



UvA-DARE (Digital Academic Repository)

Grothendieck inequalities, nonlocal games and optimization

Briët, J.

Publication date
2011

[Link to publication](#)

Citation for published version (APA):

Briët, J. (2011). *Grothendieck inequalities, nonlocal games and optimization*. [Thesis, fully internal, Universiteit van Amsterdam]. Institute for Logic, Language and Computation.

General rights

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, P.O. Box 19185, 1000 GD Amsterdam, The Netherlands. You will be contacted as soon as possible.

Chapter 1

Nonlocal Games and Optimization

1.1 Introduction

Nonlocal games. To gain a better understanding of the physical world, physicists use mathematical frameworks to model it. Within these frameworks it is often possible to describe everything from interactions of a subatomic particles to the orbits of planets flying around the sun. Such frameworks can be used to predict what can or cannot happen in certain real-world situations and the quality of a framework can be measured by how well its their predictions match what is actually observed. Two of the most important frameworks are Classical Mechanics and Quantum Mechanics, the latter being a refinement of the former. Einstein's General Relativity is another celebrated framework, but it does not play a role in this thesis. A large part of this thesis is devoted to studying within the frameworks of both Classical and Quantum Mechanics an abstraction of a physical experiment called a *nonlocal game*, introduced first by Cleve, Høyer, Toner and Watrous [CHTW04].¹ The main reason for considering these games is that they provide an excellent way to study the most important feature unique to Quantum Mechanics: *entanglement*. A nonlocal game involves two or more players who are not allowed to communicate with each other, but do interact with an extra party usually referred to as the referee. At the start of the game the referee asks each of the players a question, upon which they each reply to him with some answer. Then, the referee decides if the players win or lose based only on the questions he asked and the answers

¹The organization of the bibliography in the back of this thesis follows alphabetical order of the abbreviations used for references in the text.

he received. The players know in advance what set of answers would cause them to win, which of course is their objective. The catch is that they only know the question that was aimed directly at them and not any of the other players' questions, so they may not have enough information to know what to answer in order to win. The players thus don't play against each other, but rather have to try to coordinate their strategies to win. The way we study nonlocal games in the frameworks of Classical and Quantum Mechanics is by analyzing the winning probabilities for optimal strategies. Probabilities come into play here because we assume that the referee randomly picks the questions and because the players' strategies may involve some random processes. It turns out that the best course of action for players who live in a world described by Classical Mechanics is the simplest kind imaginable: decide before the game begins what to answer to each question and stick with that strategy throughout the game. In a Quantum Mechanical world, more sophisticated strategies sometimes give better results. Each player can base their answer on the outcome of an experiment done on some private physical system. Such an experiment may be, for example, measuring the orientation of the intrinsic magnetic field (the spin) of an electron. Such strategies typically give rise to some randomness in the players' answers, meaning that what a player answers to a particular question is not determined in advance. But this is not what separates quantum strategies from classical strategies. The key feature of quantum strategies is that they can cause the players to produce answers that are *correlated* in ways that are impossible in a classical world, as was shown for the first time by Bell [Bel64] in a slightly different language. Physical systems that allow players to obtain such correlations are said to be *entangled*. The fact that Quantum Mechanics predicts such a phenomenon was used by Einstein, Podolski and Rosen [EPR35] to argue that this framework must be incomplete, because according to them entanglement could not be part of a reasonable description of Nature. Surprisingly, experiments done by Aspect et al. [AGR81, ADR82, AGR82] gave convincing evidence that the world we live does in fact allow for this!

Optimization. And now for something completely different. An important type of problem in computer science is that of *optimization* under constraints. One example of such a problem is finding an optimal strategy for a nonlocal game, subject to the constraint that the strategy obeys the rules of Classical Mechanics (classical strategies). For a typical nonlocal game, finding an optimal classical strategy may involve searching over a huge number of possibilities.

For example, if in a two-player nonlocal game the referee can choose from n different questions and the players can choose from two possible answers per question, then there are 2^{2n} possible (deterministic) classical strategies. Another example of such a problem that we will encounter in this thesis originates from (classical) statistical physics. Here, the problem is to optimize spatial configurations of interacting particles so as to minimize the energy of the total system. A spin can point in two possible directions, so in an array of n particles there are 2^n possible configurations. Again a huge number of possibilities to search over. Problems like the above two likely can't be solved exactly by any computer in a reasonable amount of time, where time is measured by the number of elementary steps a computer makes and where by "reasonable time" we mean a polynomial number of steps in the size of the problem.² The next-best thing to exactly solving an optimization problem in polynomial time is to *approximate* it. In this case we are willing to settle for any solution (e.g., a strategy for a nonlocal game or a configuration of spins) that is near-optimal, but can be found in a reasonable amount of time. A computer algorithm that finds such a solution in polynomial time is referred to as a polynomial-time approximation algorithm. The second major theme in this thesis deals with analyzing new approximation algorithms for a general type of optimization problem that will allow us, for example, to approximate the kind of energy-minimization problem mentioned above.

Grothendieck Inequalities. Nonlocal games and optimization may at first sight seem to be quite unrelated. However, it turns out that the problems discussed above can be treated in a very similar fashion, using mathematical tools we call *Grothendieck Inequalities*. This name derives from the fact that these tools have their origin in a celebrated paper of Grothendieck [Gro53]. Grothendieck Inequalities are the fibers pulling the other topics in this thesis together.

1.2 Quantum information theory

In this section, we give some basic mathematical background information on the aspects of quantum information theory relevant to this thesis. More infor-

²In more technical terms, if $P \neq NP$ then there exists no polynomial-time algorithm for these problems. This follows from a translation of specific instances these problems to one of Karp's [Kar72] NP-complete problems. In fact, Håstad [Hås99] showed that the situation regarding these problems is even gloomier. The details of his result will be discussed later.

mation can be found in Appendix A, the book of Nielsen and Chuang [NC00] and the excellent lectures notes of Watrous [Wat08].

1.2.1 States and quantum systems

A *state* is a complex positive semidefinite matrix ρ that satisfies $\text{Tr}(\rho) = 1$. Any n -by- n state ρ can be decomposed as

$$\rho = \sum_{i=1}^n \lambda_i |\psi_i\rangle\langle\psi_i|,$$

where $\lambda_1, \dots, \lambda_n \geq 0$ are its eigenvalues and $|\psi_1\rangle, \dots, |\psi_n\rangle$ are corresponding eigenvectors, which follows from the Spectral Theorem and positive semidefiniteness. A state ρ is *pure* if it has rank 1, that is, if $\rho = |\psi\rangle\langle\psi|$ for some complex unit vector $|\psi\rangle$. The trace of a positive semidefinite matrix equals the sum of its eigenvalues. Hence, a state is a convex combination of pure states. A state with rank greater than 1 is sometimes referred to as a *mixed state*. It is common to refer to a complex unit vector $|\psi\rangle$ as a state. What is implicitly referred to in this case is the pure state $\rho = |\psi\rangle\langle\psi|$. We will follow this custom when we are working in the context of quantum information theory.

Although a state can be treated as a purely mathematical object, it should be thought of as describing the configuration of some *quantum system*, which is an abstract physical object, or collection of objects, on which one can perform experiments. Associated with a quantum system X is a positive integer n and a copy of the vector space \mathbb{C}^n . The possible configurations of X are given by the states in $\mathbb{C}^{n \times n}$. The reason why we associate \mathbb{C}^n with a quantum system instead of $\mathbb{C}^{n \times n}$ is that we will be working mostly with pure states. The integer n is referred to as the *dimension*, or *Hilbert space dimension* of X . A quantum system X is said to be *in state* ρ .

1.2.2 Measurements and observables

Let n be a positive integer and \mathcal{A} be a finite set. A *measurement* on an n -dimensional quantum system with *outcomes* in \mathcal{A} is defined by a set of positive semidefinite matrices $\{F^a\}_{a \in \mathcal{A}} \subseteq \mathbb{C}^{n \times n}$ that satisfy

$$\sum_{a \in \mathcal{A}} F^a = I.$$

If the matrices F^a also satisfy $F^a F^b = \delta_{ab} F^a$ for every $a, b \in \mathcal{A}$, then they define a *projective measurement*.

A measurement represents an experiment that one can perform on a quantum system. A measurement $\{F^a\}_{a \in \mathcal{A}} \subseteq \mathbb{C}^{n \times n}$ performed on an n -dimensional quantum system in state $\rho \in \mathbb{C}^{n \times n}$ yields a random variable χ over the set \mathcal{A} whose probability distribution is given by

$$\Pr[\chi = a] = \text{Tr}(\rho F^a).$$

The random variable χ is referred to as the *measurement outcome*.

If the set \mathcal{A} consists of real numbers, then the expected value of the random variable resulting from a projective measurement $\{F^a\}_{a \in \mathcal{A}}$ is given by

$$\mathbb{E}[\chi] = \sum_{a \in \mathcal{A}} a \text{Tr}(\rho F^a) = \text{Tr}\left(\rho \left(\sum_{a \in \mathcal{A}} a F^a\right)\right). \quad (1.1)$$

The matrix $\sum_{a \in \mathcal{A}} a F^a$ appearing on the right-hand side of Eq. (1.1) is then called the *observable* associated to the projective measurement $\{F^a\}_{a \in \mathcal{A}}$.

We will mostly work with observables associated to projective measurements with only two outcomes. A $\{-1, 1\}$ -valued *observable* is an observable corresponding to a projective measurement with outcomes in the set $\{-1, 1\}$. We denote the set of $\{-1, 1\}$ -valued observables in $\mathbb{C}^{n \times n}$ by $\mathcal{O}(\mathbb{C}^n)$.

We note the following useful fact about $\{-1, 1\}$ -valued observables, which we use again later on. It follows from the definition that such an observable can be written as the difference $F^+ - F^-$ of two orthogonal projectors. Squaring such an observable thus gives

$$(F^+ - F^-)^2 = F^+ + F^- = I.$$

A $\{-1, 1\}$ -valued observable is therefore both Hermitian and unitary. Since any matrix that is Hermitian and unitary has its eigenvalues in $\{-1, 1\}$, the converse is also true.

1.2.3 Entangled states and local measurements

A quantum system X may consist of subsystems X_1, \dots, X_N . In this case, we associate with each subsystem X_i a copy of the vector space \mathbb{C}^{n_i} , and we associate with X the vector space $\mathbb{C}^{n_1} \otimes \dots \otimes \mathbb{C}^{n_N}$. So, if X is in state ρ then ρ is a matrix of size $n_1 \cdots n_N$.

The subsystems may be distributed among N parties, who may be located at different places anywhere in the universe. If the overall quantum system X is in state ρ , then we say that the parties *share* the state ρ .

If the first party performs measurement $\{F^{a_1}\}_{a_1 \in \mathcal{A}_1} \subseteq \mathbb{C}^{n_1 \times n_1}$ on her subsystem X_1 , while the second party performs measurement $\{F^{a_2}\}_{a_2 \in \mathcal{A}_2} \subseteq \mathbb{C}^{n_2 \times n_2}$ on his subsystem X_2 , etc., then the joint probability distribution of the N measurement outcomes $\chi_1, \chi_2, \dots, \chi_N$ is, by definition, given by

$$\Pr[\chi_1 = a_1, \chi_2 = a_2, \dots, \chi_N = a_N] = \text{Tr}(\rho F_1^{a_1} \otimes F_2^{a_2} \otimes \dots \otimes F_N^{a_N}).$$

A pure state $|\psi\rangle \in \mathbb{C}^{n_1} \otimes \dots \otimes \mathbb{C}^{n_N}$ is a *product state* if it is of the form

$$|\psi\rangle = |\psi_1\rangle |\psi_2\rangle \dots |\psi_N\rangle.$$

(Tensor product symbols are usually omitted when using Dirac notation.) If $|\psi\rangle$ is not a product state then it is said to be *entangled*. If a mixed state is a convex combination of pure product states then it is *separable*. The most famous entangled state is the so-called EPR pair

$$|\text{EPR}\rangle = \frac{|0\rangle|0\rangle + |1\rangle|1\rangle}{\sqrt{2}} \in \mathbb{C}^2 \otimes \mathbb{C}^2,$$

named after Einstein, Podolski and Rosen. This is a pure state of a pair of two-dimensional quantum systems (usually referred to as *qubits*).

The most important difference between pure product states and pure entangled states is that the former type always gives rise to product distributions on local measurement outcomes, while this may not be the case for the latter type of states. In other words, product states give uncorrelated measurement outcomes, but entangled states can give correlated measurement outcomes.

Suppose that two parties, call them Alice and Bob, share a bi-partite product state $|\psi\rangle = |\psi_A\rangle |\psi_B\rangle$ and perform measurements $\{F^a\}_{a \in \mathcal{A}}$ and $\{G^b\}_{b \in \mathcal{B}}$ on their respective quantum systems. Then, the probability that Alice's measurement outcome χ_A is a and Bob's measurement outcome χ_B is b , equals

$$\begin{aligned} \text{Tr}(|\psi\rangle\langle\psi| F^a \otimes G^b) &= \langle\psi| F^a \otimes G^b |\psi\rangle \\ &= \langle\psi_A| \langle\psi_B| F^a \otimes G^b |\psi_A\rangle |\psi_B\rangle \\ &= \langle\psi_A| F^a |\psi_A\rangle \langle\psi_B| G^b |\psi_B\rangle. \end{aligned} \quad (1.2)$$

Since $\langle\psi_A| F^a |\psi_A\rangle$ is the probability of Alice obtaining a and $\langle\psi_B| G^b |\psi_B\rangle$ is the probability of Bob obtaining b , it follows that the distribution defined by Eq. (1.2) is a product distribution and in particular, that the measurement outcomes are uncorrelated.

Below, we give some examples in which parties produce correlated measurement outcomes using entangled states.

1.3 Nonlocal games

A two-player nonlocal game is defined by four finite sets $\mathcal{A}, \mathcal{B}, \mathcal{S}$ and \mathcal{T} , a joint probability distribution $\pi : \mathcal{S} \times \mathcal{T} \rightarrow [0, 1]$ and a map $V : \mathcal{A} \times \mathcal{B} \times \mathcal{S} \times \mathcal{T} \rightarrow \{0, 1\}$. The map V is usually referred to as the *predicate*. As the underlying sets are implicit in the probability distribution π and the predicate V , a nonlocal game can be uniquely defined by π and V .

A nonlocal game $\mathcal{G} = (\pi, V)$ involves three parties: A person called the *referee* and two players, usually called Alice and Bob. The probability distribution and predicate are known to the three parties in advance. Before the game begins, Alice and Bob may come together to decide on a strategy to play the game. But after the game has begun, they are not allowed to communicate with each other anymore.

At the start of the game, the referee picks a pair $(s, t) \in \mathcal{S} \times \mathcal{T}$ according to the probability distribution π , and sends s to Alice and t to Bob. Based on their strategies, the two players then answer the referee with $a \in \mathcal{A}$ and $b \in \mathcal{B}$, respectively. The players win the game if $V(a, b, s, t) = 1$, and lose otherwise. The players' objective is of course to maximize their chance of winning.

1.3.1 Classical strategies

A *deterministic classical strategy* refers to a strategy where the players simply use deterministic maps $a : \mathcal{S} \rightarrow \mathcal{A}$ and $b : \mathcal{T} \rightarrow \mathcal{B}$ to decide what to answer the referee after receiving their questions. In this case, their probability of winning a nonlocal game $\mathcal{G} = (\pi, V)$ is given by

$$\mathbb{E}_{(s,t) \sim \pi} [V(a(s), b(t), s, t)].$$

A slightly more sophisticated classical strategy involves shared and private randomness. Here, the players flip coins (some of which both can see and others that are private) to determine their answers. However, since such a course of action results in a probability distribution over deterministic classical strategies, it cannot increase the maximal chance of winning (see for example [CHTW04]).

1.3.2 Entangled strategies

We will contrast classical strategies with entangled strategies, in which Alice and Bob may share an entangled state on which they perform local measurements to determine their answers.

An entangled strategy consists of a positive integer n , a pair of n -dimensional quantum systems X_A and X_B in some entangled state ρ and measurements $\{F_s^a\}_{a \in \mathcal{A}}$ and $\{G_t^b\}_{b \in \mathcal{B}} \subseteq \mathbb{C}^{n \times n}$. The system X_A belongs to Alice and the system X_B to Bob. The players thus share the entangled state ρ .

Upon receiving question s , Alice performs measurement $\{F_s^a\}_{a \in \mathcal{A}}$ on X_A , and upon receiving question t , Bob performs measurement $\{G_t^b\}_{b \in \mathcal{B}}$ on X_B . The answers that Alice and Bob send back to the referee are their measurement outcomes. Since the probability that Alice answers a and Bob answers b is given by $\text{Tr}(\rho F_s^a \otimes G_t^b)$, their probability of winning the game equals

$$\mathbb{E}_{(s,t) \sim \pi} \left[\sum_{a \in \mathcal{A}} \sum_{b \in \mathcal{B}} \text{Tr}(\rho F_s^a \otimes G_t^b) V(a, b, s, t) \right].$$

It follows easily from linearity of the trace function and the fact that states are convex combinations of pure states, that pure entangled states suffice in order to maximize the winning probability with an entangled strategy. Additionally, in order to possibly have any advantage over classical strategies, the state ρ should be entangled, as separable states give rise to random uncorrelated answers, that is, randomized classical strategies.

1.4 Two-player XOR games

An XOR game is a nonlocal game in which the answer sets \mathcal{A} and \mathcal{B} are $\{0, 1\}$ and the predicate V depends only on the exclusive-OR (XOR) of the answers given by the players and the value of a boolean function $f : \mathcal{S} \times \mathcal{T} \rightarrow \{0, 1\}$. More precisely, the predicate is given by $V(a, b, s, t) = [a \oplus b = f(s, t)]$ where the square brackets denote the 0/1 truth value of the statement.

The truth table of the XOR function is as follows:

\oplus	0	1
0	0	1
1	1	0

An XOR game is thus defined by a pair $\mathcal{G} = (\pi, f)$ consisting of a probability distribution π and boolean function f .

The bias and the violation ratio. In an XOR game, the players (quantum or classical) can always win with probability $1/2$ by answering every question simply by flipping an unbiased coin. For the case of XOR games it therefore

makes more sense to look at the amount by which the maximum winning probability is bounded away from $1/2$.

We define the *classical bias* of an XOR game \mathcal{G} to be the difference between the probability of winning and the probability of losing for optimal classical strategy. We denote the classical bias by $\beta(\mathcal{G})$. We define the *entangled bias* similarly, and denote it by $\beta^*(\mathcal{G})$. The (classical or entangled) bias then equals *twice* the amount by which the maximal classical winning probability is greater than $1/2$. The reason to consider this definition is given in the next paragraph.

As a measure of the advantage entangled strategies give over classical strategies we define the *violation ratio* of \mathcal{G} to be the fraction $\beta^*(\mathcal{G})/\beta(\mathcal{G})$.

Signs and observables. XOR games are more easily analyzed using the $\{-1, 1\}$ -basis instead of the $\{0, 1\}$ -basis for boolean-valued objects. Let (π, f) be some XOR game. For any classical strategy $a : \mathcal{S} \rightarrow \{0, 1\}$ and $b : \mathcal{T} \rightarrow \{0, 1\}$, the bias is given by the probability under π that $a(s) \oplus b(t) = f(s, t)$ minus the probability under π that $a(s) \oplus b(t) \neq f(s, t)$. Concisely, the bias equals

$$\begin{aligned} \mathbb{E}_{(s,t) \sim \pi} \left[(-1)^{[a(s) \oplus b(t) = f(s,t)]} \right] &= \mathbb{E}_{(s,t) \sim \pi} \left[(-1)^{a(s) \oplus b(t) + f(s,t)} \right] \\ &= \mathbb{E}_{(s,t) \sim \pi} \left[(-1)^{a(s)} (-1)^{b(t)} (-1)^{f(s,t)} \right]. \end{aligned}$$

Hence, if we define sign matrix $\Sigma_{st} = (-1)^{f(s,t)}$ and functions $\chi(s) = (-1)^{a(s)}$ and $\psi(t) = (-1)^{b(t)}$, the bias becomes

$$\mathbb{E}_{(s,t) \sim \pi} \left[\chi(s) \psi(t) \Sigma_{st} \right].$$

Let us now consider an entangled strategy consisting of a shared (pure) entangled state $|\psi\rangle$ and projective measurements $\{F_s^0, F_s^1\}$ and $\{G_t^0, G_t^1\}$. The probability that Alice answers bit a upon receiving question s and Bob answers bit b upon receiving question t equals $\langle \psi | F_s^a \otimes F_t^b | \psi \rangle$. Hence, the expected value of the sign $(-1)^{a \oplus b}$ equals

$$\begin{aligned} \Pr[a = b] - \Pr[a \neq b] &= \\ \langle \psi | F_s^0 \otimes G_t^0 | \psi \rangle + \langle \psi | F_s^1 \otimes G_t^1 | \psi \rangle - \langle \psi | F_s^0 \otimes G_t^1 | \psi \rangle - \langle \psi | F_s^1 \otimes G_t^0 | \psi \rangle &= \\ \langle \psi | (F_s^0 - F_s^1) \otimes (G_t^0 - G_t^1) | \psi \rangle. \end{aligned}$$

Defining the $\{-1, 1\}$ -valued observables $F_s = F_s^0 - F_s^1$ and $G_t = G_t^0 - G_t^1$, we get that the bias based on this strategy equals

$$\mathbb{E}_{(s,t) \sim \pi} \left[\langle \psi | F_s \otimes G_t | \psi \rangle \Sigma_{st} \right].$$

We will often replace the boolean function f by the matrix Σ , and say that the pair (π, Σ) defines an XOR game. By the above calculations, the classical bias of such a game is given by

$$\max \left\{ \mathbb{E}_{(s,t) \sim \pi} \left[\Sigma_{st} \chi(s) \psi(t) \right] : \chi : \mathcal{S} \rightarrow \{-1, 1\}, \psi : \mathcal{T} \rightarrow \{-1, 1\} \right\}$$

and the entangled bias is given by

$$\sup_{n \in \mathbb{N}} \left\{ \mathbb{E}_{(s,t) \sim \pi} \left[\Sigma_{st} \langle \psi | F_s \otimes G_t | \psi \rangle \right] : |\psi\rangle \in \mathbb{C}^n \otimes \mathbb{C}^n, F_s, G_t \in \mathcal{O}(\mathbb{C}^n) \right\}.$$

The supremum is used in the entangled bias because the possibility exists that the maximal winning probability increases indefinitely with the dimension of the quantum systems.

This reformulation will prove to be a great convenience later on. The reason why we only considered projective measurements is that general measurements do not give an advantage over projective measurements, as shown by Cleve, Høyer, Toner and Watrous [CHTW04, Proposition 2].

1.4.1 The CHSH game

The CHSH game, named after Clauser, Horne, Shimony and Holt [CHSH69], is a two-player XOR with two possible questions per player, 0 and 1. The probability distribution π on $\{0, 1\} \times \{0, 1\}$ is the uniform distribution, so every pair of questions is asked with probability $1/4$. The predicate V evaluates to 1 if and only if $a \oplus b = s \wedge t$, where \wedge denotes the AND function (which is 1 if and only if $s = t = 1$). Classical players can win this game with probability no greater than $3/4$, which can be seen by observing that the system of equations

$$\begin{aligned} a_0 \oplus b_0 &= 0 \\ a_0 \oplus b_1 &= 0 \\ a_1 \oplus b_0 &= 0 \\ a_1 \oplus b_1 &= 1 \end{aligned}$$

is overdetermined and only three equations can be satisfied simultaneously.

By sharing an EPR pair, Alice and Bob can win the CHSH game with probability $\cos(\pi/8)^2 \approx 0.85$. An entangled strategy based on $\{-1, 1\}$ -valued observables that achieves this is as follows. Define the matrices $X = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ and $Y = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}$. These matrices satisfy $X^2 = Y^2 = I$, so they are observables, and they anti-commute, meaning that $XY + YX = 0$. Define Alice's observables

for questions 0 and 1 by $F_0 = X$ and $F_1 = Y$, respectively. Define Bob's observables for questions 0 and 1 to be $G_0 = (X - Y)/\sqrt{2}$ and $G_1 = (X + Y)/\sqrt{2}$, respectively. The matrices X and Y should be thought of as being given in the basis $|0\rangle, |1\rangle$ in which the EPR pair

$$|\text{EPR}\rangle = \frac{|0\rangle|0\rangle + |1\rangle|1\rangle}{\sqrt{2}}$$

is given. The following relations then follow easily:

$$\begin{aligned} \langle \text{EPR} | X \otimes X | \text{EPR} \rangle &= 1 & \langle \text{EPR} | Y \otimes Y | \text{EPR} \rangle &= -1 \\ \langle \text{EPR} | X \otimes Y | \text{EPR} \rangle &= 0 & \langle \text{EPR} | Y \otimes X | \text{EPR} \rangle &= 0. \end{aligned}$$

From these equations we get $\langle \text{EPR} | F_s \otimes G_t | \text{EPR} \rangle = (-1)^{s \wedge t} / \sqrt{2}$ for every $s, t \in \{0, 1\}$ and it follows that the bias based on the above entangled strategy equals

$$\frac{1}{4} \sum_{s,t=0}^1 (-1)^{s \wedge t} \langle \text{EPR} | F_s \otimes G_t | \text{EPR} \rangle = \frac{1}{\sqrt{2}},$$

making the winning probability $1/2 + 1/(2\sqrt{2}) = \cos(\pi/8)^2$.

1.5 Tsirelson's Theorem

Tsirelson's Theorem [Tsi87] gives an extremely useful characterization of entangled strategies in two-player XOR games. It forms the basis of many results in this thesis. Roughly speaking, the theorem gives a correspondence relation between entangled strategies consisting of a shared entangled state and $\{-1, 1\}$ -valued observables on the one hand, and pairs of sequences of real unit vectors on the other. The correspondence relation is given by the following theorem, which is commonly referred to as Tsirelson's Theorem. We will refer to the two parts of the correspondence as the "hard direction" and the "easy direction".

1.5.1. THEOREM (TSIRELSON). (*Hard direction*) *For all positive integers n, r and any real r -dimensional unit vectors $x_1, \dots, x_n, y_1, \dots, y_n$, there exists a positive integer d that depends on r only, a state $|\psi\rangle \in \mathbb{C}^d \otimes \mathbb{C}^d$ and $\{-1, 1\}$ -observables $F_1, \dots, F_n, G_1, \dots, G_n \in \mathcal{O}(\mathbb{C}^d)$, such that for every $i, j \in \{1, \dots, n\}$, we have*

$$\langle \psi | F_i \otimes G_j | \psi \rangle = x_i \cdot y_j.$$

Moreover, $d \leq 2^{\lceil r/2 \rceil}$.

(Easy direction) Conversely, for all positive integers n, d , state $|\psi\rangle \in \mathbb{C}^d \otimes \mathbb{C}^d$ and $\{-1, 1\}$ -observables $F_1, \dots, F_n, G_1, \dots, G_n \in \mathcal{O}(\mathbb{C}^d)$, there exist a positive integer r that depends on d only and real r -dimensional unit vectors $x_1, \dots, x_n, y_1, \dots, y_n$ such that for every $i, j \in \{1, \dots, n\}$, we have

$$x_i \cdot y_j = \langle \psi | F_i \otimes G_j | \psi \rangle.$$

Moreover, $r \leq 2d^2$.

PROOF OF THEOREM 1.5.1: We start by proving the hard direction. Let

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

(These matrices are called the *Pauli matrices*.) Note that each of them squares to the identity matrix I . This implies that they have eigenvalues in $\{-1, 1\}$. Additionally, note that the last three of them, X, Y and Z , pair-wise *anti-commute*, meaning that $XY + YX = XZ + ZX = YZ + ZY = 0$.

Define for each $\ell = 1, \dots, \lceil r/2 \rceil$, the d -by- d Clifford matrices,

$$\begin{aligned} S_{2\ell+1} &= Z^{\otimes(\ell-1)} \otimes X \otimes I^{\otimes(\lceil r/2 \rceil - \ell)}, \\ S_{2\ell} &= Z^{\otimes(\ell-1)} \otimes Y \otimes I^{\otimes(\lceil r/2 \rceil - \ell)}. \end{aligned}$$

From the properties satisfied by the Pauli matrices, the Clifford matrices satisfy that they square to the identity matrix (of size d -by- d) and pair-wise anti-commute. So, for every $k, \ell \in \{1, \dots, \lceil r/2 \rceil\}$, we have $S_k S_\ell + S_\ell S_k = 2\delta_{k\ell} I$. Additionally, for every $k \neq \ell$, we have $\text{Tr}(S_k S_\ell) = 0$.

Define $F_1, \dots, F_n, G_1, \dots, G_n \in \mathbb{C}^{d \times d}$ by

$$\begin{aligned} F_i &= \sum_{k=1}^r (x_i)_k S_k, \\ G_j &= \sum_{k=1}^r (y_j)_k S_k^T. \end{aligned}$$

1. CLAIM. *The matrices $F_1, \dots, F_n, G_1, \dots, G_n$ are $\{-1, 1\}$ -observables.*

PROOF: (Hard direction) It suffices to show that $F_i^2 = G_j^2 = I$ for each $i, j \in \{1, \dots, n\}$, as this implies that the matrices have eigenvalues in $\{-1, 1\}$. To this end, consider the expansion of F_i^2 ,

$$\sum_{k, \ell=1}^r (x_i)_k (x_i)_\ell S_k S_\ell = x \cdot x I + \sum_{k > \ell} (x_i)_k (x_i)_\ell (S_k S_\ell + S_\ell S_k).$$

From the anti-commutation relations satisfied by the Clifford matrices, it follows that the second sum on the right-hand side equals zero. What remains is the identity, as x is a unit vector.

Of course, the same argument works for G_j . This proves the claim. \blacklozenge

2. CLAIM. For every $i, j \in \{1, \dots, n\}$, we have $\text{Tr}(F_i G_j^T) / d = x_i \cdot y_j$.

PROOF: Fix $i, j \in \{1, \dots, n\}$. Similarly as in the proof of the previous claim, consider the expansion of the product $F_i G_j^T$,

$$\sum_{k, \ell=1}^r (x_i)_k (y_j)_\ell S_k S_\ell. \quad (1.3)$$

Since $\text{Tr}(S_k S_\ell) = d \delta_{k\ell}$, the only terms in (1.3) that contribute nontrivially to $\text{Tr}(F_i G_j^T)$, are those for which $k = \ell$. The sum of those terms is exactly $dx \cdot y$. \blacklozenge

We now consider the expansion of $\text{Tr}(F_i G_j^T) / d$. Let $\{|1\rangle, \dots, |d\rangle\} \subseteq \mathbb{C}^d$ be an orthonormal basis for \mathbb{C}^d . Let

$$|\psi\rangle = \frac{1}{\sqrt{d}} \sum_{s=1}^d |s\rangle \otimes |s\rangle,$$

be the maximally entangled state.

We have

$$\begin{aligned} \langle \psi | F_i \otimes G_j | \psi \rangle &= \frac{1}{d} \sum_{s, t=1}^d \langle s | \otimes \langle s | F_i \otimes G_j | t \rangle \otimes | t \rangle \\ &= \frac{1}{d} \sum_{s, t=1}^d \langle s | F_i | t \rangle \langle s | G_j | t \rangle \\ &= \frac{1}{d} \text{Tr}(F_i G_j^T). \end{aligned}$$

Combining this with the two claims then proves the hard direction.

(Easy direction) Note that since $|\psi\rangle$ has norm 1 and the observables F_i and G_j are unitary operators, $F_i \otimes I |\psi\rangle$ and $I \otimes G_j |\psi\rangle$ are unit vectors in \mathbb{C}^{d^2} . Additionally, note that since F_i and G_j are Hermitian, we have that the inner product

$$(\langle \psi | F_i \otimes I) \cdot (I \otimes G_j | \psi \rangle) = \langle \psi | F_i \otimes G_j | \psi \rangle,$$

is a real number. For $v \in \mathbb{C}^{d^2}$ we let $\Re(v)$ denote its real part and $\Im(v)$ its complex part, so that

$$\begin{aligned} F_i \otimes I |\psi\rangle &= \Re(F_i \otimes I |\psi\rangle) + i \Im(F_i \otimes I |\psi\rangle) \\ I \otimes G_j |\psi\rangle &= \Re(I \otimes G_j |\psi\rangle) + i \Im(I \otimes G_j |\psi\rangle). \end{aligned}$$

Define vectors $2d^2$ -dimensional unit vectors x_i, y_j by

$$\begin{aligned} x_i &= \Re(F_i \otimes I|\psi\rangle) \oplus \Im(F_i \otimes I|\psi\rangle) \\ y_j &= \Re(G_j \otimes I|\psi\rangle) \oplus \Im(-G_j \otimes I|\psi\rangle) \end{aligned}$$

Then, since $\langle\psi|F_i \otimes G_j|\psi\rangle$ is a real number, we have

$$\begin{aligned} x_i \cdot y_j &= \Re(\langle\psi|F_i \otimes G_j|\psi\rangle) - \Im(\langle\psi|F_i \otimes G_j|\psi\rangle) \\ &= \langle\psi|F_i \otimes G_j|\psi\rangle, \end{aligned}$$

as desired. □

1.6 Multiplayer XOR games

By a multiplayer XOR game we generally mean an XOR game involving more than two players. For convenience, we will only consider N -player XOR games in which the question sets are all the same finite set \mathcal{S} . Let π be a probability distribution on \mathcal{S}^N and $f : \mathcal{S}^N \rightarrow \{0, 1\}$ be a boolean function. In an N -player XOR game $\mathcal{G} = (\pi, f)$, the referee picks an N -tuple of questions (s_1, \dots, s_N) according to π and sends s_1 to the first player, s_2 to the second, and so on. The players answer with $a_1, \dots, a_N \in \{0, 1\}^N$, respectively and win the game if

$$a_1 \oplus \dots \oplus a_N = f(s_1, \dots, s_N).$$

The classical and entangled biases are given in terms of the map $\Sigma : \mathcal{S}^N \rightarrow \{-1, 1\}$ defined by $\Sigma[s_1, \dots, s_N] = (-1)^{f(s_1, \dots, s_N)}$. The map Σ will often be referred to as a *sign tensor* and if $N = 2$ it will be called a *sign matrix*. The classical bias of the game $\mathcal{G} = (\pi, \Sigma)$ is then given by

$$\beta(\mathcal{G}) = \max \left\{ \mathbb{E}_{(s_1, \dots, s_N) \sim \pi} \left[\Sigma[s_1, \dots, s_N] \chi_1(s_1) \cdots \chi_N(s_N) \right] \right\},$$

where the maximum is taken over maps $\chi_1, \dots, \chi_N : \mathcal{S} \rightarrow \{-1, 1\}$.

Then entangled bias is given by

$$\beta^*(\mathcal{G}) = \sup \left\{ \mathbb{E}_{(s_1, \dots, s_N) \sim \pi} \left[\Sigma[s_1, \dots, s_N] \langle\psi|F_1(s_1) \otimes \cdots \otimes F_N(s_N)|\psi\rangle \right] \right\},$$

where the supremum is over positive integers n , states $|\psi\rangle \in \mathbb{C}^n \otimes \cdots \otimes \mathbb{C}^n$ and observable-valued maps $F_1, \dots, F_N : \mathcal{S} \rightarrow \mathcal{O}(\mathbb{C}^n)$.

1.6.1 Mermin's Game

Mermin [Mer90] gave a sequence of XOR games, one for every number N of players, in which the violation ratio grows exponentially with N . Entangled players can play these games perfectly by sharing an N -qubit GHZ state

$$|\text{GHZ}\rangle = \frac{|0\rangle \cdots |0\rangle + |1\rangle \cdots |1\rangle}{\sqrt{2}},$$

named after its inventors Greenberger, Horne and Zeilinger [GHZ89]. Mermin's game is described as follows. The referee picks an N -bit string $x = x_1x_2 \dots x_N$ uniformly at random from all strings with even Hamming weight $|x|$ (i.e., the number of 1s appearing in x is even). He sends x_1 to the first player, x_2 to the second, etc. In order to win the game, the players must answer bits a_1, \dots, a_N (resp.) such that $a_1 \oplus \cdots \oplus a_N = |x|/2 \pmod{2}$.

1.6.1. PROPOSITION. *The classical bias of Mermin's game is at most $2^{-(N-1)/2}$ if N is odd and at most $2^{-(N-2)/2}$ if N is even.*

PROOF: Without loss of generality, we may assume that the players use a deterministic strategy in order to play the game. Let $a_k(0)$ and $a_k(1)$ denote the answers of the k^{th} player to questions 0 and 1, respectively.

A simple calculation shows that the players' bias is given by the formula

$$\frac{1}{2^{N-1}} \sum_{x \in \{0,1\}^N: |x| \text{ even}} (-1)^{|x|/2} (-1)^{a_1(x_1) + \cdots + a_N(x_N)} = \frac{1}{2^{N-1}} \Re \left[\prod_{k=1}^N \left((-1)^{a_k(0)} + i(-1)^{a_k(1)} \right) \right],$$

where \Re denotes the real part of a complex number. Note that each complex number $(-1)^{a_k(0)} + i(-1)^{a_k(1)}$ has modulus $\sqrt{2}$ and argument a multiple of $\pi/4$. If N is odd, then the product of these complex numbers makes a 45 degree angle with the real axis in the complex plane, making their real part equal to $\pm 2^{(N-1)/2}$. If N is even, then their product is either parallel to the imaginary axis or parallel to the real axis. Hence, the real part of their product is at most $2^{N/2}$. Dividing by the above factor 2^{N-1} gives the result. \square

1.6.2. PROPOSITION. *The entangled bias of Mermin's game is 1.*

PROOF: Let $X = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ and $Y = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}$. As these matrices satisfy $X^2 = Y^2 = I$, they are $\{-1, 1\}$ -valued observables. Moreover, they satisfy

$$\begin{aligned} X|0\rangle &= |1\rangle & X|1\rangle &= |0\rangle \\ Y|0\rangle &= i|1\rangle & Y|1\rangle &= -i|0\rangle. \end{aligned}$$

We claim that N entangled players can play the game perfectly by measuring their respective qubits of the N -qubit GHZ state

$$|\text{GHZ}\rangle = \frac{|0\rangle \cdots |0\rangle + |1\rangle \cdots |1\rangle}{\sqrt{2}},$$

using observable $F(0) = X$ on question 0 and $F(1) = Y$ on question 1.

To see this, notice that we have

$$\begin{aligned} X \otimes X \otimes X \otimes \cdots \otimes X |\psi\rangle &= -|\text{GHZ}\rangle \\ Y \otimes Y \otimes X \otimes \cdots \otimes X |\psi\rangle &= |\text{GHZ}\rangle. \end{aligned}$$

In general, if the number of Y 's that appear in the tensor products above is an odd multiple of 2, then $|\psi\rangle$ is an eigenvector with eigenvalue -1 , and if the number of Y 's is a multiple of 4, then $|\psi\rangle$ has eigenvalue $+1$. Hence, for $x \in \{0, 1\}^N$ with $|x|$ even, we have

$$\bigotimes_{k=1}^N F(x_k) |\text{GHZ}\rangle = (-1)^{|x|/2} |\text{GHZ}\rangle.$$

The result now follows from the fact that the players' bias based on this strategy equals

$$\frac{1}{2^{N-1}} \sum_{x \in \{0,1\}^N: |x| \text{ even}} (-1)^{|x|/2} \langle \text{GHZ} | \bigotimes_{k=1}^N F(x_k) | \text{GHZ} \rangle = 1,$$

which completes the proof. \square

1.6.2 Stabilizer states

The GHZ state, defined in the previous subsection, is a special case of a general class of states known as stabilizer states. An N -qubit *stabilizer state* $|\psi\rangle$ is the unique common eigenvector of the elements of an abelian subgroup $S \subseteq \{I, X, Y, Z\}^{\otimes N}$ of order 2^N , such that $M|\psi\rangle = |\psi\rangle$ for every $M \in S$. Here, I, X, Y, Z are the 2-by-2 Pauli matrices (see Section 1.5) and the group operation of S is regular matrix multiplication. By a *tripartite stabilizer state*, we mean a stabilizer state whose qubits are distributed among three parties. These states are discussed in the context of multiplayer XOR games in Chapter 6.

1.7 Semidefinite programs and relaxations

A generic *semidefinite program* (SDP) has the following form. Given positive integers k, n , real n -by- n matrices A, B_1, \dots, B_k and real numbers c_1, \dots, c_k

$$\begin{aligned} & \text{maximize} && \langle A, X \rangle \\ & \text{subject to} && X \in \mathcal{S}_n^+ \\ & && \langle B_i, X \rangle = c_i, \end{aligned}$$

for $i = 1, \dots, k$. Here $\langle C, X \rangle = \text{Tr}(C^T X)$ denotes the trace inner product of the matrices C and X and \mathcal{S}_n^+ denotes the set of real n -by- n positive semidefinite matrices.

We will use the following standard terminology and facts of semidefinite programs (see for example the books of Grötschel, Lovász and Schrijver [GLS93] and Boyd and Vandenberghe [BV04], or the survey of Laurent and Rendl [LR05]).

The quantity $\langle A, X \rangle$ above is referred to as the *objective value* of the SDP. The conditions $X \in \mathcal{S}_n^+$ and $\langle B_i, X \rangle = c_i$ imposed on the matrix X are the *constraints*. If a matrix X satisfies all the constraints of an SDP, then it is said to be a *feasible solution*, or simply *feasible* for short. If a matrix X is a feasible and it maximizes the objective value, then it is said to be an *optimal solution* for the SDP, or *optimal* for short. The value $\langle A, X \rangle$ for optimal solution X is the *optimum* of the SDP.

The most important fact about SDPs is that their optimum can be approximated to within arbitrary fixed precision in polynomial time, as testing whether a rational matrix is positive semidefinite can be done efficiently using for example Gaussian elimination.

1.7.1 Approximation algorithms

One of the most important uses of semidefinite programs is in approximation algorithms for combinatorial optimization problems that are unknown to be solvable exactly in polynomial time. The philosophy behind such algorithms is that it is often good enough to have a solution that is close to optimal. The advantage gained by relaxing exact optimality is that near-optimal solutions can sometimes be found much faster.

We distinguish semidefinite programs from approximation algorithms by requiring from the latter that they return a feasible solution for the optimization problem they approximate. A semidefinite program which serves as a

relaxation for an optimization problem can sometimes be turned into an approximation algorithm by adding a procedure which turns an optimal solution to the SDP (some positive semidefinite matrix) into a feasible one for the optimization problem.

If the optimum of an SDP is c times the optimum of some optimization problem OPT, then we say that the SDP has *approximation ratio* c for OPT. If the output of an approximation algorithm gives a value of δ times the optimum of an optimization problem, then we say that the approximation algorithm gives a δ -*approximation*. Here, c is typically greater than 1 and δ lies in $[0, 1]$.

Below we give two examples of applications of semidefinite programs for well-known combinatorial optimization problems: the maximum cut problem and the problem of computing the chromatic number of a graph.

1.7.2 MAX CUT

The *maximum cut problem* (MAX CUT) refers to the following combinatorial optimization problem. Given an undirected graph $G = (V, E)$ with finite vertex set V and edge set $E \subset V \times V$ (with no self-loops), find a bi-partitioning of V such that the number of edges crossing the partition is maximal. Such a bi-partitioning is also referred to as a *cut*, and the number of edges crossing it as the *size* of the cut.

The MAX CUT problem is one of Karp's 21 NP-complete problems [Kar72] (see also [GJ76]). It is therefore unlikely that a polynomial-time algorithm exists that solves it exactly in the worst case. To make matters worse, Håstad [Hås99] proved that even finding a cut of size $16/17 - \epsilon$ times the size of a maximum cut, for any constant $\epsilon > 0$, cannot be done in polynomial time unless $P=NP$.

Good upper bounds on the size of a maximum cut of a graph can be found using a semidefinite program and a matrix called the Laplacian. Given a graph $G = (V, E)$, its *Laplacian* $A : V \times V \rightarrow \mathbb{R}$ is defined by

$$A(u, v) = \begin{cases} \deg(u) & \text{if } v = u \\ -1 & \text{if } \{u, v\} \in E \\ 0 & \text{otherwise,} \end{cases}$$

where $\deg(u) = |\{v \in V : \{u, v\} \in E\}|$ denotes the *degree* of vertex u .

The semidefinite program is then given by:

$$\begin{aligned} & \text{maximize} && \frac{1}{4} \langle A, X \rangle \\ & \text{subject to} && X \in \mathcal{S}_V^+ \\ & && X(u, u) = 1 \text{ for every } u \in V, \end{aligned} \tag{1.4}$$

where \mathcal{S}_V^+ denotes the set of real positive semidefinite matrices whose rows and columns are indexed by the vertices of G . The fact that the optimum of this SDP upper bounds the size of a maximum cut can be shown as follows. Suppose that $S \subseteq V$ defines a cut $(S, V \setminus S)$ of maximal size. Define the function $\chi : V \rightarrow \{-1, 1\}$ by setting $\chi(u) = +1$ if $u \in S$ and $\chi(u) = -1$ otherwise. Then, the matrix $X(u, v) = \chi(u)\chi(v)$ is feasible for SDP (1.4) since it is positive semidefinite and has ones on the diagonal. For its objective value we compute

$$\begin{aligned} \langle A, X \rangle &= \sum_{u, v \in V} A(u, v) \chi(u) \chi(v) \\ &= \sum_{u \in V} \deg(u) - 2 \sum_{\{u, v\} \in E} \chi(u) \chi(v) \\ &= 2 \sum_{\{u, v\} \in E} (1 - \chi(u) \chi(v)). \end{aligned} \tag{1.5}$$

Each of the terms $1 - \chi(u)\chi(v)$ in the last sum equals 2 if the edge $\{u, v\}$ crosses the cut and 0 otherwise. Hence, the objective value of X is exactly the size of the maximum cut. Note that the optimum of SDP (1.4) may be higher.

In a celebrated paper, Goemans and Williamson [GW94] turned SDP (1.4) into a .878-approximation algorithm for MAX CUT, Algorithm 1.1 shown below. The description of the algorithm uses that for any $X \in \mathcal{S}_V^+$ satisfying $X(u, u) = 1$ there is a function $f : V \rightarrow S^{|V|-1}$ such that $X(u, v) = f(u) \cdot f(v)$ for every $u, v \in V$, where

$$S^{n-1} = \{x \in \mathbb{R}^n : x \cdot x = 1\}$$

denotes the real n -dimensional unit sphere (see for example Appendix A).

To analyze Algorithm 1.1 we define a function $\chi : V \rightarrow \{-1, 1\}$ by setting $\chi(u) = +1$ if u belongs to the set S returned by the algorithm and setting $\chi(u) = -1$ otherwise. Based on the vector z sampled in the algorithm we have

$$\chi(u) = \text{sign}(z \cdot f(u)).$$

Let $A : V \times V \rightarrow \mathbb{R}$ be the Laplacian matrix of the graph G given to the algorithm. By running the sequence of equations in Eq. (1.5) backwards we

Algorithm 1.1 (Goemans and Williamson) Takes as input a graph $G = (V, E)$ and returns a cut $(S, V \setminus S)$ for some $S \subseteq V$ in G .

- (1) Solve SDP (1.4), obtaining a function $f : V \rightarrow S^{|\mathbb{S}|-1}$.
 - (2) Sample a vector $z \in \mathbb{R}^{|\mathbb{S}|}$ such that the entries of z are independently distributed Gaussian random variables with mean 0 and variance 1.
 - (3) Put $u \in S$ if and only if $z \cdot f(u) \geq 0$.
-

get that on expectation over the vector z , the size of the cut returned by the algorithm is given by

$$\begin{aligned} \mathbb{E}_z \left[\frac{1}{2} \sum_{\{u,v\} \in E} (1 - \chi(u)\chi(v)) \right] &= \mathbb{E}_z \left[\frac{1}{4} \sum_{u,v \in V} A(u,v) \chi(u)\chi(v) \right] \\ &= \frac{1}{4} \sum_{u,v \in V} A(u,v) \mathbb{E}_z [\chi(u)\chi(v)], \end{aligned} \quad (1.6)$$

where we used linearity of expectation for the second identity.

The next step of the analysis uses a useful identity often referred to as Grothendieck's Identity, as it appeared first in [Gro53, Proposition 4, p. 63].

1.7.1. LEMMA (GROTHENDIECK'S IDENTITY). *Let x, y be real unit vectors and let z be a random Gaussian vector with independently distributed entries that have mean 0 and variance 1. Then, we have*

$$\mathbb{E}_z [\text{sign}(z \cdot x) \text{sign}(z \cdot y)] = \frac{2}{\pi} \arcsin(x \cdot y).$$

PROOF: We have $\text{sign}(z \cdot x) \text{sign}(z \cdot y) = +1$ if and only if the vectors x and y lie on the same side of the hyperplane orthogonal to the vector z . Now we project this n -dimensional situation to the plane spanned by x and y . Then the projected random hyperplane becomes a random line. This random line is distributed according to the uniform probability measure on the unit circle because z is normally distributed. We obtain the result by measuring regions on the unit circle and using the identity $\arcsin(t) = \pi/2 - \arccos(t)$: The probability that x and y lie on the same side of the line is $1 - \arccos(x \cdot y)/\pi$. \square

Using Grothendieck's Identity and $\chi(u) = \text{sign}(z \cdot f(u))$, we get that the sum appearing on the right-hand side of Eq. (1.6) equals

$$\sum_{u,v \in V} A(u,v) \mathbb{E}_z [\chi(u)\chi(v)] = \sum_{u,v \in V} A(u,v) \frac{2}{\pi} \arcsin(f(u) \cdot f(v)).$$

The fact that the matrix A satisfies $\sum_{u \in V} A(u, v) = 0$ for every $v \in V$ then gives

$$\sum_{u, v \in V} A(u, v) \frac{2}{\pi} \arcsin(f(u) \cdot f(v)) = \sum_{u, v \in V} (-A(u, v)) \left(1 - \frac{2}{\pi} \arcsin(f(u) \cdot f(v))\right). \quad (1.7)$$

Define

$$\alpha_{\text{GW}} = \min \left\{ \frac{\arccos(t)}{1-t} : t \in [-1, 1] \right\} = .878\dots$$

Using the trigonometric identity $1 - 2 \arcsin(t)/\pi = \arccos(t)$ and $A(u, v) \leq 0$ for all $u \neq v$, we can now write and bound the right-hand side of Eq. (1.7) as

$$\sum_{\{u, v\} \in E} (-A(u, v)) \left(\frac{\arccos(f(u) \cdot f(v))}{1 - f(u) \cdot f(v)} \right) (1 - f(u) \cdot f(v)) \geq \alpha_{\text{GW}} \sum_{\{u, v\} \in E} (-A(u, v)) (1 - f(u) \cdot f(v)).$$

Now using $1 - f(u) \cdot f(u) = 0$ and $A(u, v) = 0$ for all $\{u, v\} \notin E$ allows us to sum over all pairs of vertices, making the above sum equal to

$$\sum_{u, v \in V} (-A(u, v)) (1 - f(u) \cdot f(v)) = \sum_{u, v \in V} A(u, v) f(u) \cdot f(v),$$

where in the identity we again used that $\sum_{v \in V} A(u, v) = 0$ for all $u \in V$. The last sum above is simply 4 times the optimum of SDP (1.4), which is in turn at least as large as the size of a maximum cut. Collecting the factor $1/4$ left behind in Eq. (1.6) gives that the expected size of a cut returned by Algorithm 1.1 is at least $.878\dots$ times the the size of a maximum cut.

Optimality of Goemans and Williamson's approximation algorithm. By exhibiting an explicit family of graphs, Karloff [Kar96], and later Feige and Schechtman [FS02], proved that Goemans and Williamson's analysis of their algorithm is in fact optimal, showing that strange-appearing number $.878\dots$ is an upper bound on the approximation ratio of the algorithm for those graphs. Khot, Kindler, Mossel and O'Donnell [KKMO04] showed that based on the assumption of a complexity-theoretic conjecture known as the Unique Games Conjecture (cf. Section 1.7.4), $.878\dots$ is in fact the best-possible approximation ratio achievable by *any* polynomial-time approximation algorithm.

1.7.3 The chromatic number and the Lovász theta number

The *chromatic number* of a graph is defined as the smallest number of colors needed to color its vertices such that no two adjacent vertices receive the same color. A coloring of the vertices that assigns different colors to adjacent pairs and uses k colors is said to be a *proper k -coloring* of the graph. Computing the chromatic number is a well-known NP-hard problem.

The theta number refers to the optimum of a celebrated semidefinite program introduced by Lovász [Lov79]. One of its many applications is that it gives a lower bound on the chromatic number of a graph. For this, we consider the *complement* of a graph $G = (V, E)$, denoted \overline{G} , which is the graph with vertex set V in which a pair of distinct vertices are an edge if and only if they are not an edge in G . The *theta number* of the complement of a graph $G = (V, E)$, denoted by $\vartheta(\overline{G})$, is the optimum of the following semidefinite program:

$$\begin{aligned} & \text{minimize} && \lambda \\ & \text{subject to} && Z \in \mathcal{S}_V^+ \\ & && Z(u, u) = \lambda - 1 \text{ for every } u \in V \\ & && Z(u, v) = -1 \text{ for every } \{u, v\} \in E, \end{aligned}$$

where \mathcal{S}_V^+ denotes the set of real positive semidefinite matrices whose rows and columns are indexed by the vertices of G .

The fact that the value $\vartheta(\overline{G})$ provides a lower bound for the chromatic number of G can be seen as follows. Suppose that G has a proper k -coloring. We associate with each vertex $v \in V$ a vector $f(v) \in \mathbb{R}^{\binom{k}{2}}$ whose coordinates are indexed by all unordered pairs $\{i, j\} \in \binom{\{1, \dots, k\}}{2}$.³ If the coloring assigns color i to v then we define $f(v)$ by

$$f(v)_{\{i,j\}} = \begin{cases} 1 & \text{if } j > i \\ -1 & \text{if } j < i \end{cases}$$

and setting all other entries to zero. The matrix $Z(u, v) = f(u) \cdot f(v)$ is feasible for the above SDP and has objective value k . It follows that $\vartheta(\overline{G}) \leq \chi(G)$.

Notice that there are only k different vectors in the set $(f(v))_{v \in V}$. So, although the vectors $f(v)$ have dimension $\binom{k}{2}$, they only span a k -dimensional space. Geometrically, the vectors $f(v)$ define a $(k - 1)$ -dimensional regular simplex whose vertices lie in a sphere of radius $\sqrt{k - 1}$: Vertices in the graph having the same color are sent to the same vertex in the regular simplex and vertices of different colors are sent to different vertices in the regular simplex.

³Throughout we denote by $\binom{S}{t}$ the family of all t -element subsets of a finite set S .

1.7.4 A little on the Unique Games Conjecture

In 2002 Khot [Kho02] introduced the Unique Games Conjecture (UGC) in order to make progress on the problem of obtaining hardness of approximation results for NP-complete problems. Before that, Håstad [Hås99] made significant advances in this area. However, for many problems exact approximation results remained unknown. Since its introduction, it has been shown that the UGC would imply many inapproximability results unknown to be obtainable otherwise [KN08, KN09]. Often such results are highly accurate, matching the approximation ratios of known algorithms. Examples of problems where exact UGC hardness results are known are MAX CUT [KKMO04], minimum vertex cover [KR08], kernel clustering [KN10], max- k CSP [ST09]. Perhaps the most striking result is due to Raghavendra [Rag08], who showed that truth of the UGC implies that there is a single generic SDP-based polynomial-time approximation algorithm for all constraint satisfaction problems that achieves the optimal approximation ratio.

One of several equivalent formulations of the UGC [Kho10] is as follows. For positive integer n , an instance of a *linear unique game over \mathbb{Z}_n* is a two-player nonlocal game given by a positive integer N and numbers $c_{ij} \in \mathbb{Z}_n$ for $i, j \in \{1, \dots, N\}$. At the start of the game a referee uniformly samples a pair i, j from the set $\{1, \dots, N\}$ and sends question “ i ” to Alice and question “ j ” to Bob. The players answer $a_i, b_j \in \mathbb{Z}_n$, respectively, and win if $a_i - b_j = c_{ij} \pmod{n}$.

1.7.2. CONJECTURE (UNIQUE GAMES CONJECTURE). *For any $0 < \varepsilon < 1$, there exists positive integer $n = n(\varepsilon)$ such that given a linear unique game over \mathbb{Z}_n with maximum classical winning probability $1 - \varepsilon$, there is no polynomial-time algorithm that finds a classical strategy whose winning probability is greater than ε .*

Recently, Arora, Barak and Steurer [ABS10] gave a sub-exponential-time algorithm with performance guarantee better than is allowed in the conjecture for any polynomial-time algorithm. Though this does not disprove the conjecture, it does show that it is on somewhat shaky ground.

In the context of nonlocal games it is natural to ask what happens to the UGC when we allow for entangled strategies. Kempe, Regev and Toner [KRT08] examined exactly this situation and showed that in this case, conjecture is false.