{MC}}$ .*
3. *Assuming the UGC, it is NP-hard to approximate  $\text{MAX-CUT}_k(G)$  to within a factor of  $\alpha_{k\text{MC}} + \epsilon$ , for all  $\epsilon > 0$ .*

## References

- [AA88] Daniel Arovas and Assa Auerbach. Functional integral theories of low-dimensional quantum Heisenberg models. *Physical Review B*, 38(1):316, 1988. [9](#)
- [AAV13] Dorit Aharonov, Itai Arad, and Thomas Vidick. Guest column: the quantum PCP conjecture. *ACM SIGACT News*, 44(2):47–79, 2013. [1](#)
- [ABH<sup>+</sup>05] Sanjeev Arora, Eli Berger, Elad Hazan, Guy Kindler, and Muli Safra. On non-approximability for quadratic programs. In *Proceedings of the 46th Annual IEEE Symposium on Foundations of Computer Science*, pages 206–215, 2005. [B.1](#)
- [AGKS21] Anurag Anshu, David Gosset, Karen Morenz Korol, and Mehdi Soleimanifar. Improved approximation algorithms for bounded-degree local Hamiltonians. *arXiv preprint arXiv:2105.01193*, 2021. [1](#), [5](#)
- [AGM20] Anurag Anshu, David Gosset, and Karen Morenz. Beyond product state approximations for a quantum analogue of Max Cut. In *Proceedings of the 15th Conference on the Theory of Quantum Computation, Communication and Cryptography*, pages 7:1–7:15, 2020. [1](#), [1](#), [1](#), [1](#), [2.6](#), [2.6](#), [3](#)
- [ALM<sup>+</sup>98] Sanjeev Arora, Carsten Lund, Rajeev Motwani, Madhu Sudan, and Mario Szegedy. Proof verification and the hardness of approximation problems. *Journal of the ACM*, 45(3):501–555, 1998. [1](#)
- [AS72] Milton Abramowitz and Irene Stegun, editors. *Handbook of mathematical functions: with formulas, graphs, and mathematical tables*, volume 55 of *Applied mathematics series*. National Bureau of Standards, 1972. [4.1](#), [1](#), [2](#), [3](#), [B.3](#)
- [AS98] Sanjeev Arora and Shmuel Safra. Probabilistic checking of proofs: a new characterization of NP. *Journal of the ACM*, 45(1):70–122, 1998. [1](#)
- [BBT09] Nikhil Bansal, Sergey Bravyi, and Barbara Terhal. Classical approximation schemes for the ground-state energy of quantum and classical Ising spin Hamiltonians on planar graphs. *Quantum Information & Computation*, 9(7):701–720, 2009. [1](#)

- [BdOFV10] Jop Briët, Fernando Mário de Oliveira Filho, and Frank Vallentin. The positive semidefinite Grothendieck problem with rank constraint. In *Proceedings of the 37th International Colloquium on Automata, Languages and Programming*, pages 31–42, 2010. [2](#), [\(document\)](#), [1](#), [2](#), [2.2](#), [2.2](#), [2.4](#), [2.14](#), [2.4](#), [2.7](#), [7.15](#), [C](#), [C](#), [C.2](#)
- [Bet31] Hans Bethe. Zur theorie der metalle. *Zeitschrift für Physik*, 71(3-4):205–226, 1931. [9](#), [7.2](#)
- [BGKT19] Sergey Bravyi, David Gosset, Robert König, and Kristan Temme. Approximation algorithms for quantum many-body problems. *Journal of Mathematical Physics*, 60(3):032203, 2019. [1](#), [7.2](#)
- [BH16] Fernando Brandão and Aram Harrow. Product-state approximations to quantum ground states. *Communications in Mathematical Physics*, 342(1):47–80, 2016. [1](#), [1](#), [2.6](#), [2.7](#), [7.10](#), [7.11](#), [8](#)
- [BK10] Nikhil Bansal and Subhash Khot. Inapproximability of hypergraph vertex cover and applications to scheduling problems. In *Proceedings of the 37th International Colloquium on Automata, Languages and Programming*, pages 250–261, 2010. [11](#)
- [Blo30] Felix Bloch. Zur theorie des ferromagnetismus. *Zeitschrift für Physik*, 61(3-4):206–219, 1930. [9](#)
- [Bor85] Christer Borell. Geometric bounds on the Ornstein-Uhlenbeck velocity process. *Probability Theory and Related Fields*, 70(1):1–13, 1985. [1](#), [II](#)
- [BS16] Boaz Barak and David Steurer. Lecture notes from CS229r/6.S898: Proofs, beliefs, and algorithms through the lens of Sum-of-Squares. Found at <https://www.sumofsquares.org/public/index.html>, 2016. [A](#)
- [CM16] Toby Cubitt and Ashley Montanaro. Complexity classification of local Hamiltonian problems. *SIAM Journal on Computing*, 45(2):268–316, 2016. [1](#), [7.2](#)
- [CW04] Moses Charikar and Anthony Wirth. Maximizing quadratic programs: extending Grothendieck’s inequality. In *Proceedings of the 45th Annual IEEE Symposium on Foundations of Computer Science*, pages 54–60, 2004. [7.2](#)
- [DX13] Feng Dai and Yuan Xu. Spherical harmonics. In *Approximation theory and harmonic analysis on spheres and balls*, pages 1–27. Springer, 2013. [4.1](#), [4.1](#), [4.3](#), [4.1](#), [4](#)
- [Eld15] Ronen Eldan. A two-sided estimate for the Gaussian noise stability deficit. *Inventiones mathematicae*, 201(2):561–624, 2015. [II](#)
- [FBB79] J. N. Fields, Hendrik Blöte, and Jill Bonner. Renormalization group and other calculations for the one-dimensional spin-1/2 dimerized Heisenberg antiferromagnet. *Journal of Applied Physics*, 50(B3):1807–1809, 1979. [9](#)
- [FE12] Christopher Frye and Costas Efthimiou. *Spherical Harmonics in p Dimensions*. 2012. [4](#), [4.1](#)
- [Fra17] Fabio Franchini. *An introduction to integrable techniques for one-dimensional quantum systems*. Springer, 2017. [2](#)

- [FS02] Uriel Feige and Gideon Schechtman. On the optimality of the random hyperplane rounding technique for MAX CUT. *Random Structures & Algorithms*, 20(3):403–440, 2002. [1](#), [2.7](#), [2](#), [B.1](#)
- [GHLS15] Sevag Gharibian, Yichen Huang, Zeph Landau, and Seung Woo Shin. Quantum Hamiltonian complexity. *Foundations and Trends in Theoretical Computer Science*, 10(3):159–282, 2015. [1](#)
- [GK12] Sevag Gharibian and Julia Kempe. Approximation algorithms for QMA-complete problems. *SIAM Journal on Computing*, 41(4):1028–1050, 2012. [1](#)
- [GP19] Sevag Gharibian and Ojas Parekh. Almost optimal classical approximation algorithms for a quantum generalization of Max-Cut. In *Proceedings of the 22nd International Workshop on Approximation Algorithms for Combinatorial Optimization Problems*, pages 31:1–31:17, 2019. [1](#), [1](#), [1](#), [1](#), [2.2](#), [2.4](#), [2.15](#), [10](#), [7.2](#)
- [GW95] Michel Goemans and David Williamson. Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *Journal of the ACM*, 42(6):1115–1145, 1995. [1](#), [2.1](#), [2.4](#), [2.7](#), [7.2](#)
- [Hec17] Erich Hecke. Über orthogonal-invariante Integralgleichungen. *Mathematische Annalen*, 78(1):398–404, 1917. [4.3](#)
- [Hei28] Werner Heisenberg. Zur Theorie des Ferromagnetismus. *Zeitschrift für Physik*, 49(9-10):619–636, 1928. [1](#), [7.2](#)
- [HLP20] Sean Hallgren, Eunou Lee, and Ojas Parekh. An approximation algorithm for the max-2-local Hamiltonian problem. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2020)*. Schloss Dagstuhl-Leibniz-Zentrum für Informatik, 2020. [1](#)
- [HM17] Aram W Harrow and Ashley Montanaro. Extremal eigenvalues of local Hamiltonians. *Quantum*, 1:6, 2017. [1](#)
- [HO22] Matthew B Hastings and Ryan O’Donnell. Optimizing strongly interacting fermionic hamiltonians. In *Proceedings of the 54th Annual ACM SIGACT Symposium on Theory of Computing*, pages 776–789, 2022. [1](#)
- [HT20] Steven Heilman and Alex Tarter. Three candidate plurality is stablest for small correlations. Technical report, arXiv:2011.05583, 2020. [1](#), [II](#)
- [IM12] Marcus Isaksson and Elchanan Mossel. Maximally stable Gaussian partitions with discrete applications. *Israel Journal of Mathematics*, 189(1):347–396, 2012. [1](#), [10](#), [10.2](#)
- [Kar72] Richard Karp. Reducibility among combinatorial problems. In *Proceedings of the Symposium on the Complexity of Computer Computations*, pages 85–103, 1972. [2.1](#)
- [Kar99] Howard Karloff. How good is the Goemans–Williamson MAX CUT algorithm? *SIAM Journal on Computing*, 29(1):336–350, 1999. [9](#)
- [Kho02] Subhash Khot. On the power of unique 2-prover 1-round games. In *Proceedings of the 34th Annual ACM Symposium on Theory of Computing*, pages 767–775, 2002. [1](#), [11.2](#)

- [KKMO07] Subhash Khot, Guy Kindler, Elchanan Mossel, and Ryan O’Donnell. Optimal inapproximability results for Max-Cut and other 2-variable CSPs? *SIAM Journal on Computing*, 37(1):319–357, 2007. [1](#), [2.1](#), [2.6](#), [11](#)
- [KO09] Subhash Khot and Ryan O’Donnell. SDP gaps and UGC-hardness for Max-Cut-Gain. *Theory of Computing*, 5:83–117, 2009. [B.1](#)
- [KR08] Subhash Khot and Oded Regev. Vertex cover might be hard to approximate to within  $2 - \epsilon$ . *Journal of Computer and System Sciences*, 74(3):335–349, 2008. [11](#)
- [KSV02] Alexei Kitaev, Alexander Shen, and Mikhail Vyalyi. *Classical and quantum computation*. American Mathematical Society, 2002. [1](#)
- [LM62] Elliott Lieb and Daniel Mattis. Ordering energy levels of interacting spin systems. *Journal of Mathematical Physics*, 3(4):749–751, 1962. [7.2](#)
- [LRS15] James Lee, Prasad Raghavendra, and David Steurer. Lower bounds on the size of semidefinite programming relaxations. In *Proceedings of the 47th Annual ACM Symposium on Theory of Computing*, pages 567–576, 2015. [2.1](#)
- [Men13] J Ricardo G Mendonça. Exact eigenspectrum of the symmetric simple exclusion process on the complete, complete bipartite and related graphs. *Journal of Physics A: Mathematical and Theoretical*, 46(29):295001, 2013. [2](#)
- [MN15] Elchanan Mossel and Joe Neeman. Robust optimality of Gaussian noise stability. *Journal of the European Mathematical Society*, 17(2):433–482, 2015. [II](#)
- [MN18] Emanuel Milman and Joe Neeman. The Gaussian double-bubble conjecture. Technical report, arXiv:1801.09296, 2018. [II](#)
- [MOO10] Elchanan Mossel, Ryan O’Donnell, and Krzysztof Oleszkiewicz. Noise stability of functions with low influences: invariance and optimality. *Annals of Mathematics*, 171(1):295–341, 2010. [1](#), [10](#)
- [MR15] Matthew McGonagle and John Ross. The hyperplane is the only stable, smooth solution to the isoperimetric problem in Gaussian space. *Geometriae Dedicata*, 178(1):277–296, 2015. [II](#)
- [O’D08] Ryan O’Donnell. Lecture 16 from 18-854B: advanced approximation algorithms. Found at <https://www.cs.cmu.edu/~anupamg/adv-approx/lecture16.pdf>, 2008. [9](#)
- [O’D14] Ryan O’Donnell. *Analysis of Boolean functions*. Cambridge University Press, 2014. [4.1](#), [5](#), [7.7](#), [7.21](#), [10](#), [B.3](#)
- [OW08] Ryan O’Donnell and Yi Wu. An optimal SDP algorithm for Max-Cut, and equally optimal Long Code tests. In *Proceedings of the 40th Annual ACM Symposium on Theory of Computing*, pages 335–344, 2008. [1](#), [8](#), [8.1](#), [B.1](#)
- [PM17] Stephen Piddock and Ashley Montanaro. The complexity of antiferromagnetic interactions and 2D lattices. *Quantum Information & Computation*, 17(7-8):636–672, 2017. [1](#), [7.2](#)

- [PT21a] Ojas Parekh and Kevin Thompson. Application of the Level-2 Quantum Lasserre Hierarchy in Quantum Approximation Algorithms. In *48th International Colloquium on Automata, Languages, and Programming (ICALP 2021)*, volume 198 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 102:1–102:20, 2021. [1](#), [1](#), [2.6](#), [2.6](#), [3](#)
- [PT21b] Ojas Parekh and Kevin Thompson. Beating Random Assignment for Approximating Quantum 2-Local Hamiltonian Problems. In *29th Annual European Symposium on Algorithms (ESA 2021)*, volume 204 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 74:1–74:18, 2021. [1](#), [10](#), [5](#)
- [PT22] Ojas Parekh and Kevin Thompson. An optimal product-state approximation for 2-local quantum hamiltonians with positive terms. Technical report, arXiv: 2206.08342, 2022. [1](#), [10](#)
- [Rag08] Prasad Raghavendra. Optimal algorithms and inapproximability results for every CSP? In *Proceedings of the 40th Annual ACM Symposium on Theory of Computing*, 2008. [1](#), [2.6](#)
- [Rag09] Prasad Raghavendra. *Approximating NP-hard problems: efficient algorithms and their limits*. PhD thesis, University of Washington, 2009. [1](#)
- [Sze39] Gabor Szegő. *Orthogonal polynomials*, volume 23. American Mathematical Soc., 1939. [2](#)
- [TSSW00] Luca Trevisan, Gregory Sorkin, Madhu Sudan, and David Williamson. Gadgets, approximation, and linear programming. *SIAM Journal on Computing*, 29(6):2074–2097, 2000. [2.1](#)
- [VB96] Lieven Vandenbergh and Stephen Boyd. Semidefinite programming. *SIAM review*, 38(1):49–95, 1996. [1](#)