

CHARACTERIZING QUANTUM BEHAVIOURS

Polynomial Optimization in Quantum Information

Moisés Bermejo Morán

A thesis presented for the degree of
Doctor of Philosophy



Institute of Theoretical Physics
Jagiellonian University of Kraków
Poland
2024

Abstract

The main theme of this thesis is the characterization of *quantum behaviors* by employing algebraic techniques from the framework of polynomial optimization, specifically utilizing sum of squares decompositions that can be translated into semidefinite programs. In particular, we explore generalized Bell scenarios where the measurements act on subsystems that may overlap, as well as an algebraic approach to uncertainty relations. Additionally, we develop algebraic tools to perform computations with partially commutative variables, which can be readily applied to a variety of problems in quantum information.

The motivation for characterizing which physical models can be reproduced with quantum theory and separating them from other theories, both classical and post-quantum, is twofold. On the one hand, it is practically important to identify quantum behaviors that provide advantages in tasks such as cryptography, computation, and information processing. On the other hand, it holds fundamental significance to identify experimental settings that cannot be adequately described by the current state of quantum theory. Overcoming such limitations may hint at a more fundamental theory, such as quantum gravity, which could accommodate new phenomena responsible for such transgressions.

Streszczenie

Głównym tematem tej pracy jest charakteryzacja zachowań kwantowych za pomocą technik algebraicznych z zakresu optymalizacji wielomianowej, w szczególności wykorzystując rozkład na sumę kwadratów (tzw. formy kwadratowe), który może być przekształcony w programy półokreślone. W szczególności, zbadane zostały uogólnione scenariusze Bella, gdzie pomiary działają na podsystemach, które mogą się nakładać, oraz algebraiczne podejście do relacji nieoznaczoności. Ponadto opracowane zostały narzędzia algebraiczne do wykonywania obliczeń z częściowo przemiennymi zmiennymi, które mogą być łatwo zastosowane do różnych problemów z zakresu informacji kwantowej.

Motywacja do charakteryzacji, które modele fizyczne mogą być odtworzone za pomocą teorii kwantowej, jest dwojaka. Po pierwsze, istotne jest zidentyfikowanie zachowań kwantowych, które zapewniają przewagę w zadaniach takich jak kryptografia, obliczenia i przetwarzanie informacji. Po drugie, identyfikacja sytuacji eksperymentalnych, które nie mogą być odpowiednio opisane przez współczesną teorię kwantową, ma znaczenie dla badań podstawowych. W szczególności, obserwacja i identyfikacja łamania takich ograniczeń może prowadzić do bardziej fundamentalnej teorii, takiej jak grawitacja kwantowa, która mogłaby uwzględnić nowe zjawiska.

Contents

1	Introduction	9
1.1	Outline and contributions	9
1.2	Quantum Physics	11
1.3	Mathematical background	15
1.4	Semidefinite Programming	21
1.5	Polynomial Optimization	22
1.6	Sum of squares relaxations	27
1.7	Gelfand-Naimark-Segal construction	30
2	Special constraints	33
2.1	Gröbner bases	33
2.2	Commutation relations	44
2.3	Dimension constraints	51
2.4	Computational reductions	54
3	Correlation Settings	61
3.1	Bell scenarios	61
3.2	Correlation settings	63
3.3	Complete hierarchy	67
3.4	Settings with shared components	68
3.5	Contextuality settings	70
3.6	Local Hamiltonians	74
4	Uncertainty Relations	77
4.1	Tight uncertainty relations	77
4.2	Additive uncertainty relations	79
4.3	Lovász hierarchy	83
4.4	Applications in quantum information	85
5	Conclusions	89
	Bibliography	91

Chapter 1

Introduction

1.1 Outline and contributions

Chapter 1 introduces the basic concepts from quantum theory, algebra, analysis and convex optimization that will appear during the thesis. We also introduce the frameworks of polynomial optimization. Tools in polynomial optimization are not limited to efficient algorithms to bound the optimal value of a polynomial but also provide algebraic certificates in terms of sum of squares decompositions. The current use of polynomial optimization range from practical applications in control theory, signal processing, operational research, finances and machine learning to applications in other branches of mathematics such as combinatorics and algebraic geometry and other fields such as quantum information and biochemistry. The exposition with monoids is chosen to accommodate different polynomial optimization frameworks in the same footing, both highlighting their structural similarities and technical differences. In particular, we introduce the semidefinite programming relaxations that we will later use to characterize quantum behaviours in different scenarios.

Chapter 2 introduces an efficient approach to deal with equality constraints in polynomial optimization problems, both reducing the size and increasing the accuracy of the semidefinite programming approximations. A novel contribution in this chapter, based on the work [1], is an adapted polynomial optimization framework that interpolates between commutative and non-commutative settings exploiting the partial commutations, and other additional structure in a problem, to boost the computational performance. Partial commutations play a significant role in physical problems, since they encode the local structure in physical systems. Later, we adapt methods to impose dimension constraints on local operations with overlap. Last, we review some additional techniques in the literature to reduce the computational cost.

Chapter 3 focuses on quantum behaviours that are linear in the expectation values. That is, those that correspond to the expectation value under a quantum state of some polynomial expression in quantum measurements. First, we recall Bell scenarios, that correspond with local measurement protocols.

Linear behaviours in Bell scenarios, also called Bell correlations, had significant impact in the foundations of physics, since they rule out the possibility of local hidden variable models compatible with a *realistic* theory. That is, one for which measurements reveal predefined features of a physical system. Quantum theory predicts non classical Bell correlations that give advantages in quantum key distribution, random access codes, communication complexity and randomness generation. Bell correlations can be characterized with non-commutative polynomial optimization [2]. A novel contribution in this chapter, based on the publication [3], is the introduction of correlation settings, which generalize Bell scenarios by allowing measurements in overlapping subsystems. This gives rise to richer class of quantum behaviours that accommodate several problems in quantum information. Next, we discuss the differences with other generalizations of Bell scenarios known as contextuality settings [4].

Chapter 4 finally addresses the problem of characterizing non-linear quantum behaviours. That is, expressions involving higher order powers of expectation values under a quantum state of some polynomial expression in quantum measurements. Limitations on these quantum behaviours are typically known as *uncertainty relations*. Uncertainty relations do not only reveal fundamental limits for the information that can be accessed to in a physical system, but also find several practical applications in quantum information, from entanglement detection to security proofs in cryptography. First, we show how non linear quantum behaviours can be characterized with state polynomial optimization. The major contribution in this chapter, based on the publication [5], is a detailed analysis of additive uncertainty relations involving squares of expectation values. First, we recall the best bounds known for Pauli strings based on a combinatorial invariant of the graph encoding their commutation relations, called *Lovász number*. Next, we propose an efficient hierarchy of upper bounds based on state polynomial optimization whose first level recovers Lovász bound. Our algebraic approach based on sum of squares decompositions can be extended to more general settings. In particular, we obtain novel uncertainty relations for non-hermitian operators.

List of works

- M. B. Morán, A. Pozas Kerstjens, and F. Huber. *Bell inequalities with overlapping measurements*. Physical Review Letters **131**, 080201 (2023).
- M. B. Morán and F. Huber. *Uncertainty relations from state polynomial optimization*. Physical Review Letters **132**, 200202 (2024).
- M. B. Morán, A. Mishra & S. Pironio. *Polynomial Optimization with partially commutative variables*. Work in preparation.

Summary of contributions

- We propose a framework of partially commutative polynomial computations, interpolating between commutative and non-commutative ones [1].

In particular, we present a theory of Gröbner bases for polynomials on partially commuting letters, allowing for an efficient implementation of equality constraints.

- We propose measurement protocols that generalize Bell scenarios accounting for measurements on overlapping subsystems. We extend techniques to bound the maximal quantum value of Bell inequalities with constraints on the dimension to these scenarios and discuss applications in quantum information [3].
- We present an efficient hierarchy of upper bounds for additive uncertainty relations based on an algebraic formulation of the Lovász number [5]. We obtain novel uncertainty relations for non-hermitian operators.

1.2 Quantum Physics

A brief historical introduction to the mathematical structures for quantum mechanics can be found in [6, §1]. The most extended approach in quantum information is the Hilbert space formalism [7, §2.2.5], [8, §2].

Definition 1 (Hilbert space formalism). *Quantum observables on a finite dimensional physical system are described with self-adjoint linear operators on a finite dimensional Hilbert space H , and physical states with density matrices. A density matrix is a positive linear operator, $\rho \geq \mathbf{0}$ in $\mathcal{B}(H)$, with unit trace, $\text{tr}(\rho) = 1$. A measurement effect is an operator satisfying $\mathbf{0} \leq E \leq \mathbf{1}$, and it is called projective if $E^2 = E$. A positive operator valued measurement is a collection of effects that sum up to identity, for projective effects we say projective valued measurement. The expectation value of an observable $a \in \mathcal{B}(H)$ on a state $\rho \in \mathcal{B}(H)$ is given by Born's rule, denoted $\langle a \rangle_\rho = \text{tr}(a\rho)$, and for effects this represents the probability of observing the effect on the state.* [7, §2.2.5], [8, §2]

Observables are obliged ingredients in any physical theory, since they represent the accessible information that can be obtained from a physical system in a given state. An axiomatic approach to observables based on lattices [9] allows to build physical theories in more general structures that may better accommodate non-classical physical behaviour [10]. Later in this thesis we follow the algebraic approach to quantum theory (see Section 1.3) to exploit the connections with polynomial optimization (see Section 1.5). In this approach, quantum systems are described with C*-algebras or von Neumann algebras [6, 2.3.2], which in finite dimension reduce to the Hilbert space formalism.

Example 2 (Qudit). *The Hilbert space $H = \mathbb{C}^d$ represents a d -dimensional quantum system, called qudit. Linear operators on H can be expanded as* [11, §1]

$$\rho = \sum_{i,j=1}^d \rho_{ij} |i\rangle \langle j| \in \mathcal{B}(H) = \mathbb{C}^d \otimes (\mathbb{C}^d)^*, \quad (1.1)$$

where $\{|i\rangle\}_{i=1}^d$ is any orthonormal basis for $\mathbb{C}^d$ with dual $\{\langle i|\}_{i=1}^d$, i.e. $\langle i|j\rangle = \delta_{ij}$. Thus, they are described by $d \times d$ matrices with entries $\rho_{ij} \in \mathbb{C}$. States on H correspond with self-adjoint matrices $\rho_{ij} = \rho_{ji}^*$ with non-negative eigenvalues and satisfying $\text{tr} \rho = \sum_i \rho_{ii} = 1$. Two dimensional systems are called qubits.

The non-commutative features of quantum theory give rise to physical behaviours that can not be explained with classical theories. One prominent example is *nonlocality*, found in the ground-breaking experiments [12, 13], that shows the hypotheses of locality (space-like separated systems do not influence each other) and realism (physical properties have predefined values that are revealed during the measurement process) to be mutually incompatible [14], [15, §I.A].

[15, §II.A] **Definition 3** (Bell scenario). *The Bell scenario $B(n, m, k)$ consists of n parties with m local measurements of k outcomes each. Probability models are probability distributions of observing outcomes $\mathbf{a} = (a_1, \dots, a_n) \in [k]^n$ conditioned to measurements $\mathbf{x} = (x_1, \dots, x_n) \in [m]^n$ being performed. Thus, they satisfy $\sum_{\mathbf{a}} p(\mathbf{a}|\mathbf{x}) = 1$ for each $\mathbf{x} \in [m]^n$. A model is non-signalling if marginal distributions are well-defined, $p_1(a_1|x_1) = \sum_{\mathbf{a}} p(a_1, \mathbf{a}|x_1, \mathbf{x})$ independent of $\mathbf{x} \in [m]^{n-1}$. The collection of non-signalling models is denoted $\mathcal{N}(B)$. Deterministic models are those $p \in \mathcal{N}(B)$ with $\text{ran } p \subset \{0, 1\}$, these factorize as the product of their marginals, $p = p_1 \dots p_n$. Local models, denoted $\mathcal{L}(B)$, are the convex hull of deterministic models. Last, quantum models $\mathcal{Q}(B)$ are those that can be obtained with Born's rule. Namely, $p(\mathbf{a}|\mathbf{x}) = \text{tr}(\rho P_{a_1|x_1} \dots P_{a_n|x_n})$ for some state ρ and projectors $P_{a_i|x_i}$ on a Hilbert space H such that $\sum_{a_i} P_{a_i|x_i} = \mathbf{1}$ for each $x_i \in [m]$ and $[P_{a_i|x_i}, P_{a_j|x_j}] = \mathbf{0}$ when $i \neq j$. Quantum models with a finite dimensional realization can equivalently be realized in $H = H_1 \otimes \dots \otimes H_n$ with projectors $P_{a_i|x_i}$ acting on H_i .*

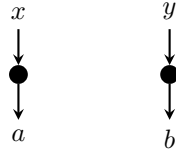


Figure 1.1: In a **bipartite Bell scenario** two parties perform measurements labelled x and y and obtain outcomes a and b with probability $p(a, b|x, y)$.

Bell experiments not only had breakthrough consequences in the foundations of physics but also find several technological applications. Nonlocality gives advantages in several informational tasks including conjugate coding [17], key distribution [18] communication complexity [19] or randomness verification [20]. Another non-classical feature of quantum theory is entanglement.

[21, §VII.A] **Definition 4** (Entanglement). *A quantum state ρ on the composite system $\otimes_s H_s$ is separable if it is a convex combination of tensor products of states on*

each local factor. Namely, $\rho = \sum_i \lambda_i \otimes_s \rho_{s,i}$ for some $0 \leq \lambda_i \leq 1$ adding up to one and states $\rho_{s,i}$ on H_s . Non separable states are called entangled.

Example 5 (Maximal entanglement). Consider the Hilbert space $H = \mathbb{C}^d \otimes \mathbb{C}^d$ [21, §II] that corresponds with two parties each with a qudit. The maximally entangled state or Bell state in H is (up to local unitaries)

$$\rho = \frac{1}{d} \sum_{i,j=1}^d |i\rangle |i\rangle \langle j| \langle j| \in \mathcal{B}(H) = (\mathbb{C}^d \otimes \mathbb{C}^d) \otimes (\mathbb{C}^d \otimes \mathbb{C}^d)^*, \quad (1.2)$$

where $\{|i\rangle\}_{i=1}^d$ is any orthonormal basis of $\mathbb{C}^d$ with dual $\{\langle i|\}_{i=1}^d$. Both reductions of a maximally entangled state are maximally mixed, i.e. $\mathbf{1}/d$. Indeed, for [7, §2.4.3] the basis above the reduction or partial trace over the second subsystem reads

$$\text{tr}_2 \rho = (\mathbf{1} \otimes \text{tr}) \rho = \frac{1}{d} \sum_i |i\rangle \langle i| = \frac{\mathbf{1}}{d}. \quad (1.3)$$

This means that we have no information about the local state of each subsystem.

Bell scenarios have close connections with bipartite entanglement. Separable states produce local models in all Bell scenarios [15, §III.A]. The converse is remarkably not true, namely, there exist entangled states that also produce local models in all Bell scenarios [22, §II]. However, all entangled states produce some model that can not be obtained from separable states in a Bell scenario assisted with ancillary quantum inputs [23, Corollary 1]. These scenarios are called *semi-quantum* and such models are termed *Buscemi nonlocal*.

Moreover, some entangled states can be uniquely determined, or *self-tested*, by the models they produce in a Bell scenario. For instance, Clauser-Horne-Shimony-Holt (CHSH) functional [12]

$$\begin{aligned} \beta = & p(00|00) - p(01|00) - p(10|00) + p(11|00) \\ & + p(00|01) - p(01|01) - p(10|01) + p(11|01) \\ & + p(00|10) - p(01|10) - p(10|10) + p(11|10) \\ & - p(00|11) + p(01|11) + p(10|11) - p(11|11), \end{aligned} \quad (1.4) \quad [15, §I.A]$$

self-tests for maximally entangled qubits in the Bell scenario $B(2, 2, 2)$. Namely, the maximum value of CHSH functional that can be obtained within quantum theory can only be attained when the parties share a pair of maximally entangled qubits and perform anti-commuting measurements [24]. Similarly, generalizations of CHSH functional self-test maximal entanglement in higher dimension d in Bell scenarios $B(2, m, d)$ with m local measurements of d outcomes [25]. Linear functionals on the correlations obtained in Bell scenarios are also called *Bell inequalities*, because they separate local from quantum and quantum from non-signalling models [15, §II.B].

Example 6 (Correlations in Bell scenario). *Examples of models for Bell scenario $\mathfrak{S} = B(2, 2, 2)$ are given by the following conditional probability tables, where the value at position (ab, xy) corresponds with $p(ab|xy)$:*

$$\begin{aligned} p_1 &= \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \in \text{ext } \mathcal{L}(\mathfrak{S}), & p_2 &= \frac{1}{2} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 \end{bmatrix} \in \mathcal{L}(\mathfrak{S}), \\ p_3 &= \frac{1}{2} \begin{bmatrix} c & c & c & \bar{c} \\ \bar{c} & \bar{c} & \bar{c} & c \\ \bar{c} & \bar{c} & \bar{c} & c \\ c & c & c & \bar{c} \end{bmatrix} \in \mathcal{Q}(\mathfrak{S}) - \mathcal{L}(\mathfrak{S}), & p_4 &= \frac{1}{2} \begin{bmatrix} 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{bmatrix} \in \mathcal{N}(\mathfrak{S}) - \mathcal{Q}(\mathfrak{S}), \end{aligned} \quad (1.5)$$

with $c = \cos^2(\pi/8)$ and $\bar{c} = 1 - c$. It is immediate to check that p_1 is non signaling and deterministic and p_2 is the convex combination of two such models, thus both are local. Model p_4 corresponds with a Popescu-Rohrlich box [26]. It is non signaling and has CHSH value $\beta(p_4) = 4$, breaking the bound for quantum models $\beta(p) \leq 2\sqrt{2}$ [27]. Similarly, model p_3 has CHSH value $\beta(p_3) = 2\sqrt{2}$, breaking the bound for local models $\beta(p) \leq 2$ [12]. Moreover, it can be obtained with a pair of maximally entangled qubits and spin measurements on directions with angles 0 and $\pi/2$ on one party and $\pi/4$ and $3\pi/4$ on the other [27]. More explicitly, the maximally entangled state of two qubits is

$$\rho = \frac{1}{2}(|0\rangle|0\rangle + |1\rangle|1\rangle)(\langle 0| \langle 0| + \langle 1| \langle 1|) \in \mathcal{B}(\mathbb{C}^2 \otimes \mathbb{C}^2). \quad (1.6)$$

The projectors to the spin components on direction with angle θ are

$$\begin{aligned} P_{0|\theta} &= (\sin \theta |0\rangle + \cos \theta |1\rangle)(\sin \theta \langle 0| + \cos \theta \langle 1|), \\ P_{1|\theta} &= (\sin \theta |0\rangle - \cos \theta |1\rangle)(\sin \theta \langle 0| - \cos \theta \langle 1|). \end{aligned} \quad (1.7)$$

The expectation values $p(a, b|x, y) = \text{tr}(\rho P_{a|\theta_x} \otimes P_{b|\theta_y})$ give the model p_3 .

In Chapter 3 we generalize Bell scenarios to protocols where the measurements act on systems that may overlap, which we call *correlation settings*. Following [2, 28, 29], we borrow techniques from polynomial optimization, summarized in Sections 1.5 and 1.6, to characterize quantum behaviours in these settings. In particular, we bound the optimal quantum value of Bell functionals with overlapping measurements [3].

Another characteristic feature of quantum theory are *uncertainty relations*. These relations manifest restrictions on the expectation values between non-commuting observables on quantum states, and find several applications in quantum technologies [30, 31, 32, 33], including quantum error correction [34] and quantum key distribution [18].

One paradigmatic example of uncertainty relation for finite dimensional systems can be found in Pauli matrices,

$$\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \quad (1.8)$$

These anti-commuting observables, $\sigma_i \sigma_j = -\sigma_j \sigma_i$ when $i \neq j$, correspond with spin measurements on the three spacial directions of a spin 1/2 particle, which is a two dimensional physical system described by the qubit space $\mathcal{B}(\mathbb{C}^2)$. Pauli matrices have eigenvalues 1 and -1, whose eigenvectors in the *computational basis* (the eigenbasis of σ_z) read

$$\begin{aligned} |+\rangle &= \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle), & |+_i\rangle &= \frac{1}{\sqrt{2}}(|0\rangle + i|1\rangle), & |0\rangle &= |0\rangle, \\ |-\rangle &= \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle), & |-_i\rangle &= \frac{1}{\sqrt{2}}(|0\rangle - i|1\rangle), & |1\rangle &= |1\rangle. \end{aligned} \quad (1.9)$$

The expectation values of Pauli matrices on a state $\rho \in \mathcal{B}(\mathbb{C}^2)$ verify [35, §5.2]

$$\langle \sigma_x \rangle^2 + \langle \sigma_y \rangle^2 + \langle \sigma_z \rangle^2 \leq 1. \quad (1.10)$$

That is, the expectation values of Pauli matrices live in a unit sphere, often called *Bloch ball*. Remarkably, points (x, y, z) in the Bloch ball are in correspondence with quantum states $\rho \in \mathcal{B}(\mathbb{C}^2)$ via $\rho = 1/2(\mathbf{1} + x\sigma_x + y\sigma_y + z\sigma_z)$. Indeed, $\{\mathbf{1}, \sigma_x, \sigma_y, \sigma_z\}$ is an orthogonal basis of $\mathcal{B}(\mathbb{C}^2)$ with respect to the inner product $\langle a, b \rangle = \text{tr}(ab^*)$ and each σ_i has zero trace, thus the component on $\mathbf{1}$ is fixed by $\text{tr}(\rho) = 1$. Positivity condition $\rho \geq \mathbf{0}$ translates to $x^2 + y^2 + z^2 \leq 1$.

The expression in Eq. (1.10) is an additive uncertainty relation. Additive uncertainty relations find applications in spectroscopy and atomic clocks [36], and quantum metrology [37]. In Chapter 4 we use the framework of state polynomial optimization [38] to characterize the maximum quantum value of polynomial expressions in the expectation values of operators satisfying prescribed algebraic relations. Moreover, we provide techniques to obtain additive uncertainty relations, which in particular generalize the relation (1.10) for Pauli matrices to higher dimensional spin operators.

1.3 Mathematical background

This section introduces the structures that will accommodate all instances of polynomial optimization we consider in this thesis as well as the algebraic approach to quantum theory. We mainly refer to standard references in algebra [39], topology [40] and functional analysis [41].

Definition 7 (Monoid). *A monoid $(\mathbb{M}, *, \mathbf{1})$ is a set equipped with an associative product $*$ and identity element $\mathbf{1}$. That is, (i) $(x * y) * z = x * (y * z)$ and (ii) $x * \mathbf{1} = x = \mathbf{1} * x$ for each $x, y, z \in \mathbb{M}$. When the monoid structure is clear from context we simply write $\mathbb{M}$. We say $\mathbb{M}$ is abelian or commutative if $x * y = y * x$ for each $x, y \in \mathbb{M}$.* [39, §1]

Monoids are very flexible structures. The first examples of a monoid are the natural numbers with respect to addition, $(\mathbb{N}, +, 0)$, or multiplication, $(\mathbb{N}, \cdot, 1)$. Monoids with inverses, for instance the integers, $(\mathbb{Z}, +, 0)$, form a special class.

[39, §I] **Definition 8** (Groups). *A group $(G, +, 0)$ is a monoid with the extra feature that every element has an inverse. Namely, for each $a \in G$ there exists $b \in G$ such that $a + b = 0$.*

Integer numbers with addition, $(\mathbb{Z}, +, 0)$, or rational numbers with multiplication, $(\mathbb{Q}, \cdot, 1)$, are abelian groups. The set $(\mathcal{S}_n, \circ, \mathbf{1})$ of permutations of n elements forms a non abelian (for $n > 2$) group under composition, whose identity is the trivial permutation, $\mathbf{1}(i) = i$ for each $i = 1, \dots, n$. In order to have linear combinations, we need a linear structure that is compatible with the monoid operation. Namely, addition distributes over taking products.

[39, §II] **Definition 9** (Rings). *A unital ring $(R, +, \mathbf{0}, *, \mathbf{1})$ is both an abelian group $(R, +, \mathbf{0})$ and a monoid $(R, *, \mathbf{1})$ that satisfies the distributive laws (i) $x*(y+z) = x*y + x*z$ and (ii) $(y+z)*x = y*x + z*x$ for each $x, y, z \in R$.*

The integers, $(\mathbb{Z}, +, 0, \cdot, 1)$, form a commutative ring, while square matrices of size n with matrix multiplication $(\text{Mat}(n), +, \mathbf{0}, \cdot, \mathbf{1})$, form a non-commutative ring. Other examples of rings that we will introduce in Section 1.3 are the commutative polynomial ring $k[X]$ and the non-commutative polynomial ring $k\langle X \rangle$ on letters X . Commutative rings for which both operations have inverses are called *fields*.

[39, §II] **Definition 10** (Fields). *A unital ring $(R, +, \mathbf{0}, *, \mathbf{1})$ for which $(R, *, \mathbf{1})$ is also an abelian group is a field. Elements of a field are called scalars.*

For instance, $(k, +, 0, \cdot, +)$, where k is either the set $\mathbb{Q}$ of rational, $\mathbb{R}$ real or $\mathbb{C}$ complex numbers, is a field. When we write k in this thesis we will mean any of these fields of numbers. Other examples of fields are the finite fields with modular addition and multiplication. Of special interest are those linear structures with scalar multiplication over a fixed field.

[39, §III] **Definition 11** (Vector space). *A vector space over a field k is an abelian group $(V, +, \mathbf{0})$ equipped with scalar multiplication such that (i) $a(x+y) = ax + ay$, (ii) $(a+b)x = ax + bx$ and (iii) $(ab)x = a(bx)$ for all $a, b \in k$ and $x, y \in V$. We say a map $f: V \rightarrow W$ between vector spaces is linear if $f(x+y) = f(x) + f(y)$ for all $x, y \in V$. The collection of linear maps from V to W is denoted $L(V, W)$.*

For instance, the complex numbers $\mathbb{C}$ form a vector space over $\mathbb{R}$ spanned by 1 and i . Finally, rings that are also vector spaces are called algebras.

[39, §II] **Definition 12** (Algebra). *A ring $(A, +, \mathbf{0}, *, \mathbf{1})$ endowed with scalar multiplication over a field k is called an algebra over k . That is, $(A, +, \mathbf{0})$ is a vector space over k , and scalar multiplication is also compatible with the product, $a(x*y) = (ax)*y$ for each $a \in k$ and $x, y \in A$. An involution is a unipotent unary linear operation, $(x^*)^* = x$ for each $x \in A$. A $*$ -algebra (also involutive*

algebra) is an algebra endowed with an involution. A map $\phi : A \rightarrow B$ between algebras compatible with addition and multiplication, $\phi(x + y) = \phi(x) + \phi(y)$ and $\phi(x * y) = \phi(x) * \phi(y)$, is called a homomorphism. A homomorphism between $*$ -algebras that is compatible with involution, $\phi(x^*) = \phi(x)^*$, is called a $*$ -homomorphism. We write $[x, y] = x * y - y * x$ for the commutator of $x, y \in A$.

The real numbers $\mathbb{R}$ and complex numbers $\mathbb{C}$ form $*$ -algebras with involution given by identity and complex conjugation respectively. Square matrices of size n with coefficients in k form a $*$ -algebra over k with involution given by entry-wise involution on the transpose. Moreover, commutative polynomials $k[X]$ and non-commutative polynomials $k\langle X \rangle$ will be examples of commutative and non-commutative $*$ -algebras over k .

Definition 13 (Ideal). An ideal of a ring $(R, +, \mathbf{0}, *, \mathbf{1})$ is a set $I \subset R$ closed under addition and multiplication with R . That is, $I + I \subset I$ and $R * I * R \subset I$. Two elements $x, y \in R$ are said to be related with respect to I if $x - y \in I$. The quotient ring of R with respect to I is the collection R/I of equivalence classes in R with respect to I . This has a natural ring structure with addition $(x + I) + (y + I) = (x + y) + I$ and multiplication $(x + I) * (y + I) = x * y + I$ for $x, y \in R$. An ideal is called simple if $\{\mathbf{0}\}$ is the only proper ideal it contains. [39, §II.1]

Intuitively, an ideal is the collection of all logical consequences of setting some elements in a ring (or algebra) to be zero. These closure properties with respect to the ring operations provide a natural ring structure over the equivalence classes.

Example 14. The quotient of the commutative polynomial ring $R = k[x, y]$ with respect to the ideal $I = \langle x^2 \rangle = x^2 * R$ is precisely $k[x, y]/\langle x^2 \rangle = k[y] + x * k[y]$.

Finally, both to consider polynomial optimization and to describe physical systems we need to enrich these objects with additional topological and analytical structure. We briefly recall some elementary concepts and key results that will play a relevant role in later chapters.

Definition 15 (Topology). We say (X, τ) is a topological space if τ is a collection of subsets of X satisfying the closure axioms: (i) contains the empty set and the whole space, $\{\emptyset, X\} \subset \tau$, (ii) closed under unions $\bigcup \mathcal{U} \in \tau$ when $\mathcal{U} \subset \tau$ and (iii) closed under finite intersections $\bigcap \mathcal{U} \in \tau$ when $\mathcal{U} \subset \tau$ is finite. Elements of τ are called open sets and their complements closed sets. A map $f : X \rightarrow Y$ between topological spaces (X, τ_X) and (Y, τ_Y) is continuous if the preimage of open sets in Y is open in X . Namely, $f^{-1}(U) \in \tau_X$ for each $U \in \tau_Y$. A collection $\mathcal{B}$ of open subsets is called a basis for the topology if every open set contains some element of the basis. A topological space X is separable if there exists a countable $Y \subset X$ intersecting all non-empty open sets and compact if every covering of X with open sets has a finite sub-covering. [40, §I.1]

Examples of topological properties include separation, countability, connectedness, compactness and metrizability.

[40, §I.6] **Definition 16** (Convergence). A directed set $(I, \leq)$ is a set equipped with an upward directed preorder. That is, for each $i, j, k \in I$ (i) $i \leq i$, (ii) $i \leq k$ when $i \leq j$ and $j \leq k$, and (iii) there exists l such that $i \leq l$ and $j \leq l$. A net on a topological space (X, τ) indexed with a directed set I is a map $x : I \rightarrow X$, typically denoted $(x_i)_{i \in I}$. We say $x \in X$ is a limit point of $(x_i)_{i \in I}$ in X , if for each open set $U^x \in \tau$ containing x , there exists $j \in I$ such that $x_i \in U^x$ for each $i \geq j$. Nets with respect to $(\mathbb{N}, \leq)$ are called sequences.

Nets are essential to characterize the topological properties of a space. For instance, a topological space is *compact* if and only if every net has a convergent sub-net. A special class of well-behaved topological spaces in which sequences are enough to capture the topological properties are *metric spaces*.

[41, §I.1] **Definition 17** (Metric). We say (X, d) is a metric space if the metric or distance $d : X \times X \rightarrow \mathbb{R}_+$ is (i) non-degenerate, $d(x, y) = 0$ if and only if $x = y$, (ii) symmetric, $d(x, y) = d(y, x)$ and (iii) satisfies triangle inequality, $d(x, z) \leq d(x, y) + d(y, z)$, for all $x, y, z \in X$. A sequence $(x_n)_n$ in a metric space (X, d) is Cauchy if for each $\varepsilon \in \mathbb{R}_+$ there is a natural number N such that $d(x_m, x_k) < \varepsilon$ for all $m, k > N$. We say (X, d) is complete if every Cauchy sequence has a limit in X .

We will focus on metric spaces with additional vector space structure.

[41, §I.1] **Definition 18** (Norm). We say $(V, \|\cdot\|)$ is a normed space if $(V, +, \mathbf{0})$ is a vector space over k and the norm function $\|\cdot\| : X \rightarrow \mathbb{R}_+$ is (i) non-degenerate, $\|x\| = 0$ if and only if $x = \mathbf{0}$, (ii) homogeneous, $\|cx\| = |c| \|x\|$ and (iii) satisfies the triangle inequality, $\|x + y\| \leq \|x\| + \|y\|$ for any $x, y \in X$ and $c \in k$.

Any normed space $(V, \|\cdot\|)$ is a metric space with distance $d(x, y) = \|x - y\|$. At the same time, a metric induces a topology on V with basis given by the open balls. Namely, $B(x, \varepsilon) = \{y \in V : d(x, y) < \varepsilon\}$ for all $\varepsilon \in \mathbb{R}_+$ and $x \in V$. Conditions (ii) and (iii) respectively ensure that scalar multiplication and addition in V are continuous in the topology induced by the norm. We ask similar compatibility conditions when the space has additional structure. Complete normed spaces are called *Banach spaces*.

[41, §I.3] The space of linear operators $T : H \rightarrow K$ between Banach spaces H and K is denoted $L(H, K)$. We say $T \in L(H, K)$ is *bounded* if the expression

$$\|T\| = \sup\{\|Tx\|_K : \|x\|_H = 1\} \quad (1.11)$$

is finite. These are precisely those linear operators that are continuous with respect to the topologies induced by the norms. The collection $\mathcal{B}(H, K)$ of bounded linear operators $T : H \rightarrow K$ is a Banach space with respect to the *operator norm* in Eq. (1.11). We write $L(H) = L(H, H)$ and $\mathcal{B}(H) = \mathcal{B}(H, H)$. Moreover, $H^* = \mathcal{B}(H, k)$ is called the (topological) *dual space* of H .

[41, §III.1] **Example 19** (Sequence space). Let k be a field and p a natural number. The space $\ell_p = \{(x_n)_{n \in \mathbb{N}} \subset k : \sum_n |x_n|^p < \infty\}$ of p -bounded sequences in k is a

Banach space with respect to the p -norm, $\|x\|_p = (\sum_n |x_n|^p)^{1/p}$. Also, the space $\ell_\infty = \{(x_n)_{n \in \mathbb{N}} : \sup_n |x_n| < \infty\}$ of bounded sequences is a Banach space with respect to the ∞ -norm $\|x\|_\infty = \sup_n |x_n|$. In particular, $(\ell_1)^* = \ell_\infty$.

The weak topology on H , denoted $\sigma(H, H^*)$ is the smallest (under inclusion) topology making the maps $\psi_f(x) = f(x)$ continuous for all $f \in H^*$. The weak* topology on H^* , denoted $\sigma^*(H^*, H)$, is the smallest topology making the maps $\text{ev}_x(f) = f(x)$ continuous for all $x \in H$. The weak* topology appears in the following important result, that guarantees that the optimal solution of certain polynomial optimization problems can be arbitrarily approximated with semidefinite programs (cf. Section 1.6). [41, §V]

Theorem 20 (Banach-Alaouglu). *The unit ball of the dual of a normed space, $\text{Ball}(H^*) = \{f \in H^* : \|f\| \leq 1\}$, is compact in the weak* topology.* [41, §V]

The Banach spaces we will consider have a multiplication that behaves well with respect to the norm. This motivates the following definition.

Definition 21 (Banach algebra). *An associative algebra $(A, +, \mathbf{0}, *, \mathbf{1})$ over k that is a complete normed space $(A, \|\cdot\|)$ and satisfies $\|x * y\| \leq \|x\| \|y\|$ for all $x, y \in A$ is called a Banach algebra.* [41, §VII.1]

Continuous functionals over compact spaces form Banach algebras.

Example 22 (Continuous functionals). *Let X be a compact Hausdorff (different points can be separated with disjoint open sets) space, for instance, a compact subspace of the Euclidean space k^n . The space $\mathcal{C}(X, k)$ of continuous functionals $f : X \rightarrow k$ is a Banach algebra with respect to pointwise multiplication, $(f * g)(x) = f(x)g(x)$ for each $x \in X$.* [41, §VII.1]

The spaces we consider additionally have a well-behaved involution.

Definition 23 (C*-algebra). *A C*-algebra is a Banach algebra $(A, +, \mathbf{0}, *, \mathbf{1}, \|\cdot\|)$ with involution satisfying the C*-identity, $\|x * x^*\| = \|x\|^2$ for all $x \in A$.* [41, §VIII.1]

Notice that C*-identity ensures continuity of involution, since $\|x\| = \|x^*\|$. In particular, *-homomorphisms preserve the C*-algebra structure. Continuous functionals $\mathcal{C}(X, k)$ on a compact and Hausdorff space X form commutative C*-algebras with involution $f^*(x) = f(x)^*$. These are, essentially, all commutative C*-algebras [6, Theorem 2.1.11]. The paradigmatic example of non-commutative C*-algebras are algebras of operators, for which the norm and the involution are defined through an inner product. In particular, these are the fundamental structures in the mathematical framework of quantum theory (cf. Section 1.2).

Definition 24 (Hilbert space). *An inner product on a vector space $(H, +, \mathbf{0})$ over k is a function $\langle \cdot, \cdot \rangle : H \times H \rightarrow k$ that is (i) positive, $\langle x, x \rangle \geq 0$, (ii) conjugate symmetric, $\langle x, y \rangle = \langle y, x \rangle^*$ and (iii) linear, $\langle ax + by, z \rangle = a\langle x, z \rangle + b\langle y, z \rangle$ for all $a, b \in k$ and $x, y, z \in H$. The inner product is non-degenerate if (iv) $\langle x, x \rangle = 0$* [41, §I.1]

if and only if $x = \mathbf{0}$. A Hilbert space is a non-degenerate inner product space complete in the metric induced by the inner product, $d(x, y) = \sqrt{\langle x - y, x - y \rangle}$.

- [41, §I.1] **Example 25** (Sequence space). *The sequence space ℓ_2 is a Hilbert space with respect to the inner product $\langle x, y \rangle = \sum_n x_n y_n^*$, and this is essentially (up to isomorphism) the only separable infinite dimensional Hilbert space.*
- [41, §VII.1] The space $\mathcal{B}(H)$ of bounded linear operators on a Hilbert space, is a Banach algebra with respect to the operator norm in Eq. (1.11) induced by the inner product. Furthermore, we can define an involution using the inner product, which gives $\mathcal{B}(H)$ the structure of C*-algebra. The *adjoint* of $T \in \mathcal{B}(H, K)$ is the unique $T^* \in \mathcal{B}(K, H)$ satisfying $\langle Tx, y \rangle_K = \langle x, T^*y \rangle_H$ for each $x \in H$ and $y \in K$, which defines an involution on $\mathcal{B}(H)$.
- [41, §VIII.1] Remarkably, any C*-algebra can be presented as a subspace of some $\mathcal{B}(H)$, closed with respect to the operator norm [42, 43]. Notice that, contrary to finite dimensional vector spaces, there are many inequivalent reasonable topologies that can be defined over an algebra of operators. The *operator topology* $(\mathcal{B}(H), \tau_{\|\cdot\|})$ is induced by the operator norm in Eq. (1.11). The *strong operator topology* $(\mathcal{B}(H), \tau_{\text{SOT}})$ is the smallest (under inclusion) topology making $\phi_x(T) = \langle Tx, x \rangle$ continuous for any $x \in H$. The *weak operator topology* $(\mathcal{B}(H), \tau_{\text{WOT}})$ is the smallest topology making $\phi_{x,y}(T) = \langle Tx, y \rangle$ continuous for any $x, y \in H$. The latter two determine the following important structure.
- [41, §IX.5] **Definition 26** (von Neumann algebra). *A von Neumann algebra over $\mathcal{B}(H)$ is a unital *-subalgebra closed with respect to the weak (equivalently, strong) operator topology. Equivalently, a von Neumann algebra is a unital *-subalgebra of $\mathcal{B}(H)$ that coincides with its double commutant, $A = A''$. The commutant of $A \subset \mathcal{B}(H)$ is defined as $A' = \{b \in \mathcal{B}(H) : \forall a \in A (a * b = b * a)\}$. The center of a von Neumann algebra A is $Z(A) = A \cap A' = \{b \in A : \forall a \in A (a * b = b * a)\}$, when $Z(A) = k\mathbf{1}$ is the field of scalars, we say A is a factor.*
- [41, §IX.7] In particular, operator algebras $\mathcal{B}(H)$ are von Neumann algebras, which are in turn C*-algebras. Moreover, von Neumann algebras are often studied in the context of local algebras, since their tensor product is unique.
- [44, §XVI.1] **Definition 27** (Tensor product). *The tensor product of vector spaces V and W over k is the vector space $V \otimes W$ of equivalence classes of finite linear combinations of pairs $(v, w) \in V \times W$ under the following identifications: (i) $(v_1 + v_2, w_1) = (v_1, w_1) + (v_2, w_1)$, (ii) $(v_1, w_1 + w_2) = (v_1, w_1) + (v_1, w_2)$, (iii) $(sv_1, w) = s(v_1, w)$ and (iv) $(v_1, sw_1) = s(v_1, w_1)$ for all $v_1, v_2 \in V$, $w_1, w_2 \in W$ and $s \in k$. We denote $v \otimes w$ the equivalence class of (v, w) . When V and W are Hilbert spaces, the closure of $V \otimes W$ with respect to the inner product $\langle v_1 \otimes w_1, v_2 \otimes w_2 \rangle = \langle v_1, v_2 \rangle \langle w_1, w_2 \rangle$ is the tensor product Hilbert space. For finite dimensional vector spaces, $(V \otimes W)^* = V^* \otimes W^*$.*
- [45, §V.3.2] The tensor product of Hilbert spaces induces a unique tensor product of von Neumann algebras, $\mathcal{B}(V) \otimes \mathcal{B}(W) = \mathcal{B}(V \otimes W)$, which coincides with the closure of the tensor product as vector spaces with respect to both the strong and weak

operator topologies. Remarkably, there are several inequivalent C^* -algebra structures that can be defined over a tensor product in infinite dimension.

Spectral theory provides a *functional calculus* for self-adjoint bounded operators. As a consequence, operators with non-negative spectrum can be characterized as (hermitian) squares. In finite dimension, this amounts to the so called Gram decomposition of a positive semidefinite matrix. [41, §VIII.2]

Definition 28 (Positivity). *Let $(A, +, \mathbf{0}, *, \mathbf{1})$ be a C^* -algebra over k . We say $b \in A$ is positive and write $b \geq \mathbf{0}$ if $b = aa^*$ for some $a \in A$, moreover one can assume a is self-adjoint. A linear functional $\lambda : A \rightarrow k$ is positive if $\lambda(aa^*) \geq 0$ for all $a \in A$ and unital if $\lambda(\mathbf{1}) = 1$. Unital positive linear functionals are called states, the collection of which is denoted $\mathcal{S}(A)$. States that can not be written as a convex combination of other states are called pure states.* [41, §VIII.3] [41, §VIII.5]

In finite dimension, every density matrix $\rho \in \mathcal{B}(H)$ determines an algebraic state on $\mathcal{B}(H)$ via $\lambda_\rho(a) = \text{tr}(a\rho) = \langle a, \rho \rangle$ for $a \in \mathcal{B}(H)$ and every algebraic state can be obtained from a density matrix. Notice that density matrices can be defined for infinite dimensional Hilbert spaces with trace-class operators, for which $\tau(\rho) = \sum_i \langle \rho e_i, e_i \rangle$ for some (any) orthonormal basis $\{e_i\}_i$ of H defines a trace. However, in infinite dimension not every algebraic state corresponds with a density matrix, and those that do are called *normal*. Density matrices of the form $\lambda_v(a) = \langle av, v \rangle$ for some $v \in H$ are called *vector states*, and are pure states. In finite dimension, all pure states are vector states. [47, §III.2.1]

In the frameworks of non-commutative and state polynomial optimization [28, 38], variables are represented with operators in some C^* -algebra and then evaluated in a state. Under mild assumptions, this algebra can be assumed to be the algebra of operators on a (possibly infinite dimensional) separable Hilbert space, $\mathcal{B}(H)$ and the state can be assumed to be a vector state. In the framework of trace polynomial optimization [48] variables are represented with von Neumann algebras of finite type instead, which admit sufficiently many finite normal tracial states [49, Theorem V.2.4].

1.4 Semidefinite Programming

We follow standard references for convex optimization [50, 51].

Definition 29 (Convex cone). *Let H be a vector space over k and $\mathcal{K} \subset H$ a subset. We say $\mathcal{K}$ is a cone if $k_+\mathcal{K} \subset \mathcal{K}$ and convex if $t\mathcal{K} + (1-t)\mathcal{K} \subset \mathcal{K}$ for each $t \in [0, 1]$. If H is a Hilbert space and $\mathcal{K} \subset H$ a convex cone, the set $\mathcal{K}^* = \{y \in H : \forall x \in \mathcal{K} \langle x, y \rangle \geq 0\}$ is again a convex cone, called the dual of $\mathcal{K}$.* [51, §2.1.5] [51, §2.6.1]

For optimization purposes we are mostly interested in real Hilbert spaces.

[50, §4.2] **Definition 30** (Conic program). *Let H be a real Hilbert space with convex cone $\mathcal{K}$. A conic program is an optimization problem of the form*

$$\begin{aligned} \max \quad & \langle C, X \rangle \\ \text{s.t.} \quad & X \in X_0 + \mathcal{L}, \\ & X \in \mathcal{K}, \end{aligned} \tag{1.12}$$

for some $C, X_0 \in H$ and some linear subspace $\mathcal{L} \subset H$. Alternatively, the linear constraints can be written as $\langle A_i, X \rangle = b_i$ for a suitable collection of $A_i \in H$ and $b_i \in \mathbb{R}$. The dual of (1.12) is the conic program

$$\begin{aligned} \min \quad & -\langle X_0, Y \rangle \\ \text{s.t.} \quad & Y \in C + \mathcal{L}^\perp, \\ & Y \in \mathcal{K}^*, \end{aligned} \tag{1.13}$$

where $\mathcal{L}^\perp$ denotes the linear subspace orthogonal to $\mathcal{L}$.

The dual program (1.13) is constructed minimizing upper bounds to the primal, thus a dual feasible point gives an upper bound to the primal. This relation is called *weak duality*. In some cases, the optimal solutions of the primal and the dual problem coincide, a property called *strong duality*. A sufficient condition for strong duality is Slater condition: that is, there exists an interior point in the (primal or dual) feasible region [50, Theorem 4.2.1]. Conic programs can be efficiently solved with inner-point methods [52].

[50, §6.4] Positive semidefinite matrices form a self-dual convex cone in the Hilbert space $\text{Mat}(n)$ of $n \times n$ matrices. A *semidefinite program* is a conic program with respect to the positive semidefinite cone. Namely,

$$\begin{aligned} \max \quad & \langle C, X \rangle \\ \text{s.t.} \quad & X \in X_0 + \mathcal{L}, \\ & X \geq 0, \end{aligned} \tag{1.14}$$

$$\begin{aligned} \min \quad & -\langle X_0, Y \rangle \\ \text{s.t.} \quad & Y \in C + \mathcal{L}^\perp, \\ & Y \geq 0, \end{aligned} \tag{1.15}$$

for some matrices C and X_0 , and linear subspace $\mathcal{L} \subset \text{Mat}(n)$. The spaces can be complex provided C and X_0 are self-adjoint. Semidefinite programs are extensively used for a wide class of applications [53].

1.5 Polynomial Optimization

First, we introduce the monoids we will consider for optimization. The exposition with monoids is chosen to accommodate the frameworks of commutative [54], non-commutative [28], trace [48] and state [38] polynomial optimization in the same footing, both highlighting their structural similarities and technical differences; as well as to set the structures to which the computational reductions discussed in Chapter 2 apply.

Example 31 (Non-commutative monoid). *The free, also non-commutative, [55, §1.3], [56, §2.1] monoid over the alphabet X , denoted $\langle X \rangle$, is the collection of finite words on the alphabet with product given by concatenation of words and identity element the empty word. An involution on the alphabet X induces an involution on $\langle X \rangle$ via $(x_{i_1} \dots x_{i_k})^* = x_{i_k}^* \dots x_{i_1}^*$. The degree $|w|$ of a word is the total number of letters it contains, and $|w|_x$ denotes the number of occurrences of letter x .*

Monoids can be presented as free monoids with an equivalence relation.

Definition 32 (Monoid presentation). *Given a relation $\mathcal{R}$ on a monoid $\mathbf{M}$, [55, §2.1], [56, §2.1] denote $\mathcal{R}^\#$ the congruence induced by $\mathcal{R}$. That is, the smallest (i) reflexive, $x\mathcal{R}^\#x$, (ii) symmetric, $x\mathcal{R}^\#y$ if $y\mathcal{R}^\#x$, and (iii) transitive, $x\mathcal{R}^\#z$ if $x\mathcal{R}^\#y$ and $y\mathcal{R}^\#z$, relation containing $\mathcal{R}$ that is (iv) compatible with the monoid operation, $(x * u * y)\mathcal{R}^\#(x * v * y)$ if $u\mathcal{R}v$ for all $x, y, z, u, v \in \langle X \rangle$. A presentation of a monoid $\mathbf{M}$ consists of a set of generators X and a relation $\mathcal{R}$ on $\langle X \rangle$ such that $\mathbf{M}$ is isomorphic to the quotient space $\langle X \rangle / \mathcal{R}^\#$ of equivalence classes of free words with respect to the congruence induced by $\mathcal{R}$. In this case, we write $\mathbf{M} = \langle X; \mathcal{R} \rangle$. We denote the equivalence class of a free word $w \in \langle X \rangle$ as $[w]$.*

Any monoid can be presented as a free monoid with additional restrictions between free words [55, Th.7.1.7], $\mathbf{M} = \{\{x_u : u \in \mathbf{M}\}; \{(x_u x_v, x_{u*v}) : u, v \in \mathbf{M}\}\}$. These restrictions amount to monomial identifications. Some monoids can be presented with finitely many relations.

Example 33 (Commutative monoid). *When the order in which the letters [55, §3.6] appear in a word is not relevant, one can take equivalence classes and obtain the free commutative monoid $[X]$ generated by the alphabet X . That is, $[X]$ has presentation $\langle X; \mathcal{C} \rangle$, where $\mathcal{C} = \{(x_i x_j, x_j x_i) : x_i, x_j \in X\}$ is the collection of commutation relations between all letters in the alphabet.*

When $\mathcal{C}$ is an incomplete collection of commutation relations, $\langle X; \mathcal{C} \rangle$ is called *partially commutative monoid* [57]. From a computational standpoint, however, it is not always the most efficient to represent monomials as free words with additional relations. For commutative words, more efficient descriptions are available. For instance, a commutative word $w \in [X]$ is completely determined by the *exponent* (number of occurrences) of each letter and multiplication amounts to addition of the exponents. Efficient normal forms are also available for partially commutative monoids [57] (see Section 2.2).

Non commutative monomials express products of operators. For instance, the monomial $x * y * x \in \langle X \rangle$ formalizes an operator that can be obtained as a product of operators in some operator algebra, $a * b * a \in A$. However, sometimes we are interested in formalizing more general expressions including the expectation value of the operators on some indeterminate state or their traces. For instance, $\lambda(a)\lambda(b) \in A$ for some state $\lambda : A \rightarrow k$ or $\tau(a)b \in A$ for some tracial state $\tau : A \rightarrow k$. In order to formalize these expressions, we need to enrich the non-commutative monoid with additional structure. For this, we use state [38] and trace monoids [58, 48].

[38, §2.1] with $\varrho(\mathbf{1}) = \mathbf{1}$ **Definition 34** (State monoid). *Given an alphabet X and a new symbol ϱ , consider the state alphabet $X_\varrho = \{\varrho(w) : w \in \langle X \rangle\}$. The state monoid over X with state symbol ϱ is the direct (commutative) product of the free monoid over X and the free commutative monoid over X_ϱ . That is, $\mathbb{S}_\varrho(X) = \langle X \rangle \times [X_\varrho]$. Elements of $\mathbb{S}_\varrho(X)$, called state monomials, can be written as $\{w_0\varrho(w_1)\dots\varrho(w_n) : w_i \in \langle X \rangle\}$ and their multiplication*

$$(w_0\varrho(w_1)\dots\varrho(w_n)) * (v_0\varrho(v_1)\dots\varrho(v_m)) = (w_0v_0)\varrho(w_0)\dots\varrho(v_m). \quad (1.16)$$

For ease of notation, we denote $\varrho : \mathbb{S}_\varrho(X) \rightarrow [X_\varrho]$ the state projection induced by the state symbol via $\varrho(w_0\varrho(w_1)\dots\varrho(w_n)) = \varrho(w_0)\dots\varrho(w_n)$. An involution on X extends to $\mathbb{S}_\varrho(X)$ via $\varrho(w)^* = \varrho(w^*)$.

The state symbol is the formal analogue of the expectation value of operators under an algebraic state, thus they must commute. State monomials are products of operators and these formal expectation values. Similarly, one can consider formal analogues for tracial states, which also find several applications in quantum information [48].

[48, §2.1] with $\theta(\mathbf{1}) = \mathbf{1}$ **Definition 35** (Trace monoid). *Given an alphabet X and a new symbol θ , consider the trace alphabet $\dot{X}_\theta = \{\theta(w) : w \in \langle X \rangle/T\}$, where*

$$T = \{(x_1\dots x_n, x_{\sigma(1)}\dots x_{\sigma(n)}) : x_1\dots x_n \in \langle X \rangle, \sigma = (12\dots n)^l \in \mathcal{S}_n\}. \quad (1.17)$$

is the equivalence relation that identifies free words that differ in a cyclic permutation of their letters. The trace monoid over X with trace symbol θ is the direct (commutative) product of the free monoid over X and the free commutative monoid over $\dot{X}_\theta$. That is, $\mathbb{T}_\theta(X) = \langle X \rangle \times [\dot{X}_\theta]$. Elements of $\mathbb{T}_\theta(X)$, called trace monomials, can be written as $\{w_0\theta(w_1)\dots\theta(w_n) : w_i \in \langle X \rangle/T\}$ and their multiplication

$$(w_0\theta(w_1)\dots\theta(w_n)) * (v_0\theta(v_1)\dots\theta(v_m)) = (w_0v_0)\theta(w_0)\dots\theta(v_m). \quad (1.18)$$

For ease of notation, we denote $\theta : \mathbb{T}_\theta(X) \rightarrow [\dot{X}_\theta]$ the trace projection induced by the state symbol via $\theta(w_0\theta(w_1)\dots\theta(w_n)) = \theta(w_0)\dots\theta(w_n)$. An involution on X extends to $\mathbb{T}_\theta(X)$ via $\theta(w)^* = \theta(w^*)$.

The trace symbol is the formal analogue of the expectation value of operators under a tracial algebraic state, thus they must commute and satisfy the equivalence under cyclic permutations in Eq. (1.17). Trace monomials are products of operators and these formal expectation values under a tracial state.

[39, §II.3], [59, §4] **Definition 36** (Polynomial). *The monoid algebra of $\mathbb{M}$ over k , is the k -linear span of $\mathbb{M}$, denoted $k\langle\mathbb{M}\rangle$. An element $p \in k\langle\mathbb{M}\rangle$ is a finite linear combination of terms, $p = \sum_i c_i m_i$, each containing a coefficient $c_i \in k$ and a monomial $m_i \in \mathbb{M}$. The support of p is the collection of $m_i \in \mathbb{M}$ for which $c_i \neq 0$. An involution on $\mathbb{M}$ extends k -linearly to $k\langle\mathbb{M}\rangle$, giving it structure of $*$ -algebra. The monoid algebras over the free commutative, non-commutative, state and trace monoid are called commutative, non-commutative, state and trace polynomial algebras. Their elements are called polynomials.*

Non commutative polynomials can express arithmetical relations between operators. State polynomials can express arithmetical relations between operators and their expectation values. Finally, trace polynomials can express arithmetical relations between operators and their traces. For instance, the polynomial $x^2 - x \in k\langle X \rangle$ formalizes operators $a \in A$ in some C^* -algebra satisfying $a^2 = a$. The state polynomial $\varrho(x)\varrho(x^*) - 1 \in k\langle \mathbb{S}_\varrho(X) \rangle$ formalizes operators $a \in A$ whose expectation value under any state $\lambda : A \rightarrow k$ satisfies $|\lambda(a)|^2 = 1$. Finally, the trace polynomial $\theta(xy) - 1 \in k\langle \mathbb{T}_\theta(X) \rangle$ formalizes operators $a, b \in A$ whose expectation under any tracial state $\tau : A \rightarrow k$ satisfies $\tau(a * b) = \tau(b * a) = 1$. These polynomials are formal analogues of operators in the sense that they do not fix an operator algebra A nor states λ or τ .

The *evaluation* of a commutative polynomial on $a \in k^n$ is the image under [39, §II.3] the homomorphism $\text{ev}_a : k[X] \rightarrow k$ induced via $\text{ev}_a x_i = a_i$. For example, the evaluation of $p = x^2 + xy + 1 \in k[x, y]$ on $(1, 2)$ is $\text{ev}_{(1,2)} p = 1^2 + 1 \cdot 2 + 1 = 4$. We can define an evaluation for non-commutative polynomials representing non-commuting variables with operators to which we assign values with states, so that the partial order induced by positivity is preserved. This motivates the following definition.

Definition 37 (Evaluation). *An evaluation on a monoid $\mathbb{M}$ is a pair (λ, ψ) where $\psi : k\langle \mathbb{M} \rangle \rightarrow \mathcal{B}(H)$ is a $*$ -representation and $\lambda : \mathcal{B}(H) \rightarrow \mathbb{C}$ is a state on $\mathcal{B}(H)$. That is, ψ is a unital linear $*$ -homomorphism and λ a unital positive linear functional. The composition $\lambda \circ \psi : k\langle \mathbb{M} \rangle \rightarrow \mathbb{C}$ linearly assigns scalar values to polynomials. For state monomials we require the state to preserve the state symbol, that is, $\lambda \circ \psi(w_0 \varrho(w_1) \dots \varrho(w_n)) = \lambda \circ \psi(w_0) \lambda \circ \psi(w_1) \dots \lambda \circ \psi(w_n)$. For trace monomials, evaluations are restricted to finite von Neumann algebras [38, §2.2] with finite normal tracial states that preserve the trace symbol. A collection of operators $a = \{a_x \in \mathcal{B}(H) : x \in X\}$ induces a representation $\psi : k\langle X \rangle \rightarrow \mathcal{B}(H)$ via $\psi(x) = a_x$ for x in the alphabet X of $\mathbb{M}$. For a given state $\lambda : \mathcal{B}(H) \rightarrow \mathbb{C}$, this extends to state (and trace) polynomials via $\psi(\varrho(w)) = \lambda(\psi(w))$, and we often write $p(\lambda, a) = \psi p$, or simply $p(a) = \psi p$ if there are no state symbols.* [48, §2.2]

Example 38. Consider the polynomial $p = xz + zx \in k\langle x, z \rangle$ and the representation $\psi : k\langle x, z \rangle \rightarrow \mathcal{B}(\mathbb{C}^2)$ induced via $\psi x = X$ and $\psi z = Z$, where

$$X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \quad (1.19)$$

One verifies $\psi p = p(X, Z) = XZ + ZX = \mathbf{0}$. Then, for any state $\lambda : \mathcal{B}(\mathbb{C}^2) \rightarrow \mathbb{C}$, the evaluation (λ, ψ) on p is $\lambda \circ \psi p = 0$.

Example 39. Consider the state polynomial $p = x^2 - x\varrho(x) \in k\langle \mathbb{S}_\varrho(x) \rangle$ and the representation $\psi_0 : k\langle x \rangle \rightarrow \mathcal{B}(\mathbb{C}^2)$ induced via $\psi_0 x = X$ in Eq. (1.19). Given a state $\lambda : \mathcal{B}(\mathbb{C}^2) \rightarrow \mathbb{C}$, this extends to a representation $\psi : k\langle \mathbb{S}_\varrho(x) \rangle \rightarrow \mathcal{B}(\mathbb{C}^2)$ induced by $\psi\varrho(w) = \lambda(\psi_0 w)$ for each monomial $w \in \langle x \rangle$. We have $p(\lambda, X) = X^2 - X\lambda(X)$ and $\lambda(p(\lambda, X)) = \lambda \circ \psi p = \lambda(X^2) - \lambda(X)^2$. In particular, the state defined as $\lambda(a) = a_{11}$ for $a \in \mathcal{B}(\mathbb{C}^2)$ gives $\lambda(X^2) = 1$ and $\lambda(X) = 0$, thus

$\lambda(p(\lambda, X)) = \lambda \circ \psi p = 1$. Now let $\tau = 1/2 \operatorname{tr}$ be the unique tracial state on $\mathcal{B}(\mathbb{C}^2)$, the evaluation (τ, ψ) on p gives $\tau(p(\tau, X)) = \tau \circ \psi p = \tau(\mathbf{1}) - \tau(X)^2 = 1$.

Evaluations associate complex numbers to polynomials preserving the partial order induced via positivity. Namely, $\lambda \circ \psi(p) - \lambda \circ \psi(q) \geq 0$ when $p - q \geq \mathbf{0}$ in $k\langle \mathbb{M} \rangle$ for any evaluation (λ, ψ) . Moreover, the evaluation of a symmetric polynomial, $p = p^*$, is a real number since $\lambda \circ \psi(p)^* = \lambda \circ \psi(p^*) = \lambda \circ \psi(p)$.

In the non-commutative case, evaluations correspond with eigenvalues of bounded operators on a separable Hilbert space, as considered in [28, §1]. For state monomials, evaluations correspond with expectation values of bounded operators over an algebraic state, as considered in [38, §2.2]. For trace polynomials, evaluations on finite von Neumann algebras with finite normal tracial states are considered in [48, §2.2]. For commutative polynomials, evaluations reduce to probability measures as considered in [54, Proposition 2.1]. Indeed, unital commutative algebras of operators are exactly spaces of continuous functionals over a compact Hausdorff space [41, §VIII.2], for which states correspond with probability distributions by Riesz representation theorem [41, Theorem III.5.7]. Then, $\lambda \circ \psi(w)$ corresponds with the *moment* of the monomial w , as considered in [54, §2].

[60, §6.6] for $k\langle X \rangle$ **Definition 40** (Polynomial optimization problem). *A polynomial optimization problem over $k\langle \mathbb{M} \rangle$ is a triple (p, G, S) , where $p = p^* \in k\langle \mathbb{M} \rangle$ is a symmetric polynomial to maximize over a region specified by lists of polynomial equality constraints $G \subset k\langle \mathbb{M} \rangle$ and inequality constraints $S \subset k\langle \mathbb{M} \rangle$. The optimal value of the polynomial optimization problem (p, G, S) is given by*

$$\begin{aligned} \max_{(\lambda, \psi)} \quad & \lambda \circ \psi p & (1.20) \\ \text{s. t.} \quad & \psi g = \mathbf{0} & \forall g \in G \\ & \psi s \geq \mathbf{0} & \forall s \in S. \end{aligned}$$

The maximization runs over all evaluations (λ, ψ) of the monoid $\mathbb{M}$ on some separable Hilbert space. It is common to write the problem simply as

$$\begin{aligned} \max \quad & p & (1.21) \\ \text{s. t.} \quad & g = \mathbf{0} & \forall g \in G \\ & s \geq \mathbf{0} & \forall s \in S. \end{aligned}$$

When $p = \mathbf{0}$, we say $(\mathbf{0}, G, S)$ is a polynomial feasibility problem.

Although not strictly necessary, we assume the polynomial constraints to be symmetric. Notice that equality constraints can be presented as pairs of inequality constraints with opposite signs. Thus, (p, G, S) and $(p, \emptyset, S \cup G \cup -G)$ represent the same polynomial optimization problem. We distinguish equality constraints because these can be implemented more efficiently with the techniques discussed in Chapter 2. Although most algebraic computations can be performed in more general monoids and polynomial algebras, we restrict the

optimization problems to four polynomial algebras for which we have powerful positivity certificates that give rise to efficient relaxations. These are:

- [54] **Commutative** polynomials $k\langle X; \mathcal{C} \rangle$.
- [28] **Non commutative** polynomials $k\langle X \rangle$.
- [38] **State** polynomials $k\langle \mathbb{S}_\varrho(X) \rangle$.
- [48] **Trace** polynomials $k\langle \mathbb{T}_\theta(X) \rangle$.

For applications in quantum information, it is desirable to restrict the optimization to *physical states*. That is, those that correspond with density matrices. It turns out that, under mild regularity assumptions on the sets of constraints (see Section 1.6) the optimal value can be obtained with physical states from Gelfand-Naimark-Segal construction. In particular, this is the case for all problems in quantum information we consider in Chapters 3 and 4. Thus, we can lift such assumptions in the optimization.

1.6 Sum of squares relaxations

For a collection $S \subset k\langle \mathbb{M} \rangle$ of polynomials, the *feasible evaluations* for S are those evaluations (λ, ψ) for which ψ assigns non-negative values to all polynomials in S , denoted $\mathcal{F}(S) = \{(\lambda, \psi) : \psi s \geq \mathbf{0} \forall s \in S\}$. In particular, for pairs $\{-s, s\} \subset S$, we have $\psi s = \mathbf{0}$ when $(\lambda, \psi) \in \mathcal{F}(S)$. Notice that λ enters the condition on ψ above for state and trace polynomials, for which ψ replaces the state and trace symbols by λ . The images of the set X of generators of $\mathbb{M}$ under feasible representations form the so called *semialgebraic set* defined by S denoted $\mathcal{D}(S) = \{\psi(X) = (\psi x_1, \dots, \psi x_n) : (\lambda, \psi) \in \mathcal{F}(S)\}$ [60, §6.1], [38, §2.2] and [48, §2.2]. We say a polynomial $p \in k\langle \mathbb{M} \rangle$ is *positive* over S if $\psi p \geq \mathbf{0}$ for each feasible evaluation $(\lambda, \psi) \in \mathcal{F}(S)$ [60, §6.1].

Definition 41 (Quadratic module). *For a collection $S \subset k\langle \mathbb{M} \rangle$ of polynomials, [60, §6.1] the quadratic module generated by S , denoted $\mathcal{Q}(S)$, is the closure of $S \cup \{\mathbf{1}\}$ under addition, scalar multiplication and hermitian multiplication in the sense that $fsf^* \in \mathcal{Q}(S)$ for $s \in \mathcal{Q}(S)$ and $f \in k\langle \mathbb{M} \rangle$. For state (trace) polynomials, we additionally require $\mathcal{Q}(S)$ to be closed under state (trace) projection. Namely, [38, §5.1], [48, §4] $\varrho(s) \in \mathcal{Q}(S)$ ($\theta(s) \in \mathcal{Q}(S)$) for each $s \in \mathcal{Q}(S)$.*

In particular, any $p \in \mathcal{Q}(\emptyset) \subset k\langle X \rangle$ is a sum of hermitian squares ff^* . Thus, $\psi p \geq \mathbf{0}$ for each evaluation (λ, ψ) . Similarly, $p \in \mathcal{Q}(S)$ is a sum of terms of the form fsf^* , where $s \in S \cup \{\mathbf{1}\}$. Thus, $\psi p \geq \mathbf{0}$ for each evaluation $(\lambda, \psi) \in \mathcal{F}(S)$ feasible for S . The situation is similar for state (trace) polynomials. In particular, pure state (trace) polynomials $\varrho(p) \in \mathcal{Q}(S) \subset k\langle \mathbb{S}_\varrho(X) \rangle$ are sums of terms of the form $\varrho(f_0 s f_0^*) \varrho(f_1 f_1^*) \dots \varrho(f_n f_n^*)$, where $s \in S \cup \{\mathbf{1}\}$ [38, Lemma 5.6] ([48, Lemma 4.1]). Thus, $\psi \varrho(p) \geq 0$ for each $(\lambda, \psi) \in \mathcal{F}(S)$.

Remarkably, the converse is also true under mild assumptions, for instance, when the quadratic module $\mathcal{Q}(S)$ is *archimedean*, i.e. $N - \sum_{x \in X} x x^* \in \mathcal{Q}(S)$ [60, §6.1]

for some natural number N and set of generators X . That is, a polynomial positive over S belongs to the quadratic module $\mathcal{Q}(S)$. These certificates for positivity, also called *Positivstellensätze*, were introduced in [61, Lemma 4.1] for commutative polynomials, [62, Theorem 1.2] for non-commutative polynomials, [38] for state polynomials and [48] for trace polynomials and used in [54, 28, 38, 48] for polynomial optimization.

[60, Theorem 2.3] **Proposition 42** (Putinar). *A polynomial $p \in k[X]$ is positive over S if and only if $p + \varepsilon \in \mathcal{Q}(S)$ for each $\varepsilon > 0$. That is, it decomposes as*

$$p + \varepsilon = \sum_i f_i^2 s_i \quad (1.22)$$

for some polynomials $s_i \in S \cup \{1\}$ and $f_i \in k[X]$.

Also for non-commutative polynomials [62].

[60, Theorem 6.2] **Proposition 43** (Helton-McCullough). *A polynomial $p \in k\langle X \rangle$ is positive over S if and only if $p + \varepsilon \in \mathcal{Q}(S)$ for each $\varepsilon > 0$. That is, it decomposes as*

$$p + \varepsilon = \sum_i f_i s_i f_i^* \quad (1.23)$$

for some polynomials $s_i \in S \cup \{1\}$ and $f_i \in k\langle X \rangle$.

Recently, also for trace polynomials.

[48, Theorem 4.4] **Proposition 44** (Klep-Magron-Volčič). *A trace polynomial $\theta(p) \in k\langle \mathbb{T}_\theta(X) \rangle$ is positive over S if and only if $\theta(p) + \varepsilon \in \mathcal{Q}(S)$ for each $\varepsilon > 0$. That is,*

$$\theta(p) + \varepsilon = \sum_i \theta(f_{i,0} s_i f_{i,0}^*) \theta(f_{i,1} f_{i,1}^*) \dots \theta(f_{i,n_i} f_{i,n_i}^*) \quad (1.24)$$

for some polynomials $s_i \in S \cup \{1\}$ and $f_{i,j} \in k\langle \mathbb{T}_\theta(X) \rangle$.

The certificate in Proposition 44 does not hold for non-purely tracial polynomials [48, Example 4.6]. The situation is similar for state polynomials.

[38, Corollary 5.8] **Proposition 45** (Klep-Magron-Volčič-Wang). *A state polynomial $\varrho(p) \in k\langle \mathbb{S}_\varrho(X) \rangle$ is positive over S if and only if $\varrho(p) + \varepsilon \in \mathcal{Q}(S)$. That is, it decomposes as*

$$\varrho(p) + \varepsilon = \sum_i \varrho(f_{i,0} s_i f_{i,0}^*) \varrho(f_{i,1} f_{i,1}^*) \dots \varrho(f_{i,n_i} f_{i,n_i}^*) \quad (1.25)$$

for some polynomials $s_i \in S \cup \{1\}$ and $f_{i,j} \in k\langle \mathbb{S}_\varrho(X) \rangle$.

Positivity certificates are an intense area of research in real algebraic geometry. One of their several applications is transforming polynomial optimization problems into *sum of squares* decomposition. In particular, those certificates for archimedean quadratic modules in Propositions 42, 43, 44 and 45 allow to write a polynomial optimization problem as an infinite semidefinite program, which can be approximated with semidefinite programs of finite size [54, 28, 48, 38].

Indeed, the maximal value of a polynomial p over S can be expressed as [60, §2.3, §6.6] the smallest real number t such that $t - p$ is positive over S . Equivalently, the smallest t such that $t - p \in \mathcal{Q}(S)$ admits a decomposition into sums of squares. Sum of squares in $k\langle \mathbb{M} \rangle$ correspond with quadratic forms on $k\langle \mathbb{M} \rangle$ as a vector space via the eigenvalue decomposition. In particular, for a fixed basis this corresponds with positive semidefinite matrices. Namely, a positive semidefinite (infinite) matrix $Q \geq 0$ with eigenvalue decomposition $Q = P\Lambda P^*$ produces the sum of squares $\sum_i \Lambda_{ii} P_i P_i^*$, where P_i denotes the i -th eigenvector. Moreover, sum of squares decompositions restricted to finite subspaces $V \subset k\langle \mathbb{M} \rangle$ correspond with positive semidefinite matrices indexed with bases $\mathcal{I}$ of V . Optimizing over such matrices with linear constraints is a semidefinite program of size $|\mathcal{I}| = \dim V$. For the ease of exposition, we consider a linear subspace V with a basis $\mathcal{I}$ of monomials. In such a case, the relaxation of problem $(p, \emptyset, \emptyset)$ corresponds with the semidefinite program

$$\begin{aligned} \min \quad & t \\ \text{s. t.} \quad & (t - p)_w = \sum_{u^* v = w} Q_{u,v}, \\ & Q \geq 0, \end{aligned} \tag{1.26}$$

where $(t - p)_w$ denotes the coefficient of monomial w in $t - p$. Notice that, in order for a subspace V to admit a feasible solution $t - p$ must belong to the linear subspace spanned by V^*V . Equivalently, the support of $t - p$ must be contained in $\mathcal{I}^*\mathcal{I}$. We call such a semidefinite program the *sum of squares* relaxation of $(p, \emptyset, \emptyset)$ restricted to V or indexed by $\mathcal{I}$.

Similarly, terms fsf^* for $s = \sum_x s_x x \in S$ where f is restricted to live in a subspace V^s are introduced with an additional positive semidefinite matrix Q^s such that $Q_{u,v}^s$ contributes to the monomials u^*xv with weight s_x for each x in the support of s . More concretely, for monomial bases $\mathcal{I}^s$ of the linear subspaces V^s , with $s \in S \cup \{1\}$, the relaxation of problem $(p, \emptyset, S)$ reads

$$\begin{aligned} \min \quad & t \\ \text{s. t.} \quad & (t - p)_w = \sum_{u^* v = w} Q_{u,v} + \sum_{s \in S} \sum_{u^* x v = w} s_x Q_{u,v}^s, \\ & Q, Q^s \geq 0. \end{aligned} \tag{1.27}$$

Typically, $\mathcal{I}^s$ is taken to be the collection of those monomials $u, v \in \mathcal{I} = \mathcal{I}^1$ such that the support of u^*sv is contained in $\mathcal{I}^*\mathcal{I}$. In such a case, we call semidefinite program (1.27) the *sum of squares* relaxation of $(p, \emptyset, S)$ restricted to $V = V^1$.

Alternatively, we can consider the dual of Problem (1.26) exchanging the role of constraints and variables. That is, for each monomial w we introduce a real (or complex) scalar variable y_w and arrange them in a *moment matrix*, denoted M , with entries $M_{u,v} = y_{u^*v}$. Let $\mathcal{M}$ be the linear subspace of moment matrices indexed by a monomial basis $\mathcal{I}$ of V ,

$$\mathcal{M} = \{M \in \text{Mat}(|\mathcal{I}|) : \forall u, v \in \mathcal{I} (M_{u,v} = y_{u^*v})\}. \tag{1.28}$$

The dual semidefinite program of 1.26 reads

$$\begin{aligned} \max \quad & \sum_w p_w y_w \\ \text{s. t.} \quad & y_1 = 1, \\ & M \geq 0, \\ & M \in \mathcal{M}. \end{aligned} \tag{1.29}$$

Similarly, for each constraint $s = \sum_x s_x x \in S$ we introduce the moment matrix *localized* at s , denoted M^s , with entries $M_{u,v}^s = \sum_x s_x y_{u^* x v}$. We denote $\mathcal{M}^s$ the linear subspace of moment matrices localized at s and indexed by a monomial basis $\mathcal{I}^s$ of V^s ,

$$\mathcal{M}^s = \{M^s \in \text{Mat}(|\mathcal{I}^s|) : \forall u, v \in \mathcal{I}^s (M_{u,v}^s = \sum_x s_x y_{u^* x v})\}. \tag{1.30}$$

Then, the dual semidefinite program of 1.27 reads

$$\begin{aligned} \max \quad & \sum_w p_w y_w \\ \text{s. t.} \quad & y_1 = 1, \\ & M \geq 0, M \in \mathcal{M}, \\ & M^s \geq 0, M^s \in \mathcal{M}^s. \end{aligned} \tag{1.31}$$

This offers an alternative interpretation and intuition for polynomial optimization problems. Namely, a collection of moments y_w arranged in a positive semidefinite matrix approximate feasible evaluations of actual monomials.

In the case all polynomials are considered in the sum of squares decomposition, $V = k\langle \mathbb{M} \rangle$, Gelfand-Naimark-Segal construction (see Section 1.7) produces an explicit evaluation (λ, ψ) on a separable Hilbert space. This gives a proof of completeness for the relaxations from functional analysis complementary to the ones from real algebraic geometry using Positivstellensätze. It is important to remark that there is no available procedure to choose the subspace V efficiently for a general polynomial optimization problem. Polynomial optimization is NP-hard already for commutative variables with quadratic constraints [63]. The general approach, as proposed in [54, 28, 48, 38], is to consider the hierarchy of subspaces V_d spanned by all monomials up to degree d . For increasing degree, the optimal values of these relaxations approach from above the optimal value of the polynomial optimization problem, since $V_d \subset V_{d+1}$ and $\lim_d V_d = k\langle \mathbb{M} \rangle$.

In Chapter 2 we use Gröbner bases to efficiently deal with equality constraints. The benefits of this are twofold. Not only we do not need to introduce extra localizing variables and constraints in the problem, but we can additionally get rid of some redundant monomials reducing the size of the relaxations.

1.7 Gelfand-Naimark-Segal construction

We present the proof for the convergence [28] of the semidefinite programming relaxation 1.27 for the polynomial optimization problem $(p, \emptyset, S)$ in $k\langle X \rangle$. It

Semidefinite programming relaxations

The *sum of squares relaxation* of the polynomial optimization problem $(p, \emptyset, S)$ to the subspace V with basis $\mathcal{I}$ is the semidefinite program

$$\begin{aligned} \min \quad & t \\ \text{s. t.} \quad & (t - p)_w = \sum_{u^*v=w} Q_{u,v} + \sum_{s \in S} \sum_{u^*xv=w} s_x Q_{u,v}^s, \\ & Q, Q^s \geq 0. \end{aligned} \tag{1.32}$$

The dual *moment relaxation* is the semidefinite program

$$\begin{aligned} \max \quad & \sum_w p_w y_w \\ \text{s. t.} \quad & y_1 = 1, \\ & M \geq 0, M \in \mathcal{M}, \\ & M^s \geq 0, M^s \in \mathcal{M}^s. \end{aligned} \tag{1.33}$$

Where the support of p is contained in the linear subspace spanned by V^*V . Strong duality holds when the feasible regions have non-empty interior. When $V = k\langle \mathbf{M} \rangle$, the optimal solutions of both relaxations coincide with the optimal value of the problem $(p, \emptyset, S)$.

introduces a technique that we will use several times along the thesis, the Gelfand-Naimark-Segal (GNS) construction [42, 43]. Again, we require that the quadratic module $\mathcal{Q}(S)$ is archimedean. That is, $N - \sum_{x \in X} x^*x$ for some natural number N and set of generators X .

Assume that the moment relaxation is feasible for every monomial basis $\mathcal{I}$, and arrange the moments y_w for $w \in \mathcal{I}$ in a matrix such that $M(y)_{u,v} = y_{u^*v}$.

Claim 46. *There exists a limiting sequence of moments $\hat{y}$.*

[28, Lemma 5]

Proof. Assume $\sum_{x \in X} x^*x \leq N$. Thus $\sum_x y_{u^*x^*xu} \leq N y_{u^*u}$ for each monomial u , but all the terms in the left hand side are non-negative, thus $y_{u^*x^*xu} \leq N y_{u^*u}$ for each i . Since $y_1 = 1$, by induction it follows that $y_{u^*u} \leq N^{|u|}$, where $|u|$ indicates the degree of the monomial u . Also, the matrix of moments being positive semidefinite implies in particular that $y_{u^*u} - |y_u|^2 \geq 0$, thus $|y_u| \leq N^{|u|/2}$. Filling with 0 for all the monomials not appearing in $\mathcal{I}$, we get a sequence $y \in \ell_\infty$. And normalizing $z_u = y_u / N^{|u|/2}$ we get a sequence in $\text{Ball}(\ell_\infty)$.

By Banach-Alaoglu (see Theorem 20), $\text{Ball}(\ell_\infty)$ is $\sigma(\ell_\infty, \ell_1)$ -compact, where $\sigma(\ell_\infty, \ell_1)$ is the topology induced by the seminorms $\rho_a(x) = |a(x)|$ for each $a \in \ell_1$. In particular, the basis elements $e_n \in \ell_1$ suffice to induce the topology. Thus the topology is induced by countably many seminorms and the space is metrizable [40, Lemma 4.4.6]. Then the notion of compactness is equivalent

to sequential compactness, and there exists a subsequence of $z^{\mathcal{I}}$ converging to some $\hat{z} \in \text{Ball}(\ell_\infty)$ in the $\sigma(\ell_\infty, \ell_1)$ topology. Moreover, taking $e_n \in \ell_1$, $z_n^{\mathcal{I}} = e_n(z^{\mathcal{I}})$ converges to $\hat{z}_n = e_n(\hat{z})$ for each n . Taking $\hat{y}_u = \hat{z}_u N^{|u|/2}$, we conclude $y_{\mathcal{I}}$ converges pointwise to $\hat{y}$ in the limit $\mathcal{I}$ includes all monomials. $\square$

Let $\hat{p} = \lim_{\mathcal{I}} p_{\mathcal{I}}$ be the limiting value of the sequence of optimal values $p_{\mathcal{I}}$ restricted to $\mathcal{I}$. The next step is to construct from $\hat{y}$, a triple $(\hat{H}, \hat{\mathbf{X}}, \hat{\psi})$ attaining the value $\hat{p}$. For this, we perform GNS construction.

[28, Theorem 1] **Claim 47.** *There is a triple $(\hat{H}, \hat{\mathbf{X}}, \hat{\psi})$ attaining the value $\hat{p}$.*

Proof. Since $M(\hat{y}) \geq 0$, there are (not necessarily orthonormal) vectors $\bar{u}$ for each monomial u such that $M(\hat{y})_{u,v} = \hat{y}_{u^*v} = \langle \bar{u}, \bar{v} \rangle$. Let H be the linear space these vectors span, and define the operators $X_i \bar{u} = \bar{x}_i \bar{u}$. This is well defined, since it does not depend on the particular linear expansion of a vector, and it is bounded since $\langle X_i^2 \phi, \phi \rangle \leq N \langle \phi, \phi \rangle$ for any $\phi \in H$. Thus, X_i extends uniquely to a bounded operator $\hat{X}_i$ over $\hat{H}$, the closure of H and thus separable. Moreover, X_i is symmetric, since $\langle v, X_i u \rangle = \hat{y}_{v^* x_i u} = \hat{y}_{(x_i v)^* u} = \langle X_i v, u \rangle$ for all monomials u, v . Thus, its extension $\hat{X}_i$ is self-adjoint. Last, take $\hat{\psi} = \bar{1}$, so that $\langle \hat{\psi}, \hat{\psi} \rangle = \hat{y}_1 = 1$ and $\langle p(\hat{\mathbf{X}}) \hat{\psi}, \hat{\psi} \rangle = \hat{p}$, where $p(\hat{\mathbf{X}})$ denotes the operator representation obtained from p replacing each letter x_i by the associated operator $\hat{X}_i$. $\square$

By construction, the expectation value of $p(\hat{\mathbf{X}})$ over the state $\hat{\psi}$ is $\hat{p}$. Remarkably, GNS construction can already be performed from a finite matrix of moments, provided a *flatness* condition is met [28, Theorem 2].

[28, Theorem 2] **Lemma 48.** *Let $\mathcal{I}$ be a monomial basis and $\mathcal{I}'$ its extension with all monomials up to the highest monomial in the input polynomials (and all monomials of degree one). Let M and M' be matrices of moments indexed with $\mathcal{I}$ and $\mathcal{I}'$. If $\text{rank}(M) = \text{rank}(M')$, we can perform GNS construction.*

Proof. Let H be the Hilbert space spanned by the Gram vectors, $M_{u,v} = \langle \bar{u}, \bar{v} \rangle$. By the rank condition, we can define $\bar{x}_i \bar{u}$ inside H and thus define the operators $X_i \bar{u} = \bar{x}_i \bar{u}$. Furthermore, these operators satisfy all constraints $s = \sum_w s_w w$ in S , since $s(X) \bar{u} = \sum_v s_w w(X) \bar{u} = \sum_w s_w \bar{w} \bar{u} = \mathbf{0}$ for all $\bar{u}$ spanning H . $\square$

The rank condition certifies the semidefinite programming relaxations reach the optimal solution, and GNS construction provides an explicit optimizer. The analogue of Lemma 48 also holds for state [38, Proposition 6.10] and trace [48, Proposition 5.9] polynomial optimization, provided some extremality condition that is satisfied for solutions obtained with inner point method solvers.

Chapter 2

Special constraints

In this Chapter, based on the work [1], we introduce techniques from computational algebra to efficiently deal with equality constraints in polynomial optimization problems. First, we recall the theory of Gröbner basis. Next section introduces original results for the special case of commutativity constraints, tantamount in physical applications. We provide canonical normal forms that allow efficient algebraic manipulations of polynomials in partially commutative variables, in particular, Gröbner basis computations. Last, we review known computational reductions for semidefinite programs and polynomial optimization problems.

2.1 Gröbner bases

A polynomial optimization problem (p, G, S) in $k\langle X \rangle$

$$\begin{aligned} \max \quad & p \\ \text{s. t.} \quad & g = \mathbf{0} & \forall g \in G \\ & s \geq \mathbf{0} & \forall s \in S. \end{aligned} \tag{2.1}$$

can be translated to a polynomial optimization problem $([p], \emptyset, [S])$ over the polynomial ring $k\langle X \rangle / \langle G \rangle$ of equivalence classes in $k\langle X \rangle$ under the identifications induced by the equality constraints G [64, §4]. In particular, if we have canonical forms $\lceil p \rceil$ and $\lceil s \rceil$ for the classes of p and $s \in S$, we get the problem

$$\begin{aligned} \max \quad & \lceil p \rceil \\ \text{s. t.} \quad & \lceil s \rceil \geq \mathbf{0} & \forall s \in S. \end{aligned} \tag{2.2}$$

In this section we summarize how to identify polynomials under equality constraints and systematically obtain canonical forms. For this, we make use of the theory of string replacement rules and Gröbner bases [55, 65, 66].

We first illustrate the main idea with a well known example of non-commutative polynomial optimization: a Bell inequality.

Example 49 (Bell scenario). *Consider the hermitian alphabet*

$$X = \{A_0, A_1, B_0, B_1\}, \quad (2.3)$$

whose letters are idempotent

$$\mathbf{1} = A_i^2, \quad \mathbf{1} = B_j^2, \quad (2.4)$$

and satisfy the commutativity constraints

$$\mathbf{0} = A_i B_j - B_j A_i. \quad (2.5)$$

Letters A_i and B_j correspond with dichotomic measurements performed by two separate parties as in a Bell scenario $B(2, 2, 2)$. The second degree polynomial

$$\beta(X) = A_0 B_0 + A_0 B_1 + A_1 B_0 - A_1 B_1 \quad (2.6)$$

corresponds with Clauser-Horne-Shimony-Holt functional [12]. Tsirelson's bound for quantum correlations, $\beta \leq 2\sqrt{2}$ [27], can be recovered from a sum of squares decomposition in the space of degree one polynomials [67]. Namely,

$$2\sqrt{2} - \beta = \frac{(A_0 - B_+)^2}{\sqrt{2}} + \frac{(A_1 - B_-)^2}{\sqrt{2}} \geq \mathbf{0}, \quad (2.7)$$

where $B_+ = (B_0 + B_1)/\sqrt{2}$ and $B_- = (B_0 - B_1)/\sqrt{2}$.

Equivalently, with the moment relaxation

$$2\sqrt{2} = \max_{y_{A_0 B_0} + y_{A_0 B_1} + y_{A_1 B_0} - y_{A_1 B_1}} \quad (2.8)$$

$$\text{s. t.} \quad \begin{pmatrix} 1 & y_{A_0} & y_{A_1} & y_{B_0} & y_{B_1} \\ y_{A_0} & 1 & y_{A_0 A_1} & y_{A_0 B_0} & y_{A_0 B_1} \\ y_{A_1} & y_{A_1 A_0} & 1 & y_{A_1 B_0} & y_{A_1 B_1} \\ y_{B_0} & y_{A_0 B_0} & y_{A_1 B_0} & 1 & y_{B_0 B_1} \\ y_{B_1} & y_{A_0 B_1} & y_{A_1 B_1} & y_{B_1 B_0} & 1 \end{pmatrix} \geq \mathbf{0}.$$

Equality constraints (2.4) and (2.5) are implemented in the moment matrix through the following monomial replacements:

$$A_i A_i \mapsto \mathbf{1}, \quad B_j B_j \mapsto \mathbf{1}, \quad B_j A_i \mapsto A_i B_j. \quad (2.9)$$

This reduces the number of variables in Problem (2.8) from 21 to 16 and avoids localizing matrices, reducing also the number of constraints.

Monomial replacement rules

Given a monoid with presentation $\mathbf{M} = \langle X; \mathcal{R} \rangle$, it is important to identify free words $\langle X \rangle$ under the congruence relation $\mathcal{R}^\#$ induced by $\mathcal{R}$. This is known as the *word problem* [55, §0.1]. One possible solution is to assign canonical forms $\lceil u \rceil$ and $\lceil v \rceil$ to free words u and v in $\langle X \rangle$, so that $u \mathcal{R}^\# v$ if and only if $\lceil u \rceil = \lceil v \rceil$.

Definition 50 (Replacement rules). *The set of replacement rules on $\langle X \rangle$ [55, §2.1], [57, §2.5] induced by a relation $\mathcal{R}$ on X , also called semi-Thue system or string rewriting system, is the congruence $\{x * u * y \mapsto_{\mathcal{R}} x * v * y : x, y \in \langle X \rangle, u \mathcal{R} v\}$. Denote $\mapsto_{\mathcal{R}}^*$ the transitive closure of $\mapsto_{\mathcal{R}}$. We say $\mapsto_{\mathcal{R}}$ is Nötherian if there is no infinite sequence $(u_i)_i \subset \langle X \rangle$ such that $u_i \mapsto_{\mathcal{R}} u_{i+1}$ and confluent if $u \mapsto_{\mathcal{R}}^* v$ and $u \mapsto_{\mathcal{R}}^* w$ imply $v \mapsto_{\mathcal{R}}^* t$ and $w \mapsto_{\mathcal{R}}^* t$ for some $t \in \langle X \rangle$. Replacement rules both Nötherian and confluent are called complete, and provide unique canonical forms in $\mathbb{M} = \langle X \rangle / \mathcal{R}^\#$. That is, the replacements stabilize at some finite step and the order in which they are performed does not affect the final expression. Then, $\lceil w \rceil$ is the unique expression obtained from w after all replacements.* [55, §1.1]

The definition extends to replacement rules on general monoids [55, §1.1].

Example 51. *Consider the alphabet $X = (x, y, z)$. The symmetric relation $\mathcal{R} = \{(xy, yx), (yx, xy)\}$ induces non-Nötherian replacement rules since we have the infinite chain $xy \mapsto_{\mathcal{R}} yx \mapsto_{\mathcal{R}} xy \dots$. Replacement rules induced by $\mathcal{R}' = \{(yx, xy), (zy, yz)\}$ are Nötherian but not confluent, since $zyx \mapsto_{\mathcal{R}'} yzx$ and $zyx \mapsto_{\mathcal{R}'} zxy$, which can not be further reduced to a common monomial. Finally, relation $\mathcal{R}'' = \{(yx, xy), (yz, zy)\}$ induces complete replacement rules [57, Theorem 2.2] (see Example 54) and thus unique canonical forms for free words in $\langle X \rangle$.*

Notice that symmetric relations induce loops of replacements. To avoid this, we restrict our attention to replacement rules induced by asymmetric relations. One can systematically construct an asymmetric relation from an arbitrary $\mathcal{R}$ on $\mathbb{M}$, given a method to choose one element for each symmetric pair in $\mathcal{R}$ compatible with multiplication. A monomial ordering provides such a method.

Definition 52 (Monomial ordering). *A monomial ordering or semigroup [66, §1.1] ordering for $\mathbb{M}$ is a total order on $\mathbb{M}$ compatible with the monoid operation. That is, $x * u * y < x * v * y$ for any $x, y \in \mathbb{M}$ whenever $u < v$.*

Example 53 (Degree-Lexicographic). *Given a total order on X , the degree-lexicographic ordering on $\langle X \rangle$ is defined as $u < v$ if (i) $|u| < |v|$ or (ii) $|u| = |v|$ and $u_m < v_m$ as letters in X , where m denotes the first index for which $u_m \neq v_m$.* [66, §1.1]

A relation $\mathcal{R}$ and a monomial ordering then induce replacement rules [55, §2.2]

$$\{v \mapsto_{\mathcal{R}} u : u \mathcal{R}^\# v, u < v\}. \quad (2.10)$$

Compatibility with the monoid operation guarantees that these rules define a congruence relation. Moreover, when the monomial ordering is a *well ordering* (every subset has a minimum element), the induced replacement rules are Nötherian [55, Theorem 2.2.4]. However, these are not confluent in general.

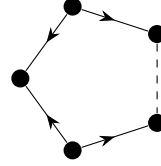
For special relations, necessary and sufficient conditions for a set of replacement rules to be complete are known. For instance, complete replacement rules for commutation relations are in correspondence with transitive orientations on the *independence graph* of the alphabet [57, Theorem 2.2]. That is, the graph whose vertices are the letters in the alphabet, that are connected whenever [57, §2.5]

they commute. An *orientation* for a graph assigns a direction to each edge, and is said to be *transitive* whenever $x \mapsto y$ and $y \mapsto z$ implies $x \mapsto z$.

[57, Definition 2.3] **Example 54** (Transitive orderings). *Take the relations $\mathcal{R}'$ and $\mathcal{R}''$ in Example 51 corresponding with the partial commutations $yx = xy$ and $zy = yz$. Relation $\mathcal{R}'$ is induced by a monomial ordering with $x < y$ and $y < z$. This gives a non transitive orientation on the dependence graph, since $z \mapsto y$ and $y \mapsto x$ but $z \not\mapsto x$. Relation $\mathcal{R}''$ is induced by a monomial ordering with $x < y$ and $z < y$ and gives a transitive orientation on the dependence graph, $y \mapsto x$ and $y \mapsto z$.*



Direct inspection reveals the smallest graph that does not admit a transitive orientation is the pentagon. Thus, there is no finite relation that induces complete replacement rules for free words up to the commutations defined by a pentagon.



The theory of Gröbner bases provide complete replacement rules, even in the more general case of constraints between polynomials.

Polynomial replacement rules

Gröbner bases, first introduced in [65] for the polynomial ring in commutative variables and later extended [66] to non-commutative variables, are special collections of polynomials that induce complete sets of replacement rules. Since then, they became a central tool in computational algebra. Among other applications, Gröbner bases induce well defined polynomial division and solve the ideal membership problem [66, Theorem 1.8]. Gröbner bases in general monoids were also considered in [66, §7] and [59, §4], and specialised for finitely generated monoids in [68]. Following [1], we present the theory of Gröbner bases in [66, §7] for general monoids and specify those results for which we make use of special properties of the non-commutative monoid.

In the same spirit of Eq. (2.10), where we used a monomial ordering on $\mathbb{M}$ to choose a replacement rule from a monomial identification, we can use the monomial ordering to choose a leading term in a general polynomial and implement similar replacement rules involving more terms.

Definition 55 (Leading term). *Any polynomial $p \in k\langle \mathbb{M} \rangle$ can be written as [66, §1.1] $p = \sum_{i=1}^n c_i m_i$, where $c_i \in k$ and $m_i \in \mathbb{M}$ in decreasing order $m_1 > \dots > m_n$. We say $\text{lc}(p) = c_1$ is the leading coefficient, $\text{lm}(p) = m_1$ is the leading monomial and $\text{lt}(p) = c_1 m_1$ is the leading term of p .*

A collection of polynomials $H \subset k\langle \mathbb{M} \rangle$ induces Nötherian replacement rules

$$\{p * \text{lt}(h) * q \mapsto_H p * (\text{lt}(h) - h) * q : h \in H, p, q \in k\langle \mathbb{M} \rangle\} \quad (2.11)$$

with respect to a monomial ordering [55, Theorem 2.2.4]. It is interesting to remark that, although monomial orderings simplify the analysis, they are not readily available for general monoids. Nonetheless, complete replacement rules can be obtained without a monomial ordering, by adding new rules to ensure that $p * h * q \mapsto_H \mathbf{0}$ for all $p, q \in \mathbb{M}$ whenever $h \mapsto_H \mathbf{0}$. This process is called *saturation* [68].

When a monomial ordering is available, the only obstacle to construct a complete set of replacement rules is confluence. To overcome this, we have to make sure $h - g \mapsto_H \mathbf{0}$ whenever $h, g \mapsto_H \mathbf{0}$. Since we implement the rules replacing leading terms, we can always reduce $h - g \mapsto_H \mathbf{0}$ when $\text{lt}(h) \neq \text{lt}(g)$. However, when $\text{lt}(h) = \text{lt}(g)$, one may need to extend the replacement rules with $\text{lt}(h - g) \mapsto_H \text{lt}(h - g) - (h - g)$, or equivalently, to extend H with the polynomial $h - g$. Algorithms to compute Gröbner bases start from a set H of generating polynomials and extend it with additional polynomials that provide all such reductions [65, 66]. Although there may be infinitely many cases to check, many of them are redundant. A sufficient collection of cases to check are given by minimal *obstructions*.

Definition 56 (Obstruction). *An obstruction, also called S -polynomial, [66, §5.3] between polynomials $f, g \in k\langle \mathbb{M} \rangle$ is an expression of the form*

$$S(f, g) = S(f, g; l, r; \lambda, \rho) = l * f * r - \lambda * g * \rho \quad (2.12)$$

with $l * \text{lt}(f) * r = \lambda * \text{lt}(g) * \rho$, which is called the joint of the obstruction and denoted $\text{lt}(f) \vee \text{lt}(g)$. We say the obstruction is minimal if there are no common left factors between l and λ or common right factors between r and ρ .

For commutative polynomials $f, g \in k[X]$ there is only one minimal obstruction, that is obtained with the least common multiple of their leading terms, $\text{lt}(f) \vee \text{lt}(g)$ [66, Theorem 3.7]. For non-commutative polynomials $f, g \in k\langle X \rangle$, there are infinitely many minimal obstructions, for instance $f * w * \text{lt}(g) - \text{lt}(f) * w * g$ for any word $w \in \langle X \rangle$. Obstructions of these form are redundant and there are finitely many minimal non-trivial obstructions to check for each pair of polynomials [66, Corollary 5.8], called *overlaps*.

Definition 57 (Overlap). *A minimal obstruction $S(f, g; l, r; \lambda, \rho)$ has overlap [66, Corollary 5.8] $w \in \mathbb{M}$ between $f, g \in k\langle \mathbb{M} \rangle$ if one of the following four scenarios happens:*

- *There are $f_0, g_0 \in \mathbb{M}$ such that $\text{lt}(f) = f_0 * w$ and $\text{lt}(g) = w * g_0$.*

- There are $f_0, g_0 \in \mathbb{M}$ such that $\text{lt}(f) = w * f_0$ and $\text{lt}(g) = g_0 * w$.
- There are $f_0, f_1 \in \mathbb{M}$ such that $\text{lt}(f) = f_0 * w * f_1$ and $\text{lt}(g) = w$.
- There are $g_0, g_1 \in \mathbb{M}$ such that $\text{lt}(f) = w$ and $\text{lt}(g) = g_0 * w * g_1$.

The meet of the obstruction is then $\text{lt}(f) \wedge \text{lt}(g) = w$. Otherwise, we say the obstruction has no overlap and set $\text{lt}(f) \wedge \text{lt}(g) = \mathbf{1}$. In such a case, either $S = f * w * \text{lt}(g) - \text{lt}(f) * w * g$ or $S = \text{lt}(g) * w * f - g * w * \text{lt}(f)$ for some $w \in \mathbb{M}$.

The sufficiency of overlaps for Gröbner bases computation was first proved by Buchberger in the commutative case [66, Theorem 3.7] and later extended to the non-commutative case [66, Theorem 5.9]. We present a proof that extends to particular classes of polynomials in general monoids [1].

[66, §5.3] **Definition 58** (Weak Gröbner representation). A weak Gröbner representation of a polynomial $f \in k\langle \mathbb{M} \rangle$ with respect to the collection $H \subset k\langle \mathbb{M} \rangle$ is a decomposition of the form

$$f = \sum_{i=1}^n c_i u_i * h_i * v_i, \quad (2.13)$$

for some coefficients $c_i \in k$, monomials $u_i, v_i \in \mathbb{M}$ and polynomials $h_i \in H$ such that $\text{lt}(f) \geq \text{lt}(u_i * h_i * v_i)$ for each $i = 1, \dots, n$. That is, there is no cancellation between the highest leading terms of polynomials in the decomposition.

[66, §1.1] **Definition 59** (Gröbner basis). A collection of polynomials $H \subset k\langle \mathbb{M} \rangle$ is a Gröbner basis for the ideal $\langle H \rangle$ if it verifies any (all) of the following equivalent conditions:

- [66, Definition 1.7] • Leading terms of H generate leading terms in $\langle H \rangle$, $\langle \text{lt}(H) \rangle = \text{lt}(\langle H \rangle)$.
- [66, Theorem 1.8] • It solves the ideal membership problem, $h \mapsto_H \mathbf{0}$ if and only if $h \in \langle H \rangle$.
- [66, Theorem 1.8] • All obstructions in H [Eq. (2.12)] have a weak Gröbner representation.

[66, Proposition 1.5] A Gröbner basis H is reduced (unique) if every polynomial $h \in H$ is monic, $\text{lc}(h) = 1$, and can not be reduced with respect to other elements, $h \mapsto_{H - \{h\}} h$.

Obstructions with no overlap are redundant in $k\langle X \rangle$ [66, Theorem 5.9].

[66, Corollary 5.8] **Proposition 60.** An obstruction with no overlap in $k\langle X \rangle$ admits a weak Gröbner representation in terms of other obstructions with overlap.

We use a structural characterization for overlaps in $\langle X \rangle$ later in the proof to inductively produce a weak Gröbner representation for obstructions with no overlap using obstructions with overlap. However, the first part of the proof holds in any monoid $\mathbb{M}$. This proof is adapted from [69, Lemma 1.3].

Proof. Take two polynomials $f = \sum_k c_k f_k$ and $g = \sum_k d_k g_k$ in $k\langle \mathbb{M} \rangle$, where the terms are ordered decreasingly. Without loss of generality, consider the obstruction $S(f, g; l, r; \lambda, \rho)$ with $l = \mathbf{1}$, $r = w * \text{lt}(g)$, $\lambda = f_1 * w$ and $\rho = \mathbf{1}$ for some $w > \mathbf{1}$,

$$S = f * w * g_1 - f_1 * w * g. \quad (2.14)$$

Using the expansions for f and g we can write

$$S = \sum_{k>1} c_k f_k * w * g_1 - \sum_{k>1} d_k f_1 * w * g_k \quad (2.15)$$

$$= \sum_{k>1} c_k f_k * w * g - \sum_{k>1} d_k f * w * g_k. \quad (2.16)$$

This gives a weak representation unless the leading terms in the sums in Eq. (2.16) cancel each other. That is, $c_2 f_2 * w * g_1 = d_2 f_1 * w * g_2$. This is only possible if $c_2 = d_2$, and $f_1 = f_2 * u$ and $g_1 = v * g_2$ such that $u * w = w * v$. When f and g have at most two terms (i.e. monomial replacement rules), this implies the obstruction is trivial $S = \mathbf{0}$. However, for more general polynomial replacement rules we make use of a structural property of $\langle X \rangle$.

Lemma 61. *For free words $u, v, w \in \langle X \rangle$, the following are equivalent: (i) $uw = vw$ and (ii) $u = xy$, $v = yx$ and $w = (xy)^a x$ for some $x, y \in \langle X \rangle$ and $a \in \mathbb{N}$.* [57, Proposition 3.1]

Using Lemma 61, we conclude that the obstruction must have the form $w = (x * y)^a * x$ for some $a > 0$ and $y \neq \mathbf{1}$ (otherwise there is overlap). Introducing this in Eq. (2.16) we find

$$S = c_2(f_1 * (xy)^{a-1} * x * g - f * (xy)^{a-1} * x * g_1) + \sum_{k>2} c_k f_k * w * g - \sum_{k>2} d_k f * w * g_k, \quad (2.17)$$

where $f_1 * (xy)^{a-1} * x * g - f * (xy)^{a-1} * x * g_1 =: S(a-1)$ is again an obstruction with $w = (xy)^{a-1} * x$. If the next leading terms in the summation cancel, $c_3 = d_3$ and Lemma 61 again implies $f_1 = f_3 * (xy)^b$ and $g_1 = (yx)^b * g_3$ for some $b > 1$. We repeat the argument until the leading terms do not cancel,

$$S = c_2 S(a-1) + c_3 S(a-b) + \dots + \sum_{k \geq m} c_k f_k * w * g - \sum_{k \geq m} d_k f * w * g_k, \quad (2.18)$$

where the leading term of the last summations is $f_m * w * g_1$ or $f_1 * w * g_m$. If we show (i) the leading term of each $S(a-b)$ is smaller than these two and (ii) each $S(a-b)$ admits a weak Gröbner decomposition, then Eq. (2.18) gives a weak Gröbner representation for S , finishing the proof.

To see (i) the bound on the leading terms of all such $S(a-b)$ in Eq. (2.18), notice that by minimality of k , for each $i < k$ we have the cancellation $f_i * (xy)^a * x * g_1 = f_1 * (xy)^a * x * g_i$, which is equivalent to the cancellation $f_i * (xy)^{a-b} * x * g_1 = f_1 * (xy)^{a-b} * x * g_i$ for each $1 < b \leq a$. Thus,

$$\text{lt}(S(a-b)) \leq f_1 * (xy)^{a-b} * x * g_m < f_1 * w * g_m, \quad (2.19)$$

$$\text{lt}(S(a-b)) \leq f_m * (xy)^{a-b} * x * g_1 < f_m * w * g_1, \quad (2.20)$$

which implies $\text{lt}(S(a-b)) < \text{lt}(S)$.

To see (ii) that each $S(a-b)$ admits a weak Gröbner representation, notice that we have reduced the existence of a weak Gröbner representation for $S(a)$ to the existence of such a representation for each $S(a-b)$ with $1 < b \leq a$. Thus, the problem inductively reduces to the existence of such a representation for $S(0) = f * x * g_1 - f_1 * x * g$. The decomposition from Lemma 61 shows that $f_1 = f_2 * x * y$ and $g_1 = y * x * g_2$ have overlap $y > 1$ when $w = x$. Thus, weak representations for obstructions with overlap produce weak representations for all obstructions. $\square$

Notice that the argument in Proposition 60 works in general monoids for polynomials with two terms. That is, $f = f_1 + c_2 f_2$ and $g = g_1 + d_2 g_2$, which induce monomial identifications up to scalar factors. However, it is an open question if for arbitrary polynomials in the partially commuting monoid all obstructions without overlap are redundant.

Complete replacement rules

Given an equivalence relation on $k\langle\mathbb{M}\rangle$, we aim to construct a complete collection of replacement rules in $k\langle\mathbb{M}\rangle$ that eventually identify all equivalent elements. For this, we need:

- The replacements finish after finitely many steps. This is the obstacle to having a Nötherian relation.
- A monomial multiple of a polynomial that reduces to zero should also reduce to zero: $p * h * q \mapsto_H \mathbf{0}$ when $h \mapsto_H \mathbf{0}$. This is the obstacle to having a congruence relation.
- Linear combinations of polynomials that reduce to zero should also reduce to zero: $h + g \mapsto_H \mathbf{0}$ when $h, g \mapsto_H \mathbf{0}$. This is the obstacle to having a confluent relation.

A well ordering provides the first condition, a monomial ordering provides the second condition and a Gröbner basis provides the third condition. However, different approaches to construct complete replacement rules and canonical forms exist.

[66, Theorem 1.8] **Proposition 62** (Buchberger). *A collection of polynomials $G \subset k\langle\mathbb{M}\rangle$ induces a complete set of replacement rules with respect to a monomial well ordering if and only if all obstructions for G reduce to zero.*

Proof. We only need to show that a non-zero $g \in \langle G \rangle$ can be reduced to zero if all non-trivial obstructions can be reduced to zero. Start with a representation $g = \sum_i \alpha_i l_i * g_i * r_i$, where $g_i \in G$ and $l_i, r_i \in \mathbb{M}$. Let $t = \max\{l_i * \text{lt}(g_i) * r_i\}$ and n the number of times t appears as a term in some $l_i * g_i * r_i$. We show

by induction on (t, n) that there is a representation of g for which $\text{lt}(g) = t$, showing that g is reducible. If $\text{lt}(g) = t$, we are done, otherwise there are $f, h \in G$ such that $\alpha_f l_f * \text{lt}(f) * r_f = t = \alpha_h l_h * \text{lt}(h) * r_h$. Take the obstruction

$$S(f, h) = \alpha_f l_f * \text{lt}(f) * r_f - \alpha_h l_h * \text{lt}(h) * r_h, \quad (2.21)$$

where $\text{lt}(S(f, h)) < t$. By assumption, there exists a representation $S(f, h) = \sum_k \gamma_k u_k * g_k * v_k$, with $g_k \in G$ and $u_k, v_k \in \mathbb{M}$ such that $\text{lt}(u_k * g_k * v_k) \leq \text{lt}(S(f, h))$ for all k . In the decomposition of g we then replace $\alpha_f l_f * f * r_f + \alpha_h l_h * h * r_h$ by

$$\begin{aligned} &= 2\alpha_f l_f * f * r_f - S(f, h) \\ &= 2\alpha_f l_f * f * r_f - \sum_k c_k u_k * g_k * v_k, \end{aligned} \quad (2.22)$$

where $\text{lt}(u_k * g_k * v_k) \leq \text{lt}(S(f, h)) < t$. This reduces the number n of terms containing t in the decomposition of g , which finishes the proof given a well ordering. $\square$

This gives the procedure to complete a collection of polynomials into a Gröbner basis, equivalently a complete set of replacement rules, described in Algorithm 1. For commutative polynomials, $\mathbb{M} = [X]$, there is only one minimal obstruction between each pair. For non-commutative polynomials, $\mathbb{M} = \langle X \rangle$, there are finitely many minimal obstructions with overlap between each pair. For general monoids identifying non-trivial obstructions is more challenging. In particular, for a partially commutative monoid, $\mathbb{M} = \langle X; \mathcal{C} \rangle$, there may be infinitely many obstructions with overlap between each pair. Nonetheless, Gröbner bases can still be obtained at the cost of performing redundant computations. Next, we discuss some procedures that consider and reduce multiple obstructions at the same time.

Gröbner basis computation

This algorithm is adapted from [66, §5.4] using Proposition 62.

Algorithm 1 GRÖBNER(G)

- 1: $P = \{(g, g') \in G \times G : \text{lt}(g) \wedge \text{lt}(g') \neq \mathbf{1}\}$
 - 2: **while** $P \neq \emptyset$ **do**
 - 3: Select and remove $(f, h) \in P$
 - 4: $S(f, h) \mapsto_G g$
 - 5: **if** $g \neq 0$ **then**
 - 6: $G = G \cup \{g\}$
 - 7: $P = P \cup \{(f, g) : f \in G, \text{lt}(f) \wedge \text{lt}(g) \neq \mathbf{1}\}$
 - 8: **end if**
 - 9: **end while**
-

Macaulay Matrices

Two of the most efficient algorithms to perform Gröbner basis computations were proposed by Faugère in [70, 71], following an earlier construction based on matrix triangularization [72, 73]. These algorithms explore several obstructions simultaneously using linear algebra arguments. This drastically reduces the cost of implementation. This section presents an approach to Gröbner bases computations based on Macaulay matrices adapted to polynomial rings on general monoids, that does not require to characterize all obstructions.

[74, §2.3] **Definition 63** (Macaulay matrix). *Given a finite list F of polynomials and an ordered list of terms T , which we take to be the union of the supports of elements in F , the Macaulay or Sylvester matrix $A = \text{Mac}(F, T)$ has as element A_{ij} the coefficient of the term t_j in the polynomial f_i (possibly zero).*

[74, §2.3] **Definition 64** (Reduced row echelon form). *A $n \times m$ matrix A is in reduced row echelon form if for each non-zero row A_i with $i \leq n$, we have (i) $A_{i,s(i)} = 1$ and (ii) $A_{j,s(i)} = 0$ for each $i \neq j \leq n$, where $s(i)$ index the first non-zero entry in A_i . A matrix A can be brought to (unique up to row permutations) reduced row echelon form with Gauss elimination, which we denote $\text{rref}(A)$.*

Clearly, A and $\text{rref}(A)$ span the same row space. For a Macaulay matrix $A = \text{Mac}(F, T)$, this row space is the linear span of the polynomials in F . The leading terms of the polynomials in $\text{rref}(A)$ are linearly independent, a feature shared with reduced Gröbner bases. Moreover, when F contains enough monomial multiples of a finite list of polynomials G , also called *shifts*, then $\text{rref}(A)$ is a Gröbner basis of $\langle G \rangle$ [75, §2.3]. In particular, this is the case if F spans all eventual obstructions in Equation (2.12). In the commutative case, there are general bounds (non-optimal) on how big F needs to be in order to obtain a Gröbner basis [75, Theorem 2.3.6] and [76, Corollary 5.4].

In the non-commutative case, Gröbner bases are not always finite [66, §6]. However, introducing in F all shifts up to some fixed degree d , the algorithm would output a truncated Gröbner basis in the spirit of the non-commutative version of Buchberger's algorithm [66]. Thus, we can readily implement Algorithm 2 to obtain truncated Gröbner bases in partially commuting variables without considering obstructions, by taking shifts

$$F = \{l * g * r : (g, l, r) \in \langle X \rangle \times G \times \langle X \rangle, \deg(l * g * r) \leq d\}. \quad (2.23)$$

Moreover, exploiting criteria to discard triples that do not produce obstructions significantly reduces the size of the Macaulay matrix and hence the cost of the Gröbner basis computation.

The algorithm F4 [70] iterates this procedure for selected subsets of obstructions, enlarging the list of polynomials with the row space of $\text{rref}(A)$ at each time, speeding up the computation. In particular, when only one pair of obstructions is considered at a time, one recovers Buchberger's algorithm. Further criteria to reduce the number of obstructions are introduced in F5 algorithm [71]. These can be adapted to non-commutative [77] and other monoids for which obstructions can easily be characterized, see Algorithms 2 and 3.

Gröbner basis computation with Macaulay matrices

Adapted Buchberger algorithm from Macaulay matrices [75, §2.3].

Algorithm 2 Gröbner basis from Macaulay matrices

Input: G finite list polynomials

Output: H Gröbner basis $\langle H \rangle = \langle G \rangle$

- 1: F big enough list of *shifts* of G
 - 2: T union support elements of F
 - 3: $A = \text{rref}(\text{Mac}(F, T))$
 - 4: H polynomials spanning row space A
-

Adapted iterated algorithm as in [70, Algorithm F4].

Algorithm 3 Iterated Gröbner basis from Macaulay matrices

Input: G finite list polynomials

Output: H Gröbner basis $\langle H \rangle = \langle G \rangle$

- 1: $H = \emptyset$
 - 2: P list of all obstructions
 - 3: **while** $P \neq \emptyset$ **do**
 - 4: F selected list of obstructions removed from P
 - 5: T union support elements of F
 - 6: $A = \text{rref}(\text{Mac}(F, T))$
 - 7: Add to H polynomials from the rows of A if new
 - 8: Update P with obstructions from those polynomials
 - 9: **end while**
-

The advantage of the approach with Macaulay matrices is that computations can readily be automatized for general monoids without characterizing minimal obstructions. In particular, localizing matrices for equality constraints G in polynomial optimization problems (see Section 1.6) contain shifts of the constraints. Indeed, the entries of a localizing moment matrix $M^g \in \mathcal{M}^g$ with monomial basis $\mathcal{I}^g$ correspond with the shifts $F_g = \{u^*gw : u, v \in \mathcal{I}^g\}$, whose moments are constrained to be zero. Taking $F = \bigcup_{g \in G} F_g$ and T the union of the supports of its elements, Algorithm 2 produces replacement rules implementing all such constraints through Eq. (2.11), avoiding localizing matrices and reducing the number of variables in the semidefinite program [1].

2.2 Commutation relations

Consider a polynomial optimization problem $(p, I \cup G, S)$ on $k\langle X \rangle$, where I is a collection of commutation relations. Namely,

$$\begin{aligned} \max \quad & p \\ \text{s.t.} \quad & g = \mathbf{0} & \forall g \in G, \\ & s \geq \mathbf{0} & \forall s \in S, \\ & x_i x_j - x_j x_i = \mathbf{0} & \forall (x_i x_j - x_j x_i) \in I. \end{aligned} \tag{2.24}$$

We can formulate problem (2.24) on the quotient ring $k\langle X \rangle / \langle I \rangle$, which we denote $(\ulcorner p \urcorner, \ulcorner G \urcorner, \ulcorner S \urcorner)$ and may reduce the computational cost. This reads

$$\begin{aligned} \max \quad & \ulcorner p \urcorner, \\ \text{s.t.} \quad & \ulcorner g \urcorner = \mathbf{0} & \forall g \in G, \\ & \ulcorner s \urcorner \geq \mathbf{0} & \forall s \in S. \end{aligned} \tag{2.25}$$

In this section, we present canonical forms for polynomials in partially commuting letters and develop a theory of Gröbner bases on $k\langle X \rangle / \langle I \rangle$ to efficiently implement the constraints $\ulcorner G \urcorner$ as replacement rules over canonical forms [1].

[55, §3.6], [57, §1] Equivalence classes of free words in $\langle X \rangle$ with respect to partial commutations, sometimes called *traces*, were studied in [78, 79] and found several applications in computer science and different areas of mathematics [80, 57]. Exploiting the combinatorial properties of traces can lead to significant computational speed ups, in particular for Gröbner basis computations. When all the variables commute, we recover the free commutative monoid $[X]$, whose words can be efficiently described with the *exponents* of the letters, that is, the number of times each letter appears in the word. The monoid operation translates into addition of the exponents. Moreover, Gröbner bases in $k[X]$ are always finite [66, §3.4] and there are more criteria to neglect redundant obstructions [66, §3.3], thus the implementation is more efficient than for the non-commutative case. In this section we provide efficient representations of partially commutative words and Gröbner bases computations that interpolate between the non-commutative $k\langle X \rangle$ and commutative $k[X]$ cases [1].

[57, §2.1] A relation $\mathcal{C} = \{(x_i x_j, x_j x_i) : (x_i, x_j) \in I\}$ on $\langle X \rangle$ of partial commutations determined by a so-called *independence set* I gives a presentation of the partially commutative monoid, $\langle X; \mathcal{C} \rangle$. We denote $[x]$ the equivalence class in $\langle X; \mathcal{C} \rangle$ of the free word $x \in \langle X \rangle$ with respect to the commutation relations. A *clique* is a subset $C \subset X$ of non-commuting letters, we say it is maximal when it is with respect to inclusion. A word $w \in \langle X; \mathcal{C} \rangle$ is fully specified by its images under the projections to their maximal cliques [57, Proposition 2.1]. That is, under the $*$ -homomorphism $\pi_C : \langle X \rangle \rightarrow \langle C \rangle$ defined by $\pi_C(x) = x$ if $x \in C$ and $\pi_C(x) = \mathbf{1}$ otherwise. Conversely, a collection of *local* words $w_C \in \langle C \rangle$ on each maximal clique corresponds to a partially commutative word $w \in \langle X; \mathcal{C} \rangle$ whenever (i) the local words are compatible on the intersections and (ii) there are no directed cycles in the graph defined by the letters with edges given by non

commuting pairs, in the order they appear in the word [81, Proposition 1.4]. Collections of local words satisfying these conditions are called *reconstructible* [81, §1], and the procedure to *reconstruct* the word in $\langle X; \mathcal{C} \rangle$ is described in the following lemma.

Lemma 65. *Given a reconstructible collection of local words $w_C \in \langle C \rangle$ on each maximal clique C , there exists a partially commutative word $w \in \langle X; \mathcal{C} \rangle$ that recovers the local data. Namely, $\pi_C(w) = w_C$ for each clique C .* [81, Proposition 1.4]

Proof. We create a graph G with one node per letter in w and label them with the corresponding letter, with a suitable numbering to distinguish multiple appearances of the same letter. That is, if $a \in X$ appears $n = |w|_a$ times, there will be nodes $a_1, \dots, a_n$. Then, iterate over all letters a_i appearing in w and connect the node a_i with the nodes of all b_j appearing after a_i in w such that $(a, b) \notin \mathcal{C}$. Finally, remove redundant edges in G , that is, if $\{(u_i, u_{i+1}) : i = 1, \dots, k\}$ is a path in G , remove the edge (u_1, u_{k+1}) . This step is well defined provided there are no directed cycles in G .

Let G_0 be the set of nodes with no incoming edge and G_i the set of those for which the largest incoming path has length i . By construction, each level contains only commuting letters, otherwise there is an edge between some letters and they can not be in the same level. Thus it is well defined to write $w_i = \prod \{a \in G_i\}$. Then, the concatenation of all these blocks, $w = w_0 \dots w_k$, is the reconstructed word. $\square$

When each level w_i is ordered, the decomposition in Lemma 65 gives a unique normal form. In particular, for the lexicographical ordering we obtain the Foata normal form [57, §2.3].

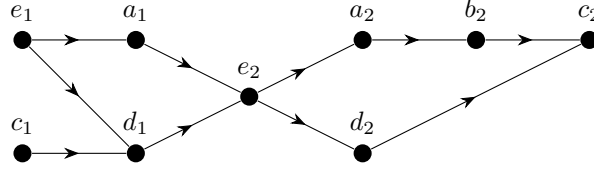
Definition 66 (Foata normal form). *The Foata normal form of $\mathbf{1} \neq w \in \langle X; \mathcal{C} \rangle$ is the unique $x_1 \dots x_n \in \langle X \rangle$ such that (i) $[x_1 \dots x_n] = w$, (ii) each x_i is the product of independent letters minimal with respect to lexicographical ordering and (iii) for each letter a in x_{i+1} there is some letter b in x_i not commuting with a .* [57, Proposition 2.5]

Example 67 (Non reconstructible word). *Take the graph with maximal cliques $(\{a, b\}, \{b, c\}, \{c, d\}, \{a, d\})$. The collection of local words (ab, bc, cd, da) is compatible on the intersections, but there is no word with those projections because there is a directed cycle, $a_1 \rightarrow b_1 \rightarrow c_1 \rightarrow d_1 \rightarrow a_1$. Interestingly, these local words correspond with the equivalence class of the free word $abcd$ up to cyclic identification. Such identifications appear in trace polynomial optimization.*

Example 68 (Reconstructible word). *Let $(\{a, b\}, \{b, c\}, \{c, d\}, \{d, e\}, \{a, e\})$ be the maximal cliques in the pentagon graph. It is immediate to check that the local words*

$$(aab, cbc, cddc, eded, eaea) \tag{2.26}$$

are compatible on intersections. The procedure described in Lemma 65 produces



from where we reconstruct the word $w = ecadeadb c$. The level decomposition is $(ce)(ad)(e)(ad)(b)(c)$, which gives the Foata normal form.

Algebraic computations between partially commutative words can be performed from their reconstructible local words. Multiplication amounts to local concatenations and involution amounts to local involutions [1]. Moreover, Gröbner basis computation can be performed in $k\langle X; \mathcal{C} \rangle$ with respect to the monomial well ordering below, introduced in [1].

Definition 69 (Degree-clique ordering). *We define the degree-clique ordering in $k\langle X; \mathcal{C} \rangle$ as $w < v$ if either (i) $|w| < |v|$ or (ii) $|w| = |v|$ and $w_{C^*} < v_{C^*}$ as monomials in $\langle C^* \rangle$, where $C^* = \min\{C : w_C \neq v_C\}$ with respect to a fixed numeration of the maximal cliques.*

Claim 70. *The degree-clique ordering is a monotone total well ordering.*

Proof. The degrees are well ordered (as natural numbers) and there are finitely many words of given degree, it is enough to show that given two different words $w \neq v$ with the same degree, one is smaller than the other. This follows from the fact that $w = v$ if and only if $w_C = v_C$ for each maximal clique C , and that each local monoid $\langle C \rangle$ has a monotone total well order. This also implies that $x * w * y < x * v * y$ for any $x, y \in \langle X; \mathcal{C} \rangle$ whenever $w < v$. $\square$

The last ingredient to obtain a Gröbner basis is to identify the minimal obstructions (see Section 2.1). That is, a minimal collection of expressions $S(f, g; l, r; \lambda, \rho) = l * f * r - \lambda * g * \rho$ with $l * \text{lt}(f) * r = \lambda * \text{lt}(g) * \rho$ to check. The main tool for this purpose is an elementary result in partially commutative monoids, known as *Levi's lemma* [57, Proposition 2.2].

[57, Proposition 2.2] **Lemma 71.** *Let $t, u, v, w \in \langle X; \mathcal{C} \rangle$, then $t * u = v * w$ if and only if $t = p * r$, $u = s * q$, $v = p * s$ and $w = r * q$ for some $p, q, r, s \in \langle X; \mathcal{C} \rangle$ such that r and s are independent (they commute and do not share common letters).*

Proof. Induction on the number of letters. $\square$

This reduces the obstructions we have to consider.

Claim 72. *Let $f, g \in k\langle X; \mathcal{C} \rangle$ with the obstruction $S(f, g) = l * f * r - \lambda * g * \rho$. Without loss of generality, we can assume l and λ are independent and also r and ρ are independent.*

Proof. If S is an obstruction, $l * \text{lt}(f) * r = \lambda * \text{lt}(g) * \rho$. By Lemma 71, $l = p * s$ and $\lambda = p * r$ for some p, r and s , the last two independent. There is no loss of generality in taking $p = \mathbf{1}$, since it is a common left factor, thus $l = r$ and $\lambda = s$ are independent. Same argument shows r and ρ are independent. $\square$

For constraints with only two monomials, this already reduces the minimal obstructions to overlaps [1], as in the non-commutative case.

Lemma 73. *All non-overlapping obstructions between polynomials with at most two terms reduce to zero.*

Proof. Let $f = f_1 + f_2$ and $g = g_1 + g_2$ be polynomials with only two terms and consider the obstruction $S = f * w * g_1 - f_1 * w * g = f_2 * w * g_1 - f_1 * w * g_2$. This can be written as $S = f_2 * w * g - f * w * g_2$. If the (leading) terms do not cancel, this gives a weak Gröbner representation and S reduces to zero. If they cancel the obstruction is trivial, $S = \mathbf{0}$, since there are no more terms. $\square$

Polynomials with two monomials correspond with monomial replacements up to scalar factors, which already cover a wide class of ubiquitous constraints. For instance, *idempotency*, $x^2 = x$, *unipotency*, $x^2 = \mathbf{1}$, *unitarity*, $xx^* = \mathbf{1}$, and *conjugation*, $xy = zx$. In particular, problems discussed in Chapters 3 and 4 only involve this kind of constraints. We illustrate with some examples how to apply the techniques developed in this chapter.

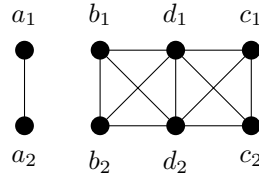
Examples

The first example of a polynomial optimization problem in partially commuting letters appears in Bell scenarios, where local operators commute. These scenarios are discussed in more detail in Section 3.1.

Example 74. *Take the hermitian alphabet*

$$X = \{a_1, a_2\} \cup \{b_1, b_2\} \cup \{c_1, c_2\} \cup \{d_1, d_2\}, \quad (2.27)$$

with partially commuting letters, according to the non-commutation graph:



Consider the collections of polynomials running over $i, j \in \{1, 2\}$,

$$\begin{aligned} G_1 = \{ & a_i a_i - a_i, \\ & b_i b_i - b_i, \\ & c_i c_i - c_i, \\ & d_i d_i - d_i \}, \end{aligned} \quad \begin{aligned} G_2 = \{ & a_i b_j - b_j a_i, \\ & a_i c_j - c_j a_i, \\ & a_i d_j - d_j a_i, \\ & b_i c_j - c_j b_i \}. \end{aligned} \quad (2.28)$$

Polynomials $G = G_1 \cup G_2$ describe the local projective effects in a Bell scenario with overlapping measurements, introduced in Sections 2.3 (see Example 79) and 3.3. The non-commutative Gröbner basis of G with respect to the degree-lexicographical ordering induced by a non-transitive ordering $d < a < b < c$ is infinite. However, it follows from Lemma 73 that G_1 is already a Gröbner basis for G in the partially commuting setting for any ordering on the letters.

Another example of polynomial optimization problem in partially commuting letters appears when considering permutations, that commute whenever their supports are disjoint. This is the case, for instance, in the quantum max-cut Hamiltonian [82, 83] discussed in Section 3.6 (see Example 109).

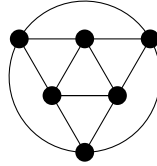
Example 75. Take the hermitian alphabet $X = \{x_{ij} : 1 \leq i < j \leq n\}$ with lexicographical ordering on the indices. Consider the following collections of polynomials over all choices of pairwise distinct indices $i, j, k, l \in [n]^1$,

$$G_1 = \{x_{ij}x_{ij} - \mathbf{1} : i, j \in [n]\}, \quad (2.29)$$

$$G_2 = \{x_{ij}x_{jk} - x_{ik}x_{ij} : i, j, k \in [n]\}, \quad (2.30)$$

$$G_3 = \{x_{ij}x_{kl} - x_{kl}x_{ij} : i, j, k, l \in [n]\}. \quad (2.31)$$

The polynomials $G = G_1 \cup G_2 \cup G_3$ characterize the algebra of transpositions of n elements, whose equivalence classes give the symmetric group $\mathcal{S}_n$. For $n = 4$, there are $|X| = 6$ letters (transpositions) with commutation graph:



As non-commutative polynomials, the reduced Gröbner basis of G with respect to the degree-lexicographical ordering contains 25 polynomials. In the partially commutative setting, it follows from Lemma 73 that $G_1 \cup G_2$ is already a Gröbner basis, which contains 18 polynomials. Table 2.1 compares the sizes of the non-commutative and partially commutative Gröbner bases for increasing n .

Finally, Gröbner bases in the partially commutative setting can be computed in parallel for collections of letters with the same commutation rules, reducing the number of obstructions and simplifying the algorithm. We present this result in the more general framework of graph products.

Graph products

Graph products [84, 56] extend the idea of independence of letters to monoids.

¹Given a natural number n , we denote $[n] = \{1, \dots, n\}$.

n	$ X $	$ G $	$ \text{NC} $	$ \text{PC} $	$ \mathcal{S}_n $
3	3	6	7	6	6
4	6	21	25	18	24
5	10	55	65	40	120
6	15	120	140	75	720
7	21	231	266	126	5040
8	28	406	462	196	40320

Table 2.1: **Gröbner bases** for the polynomials G characterizing the algebra of transpositions of n elements, with generators X . We compare the size of the non-commutative (NC) and partially commutative (PC) implementations. Last, $|\mathcal{S}_n| = n!$ is the number of inequivalent free words in the transpositions.

Definition 76 (Graph product). *Given a collection of monoids $(\mathbb{M}_v)_{v \in V}$ with presentation $\mathbb{M}_v = \langle X_v; R_v \rangle$ and an independence set I of edges between them, their graph product is the monoid with presentation $\mathbb{M}_{V,I} = \langle X; R \rangle$ where $X = \sqcup_v X_v$ and $R = \sqcup_v R_v \sqcup R_1 \sqcup R_I$. The relation $R_1 = \{\mathbf{1}_v = \mathbf{1}\}$ identifies the identities in each vertex with the identity of the graph product and the relation $R_I = \{(x_i * x_j, x_j * x_i) : x_i \in M_i, x_j \in M_j, (M_i, M_j) \in I\}$ identifies monomials that can be obtained transposing juxtaposed letters of independent monoids.* [84, §2], [56, §2.1]

In particular, when $\mathbb{M}_v = \langle v \rangle$ is the free monoid on one letter, the graph product with independence set I gives the partially commutative monoid $\langle X; \mathcal{C} \rangle$ with independence set I . When each vertex is a general monoid, it follows from Lemma 73 that Gröbner basis computation can be performed locally on each vertex monoid for monomial replacements [1].

Proposition 77. *Let $M_{V,I}$ be the graph product of monoids $(\mathbb{M}_v)_{v \in V}$, where $\mathbb{M}_v = \langle X_v; R_v \rangle$, and let $F_v \subset k(\mathbb{M}_v)$ be a collection of two terms polynomials with Gröbner basis G_v for each vertex. Then, the Gröbner basis of $F = \sqcup_v F_v$ in the graph product $\mathbb{M}_{V,I}$ with partially commutative letters is the union of the local Gröbner bases, $G = \sqcup_v G_v$.*

Proof. Since F is a collection of polynomials with two terms, Lemma (73) reduces the minimal obstructions to overlaps. Polynomials acting on different vertices do not share any letter, so there are no overlaps between them. $\square$

This means that the representation for partially commutative words already deals with all obstructions that appear between the Gröbner bases and the commutation relations.

Example 78 (Pentagon graph). *Consider the pentagon graph product $\mathbb{M}_{V,I}$, where each vertex is a free monoid $\langle X_v \rangle$ and the collection of two terms polynomials $F_v = \{x_v^2 - x_v : x_v \in X_v\}$. First, notice that F_v is a Gröbner basis, since the non-redundant obstructions, $S = x_v^2 * x_v - x_v * x_v^2 = \mathbf{0}$, are trivial. Thus,*

$F = \sqcup_v F_v$ is a Gröbner basis in the graph product $\mathbb{M}_{V,I}$ with the representation as partially commutative words. However, additional constraints are required in the non-commutative setting. In particular, when $|X_v| > 1$, there exists no finite Gröbner basis in the non-commutative setting. Already the commutation relations require an infinite Gröbner basis, since there exist no transitive orientation for the pentagon graph (recall Example 54).

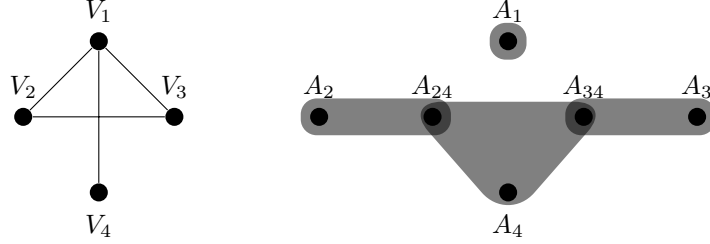
Commutations and local operations

The simplest attempt to capture a notion of independence in algebraic terms may be through commutativity. Commuting objects are independent in the sense that the order in which they appear with respect to each other is not relevant. In the context of quantum theory, the functional calculus for normal operators guarantees that commuting self-adjoint operators can be jointly diagonalized (see [41, §IX.2] or [85, Theorem 5.21]), thus represent *compatible* quantum operations that can be simultaneously performed. A different notion of independence amounts to a factorization of the operator algebra into *local algebras*, $\mathcal{B}(H) = \otimes_s \mathcal{B}(H_s)$. While local operators commute, the converse is not true. Remarkably, there exist commuting operators that can not be arbitrarily approximated with tensor factors [86, 87]. However, for finite dimensional Hilbert spaces both paradigms are equivalent [88, Theorem 1].

Given an alphabet X with some commutation relations, we can divide it into subsets of letters that reproduce the commutation relations. First, group letters with the same commutation relations into subsets so that $X = V_1 \sqcup \dots \sqcup V_k$ and let $V = (V_1, \dots, V_k)$. We obtain the *commutativity* graph, $G = (V, E)$, connecting sets of commuting letters. Then, consider the local algebra $A = \otimes_I A_I$, where I ranges over all *independent* sets in G , that is, totally disconnected subsets of vertices. The letters in V_i act locally over $A(i) = \otimes_{I(i)} A_{I(i)}$, where $I(i)$ ranges over all independent sets in G containing V_i .

Example 79. Take the graph with four vertices $V = (V_1, V_2, V_3, V_4)$ and edges $E = \{(V_1, V_2), (V_1, V_3), (V_1, V_4), (V_2, V_3)\}$. The independent sets are $\{V_1\}$, $\{V_2\}$, $\{V_3\}$, $\{V_4\}$, $\{V_2, V_4\}$ and $\{V_3, V_4\}$. Thus, we have a local algebra with six factors, $A = A_1 \otimes A_2 \otimes A_3 \otimes A_4 \otimes A_{24} \otimes A_{34}$. Operator representations for letters in V_1, V_2, V_3 and V_4 respectively act on the subalgebras

$$\begin{aligned} A(1) &= A_1, \\ A(2) &= A_2 \otimes A_{24}, \\ A(3) &= A_3 \otimes A_{34}, \\ A(4) &= A_4 \otimes A_{24} \otimes A_{34}. \end{aligned} \tag{2.32}$$



This corresponds with Example 74 collecting letters with same commutations.

2.3 Dimension constraints

Given a polynomial optimization problem, (p, G, R) on $k\langle \mathbf{M} \rangle$, we may be interested in optimizing over evaluations (λ, ψ) of bounded dimension. That is, $\psi : k\langle \mathbf{M} \rangle \rightarrow \mathcal{B}(H)$ with $\dim H \leq d$. For finite dimension, the commutativity paradigm is equivalent to the tensor product paradigm. The local structure $H = \bigotimes_{s \in S} H_s$ can be obtained with the procedure employed in Example 79. In this case, it is possible to bound the dimension of each local factor, $\dim H_s \leq d_s$. Thus, for $d : S \rightarrow \mathbb{N} \cup \{\infty\}$, we denote $(p, G, R)_d$ the corresponding polynomial optimization problem with dimension constraints. [Definition 40]

One way of ruling out undesired evaluations is to impose additional polynomial constraints on the problem that are satisfied by matrices of a fixed finite dimension [29]. That is, we can extend G with a collection of *matrix polynomial identities* satisfied by matrices of size d [58, 89, 90, 91]. For instance, all polynomial identities for matrices of size two are a consequence of *Hall* and *completely antisymmetric* polynomial identities [29, §III],

$$[[x_1, x_2]^2, x_3] = \mathbf{0} \quad \& \quad \sum_{\sigma \in \mathcal{S}_4} \text{sign}(\sigma) x_{\sigma(1)} x_{\sigma(2)} x_{\sigma(3)} x_{\sigma(4)} = \mathbf{0} \quad (2.33)$$

for all $x_1, x_2, x_3, x_4 \in \mathcal{B}(k^2)$. These have degrees five and four respectively. Thus, the optimal value of $(p, G, R)_d$ can be approximated extending G with a collection of matrix polynomial identities $\text{MPI}(d)$ for the prescribed dimensions. This corresponds with the problem $(p, G \cup \text{MPI}(d), R)$ [29, §III],

$$\begin{aligned} \max \quad & p \\ \text{s.t.} \quad & r \geq \mathbf{0} & \forall r \in R, \\ & q = \mathbf{0} & \forall q \in Q \\ & t = \mathbf{0} & \forall t \in \text{MPI}(d). \end{aligned} \quad (2.34)$$

An alternative complete hierarchy of semidefinite programming relaxations for the dimensionally constrained polynomial optimization problem was proposed in [92, 29]. This approach consists in restricting the moment matrices in the moment relaxations in Eq. (1.33) to belong to the linear space spanned moment matrices of the target dimension, which can be obtained sampling.

The drawback of this method is that numerical errors in Gram-Schmidt orthonormalization procedure become unstable at some critical basis size. More stability can be attained with Hausenholder reflections [93], or the linear subspace can be obtained with symbolic computations [29, §III]. These techniques were adapted in [3] to accommodate operators with overlapping supports. More concretely, one samples a state ρ and operators X on the target dimension and construct moment matrices M , with

$$M_{u,v} = \text{tr}[\rho u^*(X)v(X)] \quad (2.35)$$

indexed by a basis $\mathcal{I}$ of a linear subspace $V \subset k\langle \mathbb{M} \rangle$, until their linear span stabilises at $\mathcal{M}(d)$. Similarly, for each constraint $s \in S$ one samples the space $\mathcal{M}^s(d)$ spanned by the corresponding localizing matrices. The following semidefinite programming relaxations are complete [29, §IV],

$$\begin{aligned} \max \quad & \sum_w p_w M_w \\ \text{s. t.} \quad & M_1 = 1, \\ & M, M^s \geq 0, \\ & M \in \mathcal{M}(d), M^s \in \mathcal{M}^s(d). \end{aligned} \quad (2.36)$$

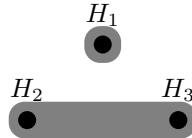
That is, for subspaces V converging to $k\langle \mathbb{M} \rangle$, the corresponding semidefinite program optimal solutions converge to the optimal solution of $(p, G, S)_d$.

Operators with overlapping supports

We illustrate convergence in the setting in Example 79. Consider a polynomial optimization problem $(p, G, S)_d$ on $k\langle X \rangle$, where the letters are divided into four subsets $X = V_1 \cup V_2 \cup V_3 \cup V_4$, each one acting on a local subalgebra. Namely V_1 , V_2 and V_3 acts on factors H_1 , H_2 and H_3 , while V_4 acts on $H_2 \otimes H_3$ (corresponds with A_2 , A_3 and A_4 being trivial in Example 79). The dimensions of the local factors are bounded by $d = (d_1, d_2, d_3, d_2 d_3)$. We now want to optimize over all evaluations (λ, ψ) given by $\psi : X \rightarrow \mathcal{B}(H_1 \otimes H_2 \otimes H_3)$ where

$$\begin{aligned} \psi : V_1 &\rightarrow \mathcal{B}(H_1) \\ \psi : V_2 &\rightarrow \mathcal{B}(H_2) \\ \psi : V_3 &\rightarrow \mathcal{B}(H_3) \\ \psi : V_4 &\rightarrow \mathcal{B}(H_2) \otimes \mathcal{B}(H_3) \end{aligned} \quad (2.37)$$

and states $\lambda \in \mathcal{S}(H_1 \otimes H_2 \otimes H_3)$ where $\dim H_i \leq d_i$ for $i = 1, 2, 3$.



We follow the sampling technique [92] and consider the dimensionally constrained semidefinite programming relaxations in Eq. (2.36). Let p^* denote the optimal value of the polynomial optimization problem $(p, G, S)_d$ and $p_{\mathcal{I}}$ the optimal value of the relaxation spanned by a basis vector $\mathcal{I}$. By construction, $p^* \leq p_{\mathcal{I}}$. We show that in the limit when the indexing sequence $\mathcal{I}$ spans the whole space, $k\langle X \rangle$, the converse also holds.

Claim 80. *There exists a limiting sequence of moments $\hat{y}$.*

Proof. Same as Claim 46. $\square$

Let $\hat{p} = \lim_{\mathcal{I}} p_{\mathcal{I}}$ denote the limiting optimal value of the relaxations. The next step is to construct from the infinite sequence of moments, $\hat{y}$, a triple $(\hat{H}, \hat{\mathbf{X}}, \hat{\psi})$ attaining the value $\hat{p}$. For this, we perform the GNS construction.

Claim 81. *There is a triple $(\hat{H}, \hat{\mathbf{X}}, \hat{\psi})$ attaining the value $\hat{p}$.*

Proof. Same as Claim 47. $\square$

By construction, the expectation value of $p(\hat{\mathbf{X}})$ over the state $\hat{\psi}$ is precisely $\hat{p}$. The last step to show that $\hat{p} \leq p^*$ is to decompose the Hilbert space and the bounded operators in a direct integral of spaces that factorize as $H_1 \otimes H_2 \otimes H_3$. This way, $\hat{p}$ is a convex combination of feasible values of $(p, G, S)_d$, and hence can not exceed the optimal value p^* . The following claim is adapted from [29, §IV] to accommodate measurements with overlapping supports.

Claim 82. *Assume each V_i is closed under involution and $\mathbf{1} \in \langle V_i \rangle / \langle G \rangle$. Then, [29, §IV] the value $\hat{p}$ is a convex combination of feasible values for $(p, G, S)_d$.*

Proof. Denote by $\mathfrak{A}_i$ the $*$ -algebra generated by the operators in $\hat{\mathbf{X}}$ corresponding to letters in V_i , which by assumption is closed under adjoints and contains the identity. Then, $\mathfrak{A}_i''$ is a von Neumann algebra over the separable Hilbert space $\hat{H}$ [49, Theorem 3.9] and hence decomposes as a direct integral of factors [49, §IV.8],

$$\mathfrak{A}_i'' = \int \mathfrak{A}_{i,z} d\mu(z). \quad (2.38)$$

By construction these factors must satisfy all matrix polynomial identities for dimension d_i , which excludes factors of type I_n for $n > d_i$ and factors of types II and III [29, Appendix A]. Thus, each $\mathfrak{A}_{i,z}$ is isomorphic to $\mathcal{B}(H_{i,z}) \otimes \mathbf{1}$ where $\dim H_{i,z} \leq d_i$, and we can assume an equality up to an isometric embedding.

In each sector z , the intersection of commuting factors must be trivial

$$\begin{aligned} \mathbb{C}\mathbf{1} &\supset \mathfrak{A}_{1,z} \cap (\mathfrak{A}_{2,z} \cup \mathfrak{A}_{3,z} \cup \mathfrak{A}_{4,z}), \\ \mathbb{C}\mathbf{1} &\supset \mathfrak{A}_{2,z} \cap (\mathfrak{A}_{1,z} \cup \mathfrak{A}_{3,z}), \\ \mathbb{C}\mathbf{1} &\supset \mathfrak{A}_{3,z} \cap (\mathfrak{A}_{1,z} \cup \mathfrak{A}_{2,z}), \\ \mathbb{C}\mathbf{1} &\supset \mathfrak{A}_{4,z} \cap \mathfrak{A}_{1,z}. \end{aligned} \quad (2.39)$$

Counting dimensions shows that the Hilbert space in each sector decomposes as $H_z = H_{1,z} \otimes H_{4,z}$ with $H_{4,z} = H_{2,z} \otimes H_{3,z}$. Moreover, for each constraint $s \in S$, $s(\hat{\mathbf{X}}) \geq \mathbf{0}$ implies $s(\hat{\mathbf{X}}_z) \geq \mathbf{0}$, where $\hat{X}_{i,z}$ corresponds with the direct component of the operator $\hat{X}_i$ acting on H_z . Thus, $\hat{p}$ is a convex combination of feasible values for $(p, G, S)_d$. $\square$

In particular, the assumptions in Claim 82 hold when the letters describe measurement effects whose support may overlap. These scenarios, considered in [3], are discussed in Chapter 3 (see Section 3.4).

2.4 Computational reductions

As we discussed in Section 1.6, there is no general method to choose a relaxation subspace giving good bounds for a polynomial optimization problem. The standard approach is to choose subspaces V_d spanned by all monomials up to degree d , which grow very fast. In such a case, one soon encounters a semidefinite program too big to handle. In this last section, we review some general tools to reduce both the size of the subspaces and the computational cost of the corresponding relaxations. We can classify them into sparsity [94, 95, 96, 60], symmetry [97, 98] and Jordan algebra [99] reductions.

Sparsity reductions

There are different notions of *sparsity*. For polynomial optimization, we distinguish *correlation sparsity* for few products in the monomials and *term sparsity* for few terms in the polynomials [95, 96]. For general semidefinite programs, we also have *data sparsity* for few non-zero terms in the input data [94]. All these forms of sparsity can be exploited both to reduce the computational cost of the semidefinite programming relaxations and to design reduced relaxations.

Correlation sparsity techniques [60, §6.2] reduce the problem neglecting monomial products that do not contribute to the sparsity pattern determined by the monomials in the support of the objective and constraint polynomials.

[60, §1.2] **Definition 83** (Sparsity pattern). *A matrix $A \in \text{Mat}(n)$ is said to have sparsity pattern given by a graph $G = ([n], E)$ when $A_{ij} = 0$ for all $(i, j) \notin E$. The collection of matrices with sparsity pattern G is denoted $\text{Mat}_G(n)$. When the graph is chordal, that is, every cycle of length more than three has an edge between non consecutive vertices, positive matrices decompose into positive blocks for each maximal clique of G (these may overlap) [60, Theorem 1.4].*

This can reduce polynomial optimization problems [60, §3.1, §6.2]. Consider a problem $(p, \emptyset, S)$ in commutative variables and a sequence of monomials $\mathcal{I}$ indexing the matrices of moments in the semidefinite programming relaxations. Start from the sparsity pattern G determined by the connections between variables in the objective p and the constraints $s \in S$. That is, $G = (X, E)$ where $(x_i, x_j) \in E$ if either (i) x_i and x_j appear in some monomial in the support of

p or some $s \in S$; or (ii) x_i and x_j appear in terms of some $s \in S$. Let G' be a chordal extension of G with maximal cliques $I_1, \dots, I_p$, and let $S_i \subset S$ be those constraints involving variables in I_i . Then, $t - p$ is positive over S if and only if there exist a sum of squares decomposition [60, Theorem 3.8]

$$t - p = \sum_i (\sigma_i + \sum_{s \in S_i} \sigma_{i,s} s), \quad (2.40)$$

where σ_i and $\sigma_{i,s}$ denote sum of squares polynomials in $k[I_i]$.

This certificate for positivity can be turned into a complete hierarchy of sparsity reduced semidefinite programming relaxations [60, §3.3],

$$\begin{aligned} \max \quad & \sum_w p_w y_w \\ \text{s. t.} \quad & y_1 = 1, \\ & M(I_i) \geq 0, M(I_i) \in \mathcal{M}(I_i), \\ & M^s(I_i) \geq 0, M^s(I_i) \in \mathcal{M}^s(I_i). \end{aligned} \quad (2.41)$$

Here, $M(I_i)$ is the matrix of moments involving variables only in I_i and $M^s(I_i)$ is the localizing matrix of a constraint $s \in S_i$ involving variables in I_i . Moments must be compatible on the intersections, namely, $M(I_i)_{u,v} = M(I_j)_{u',v'}$ whenever $u^*v = u'^*v' \in \langle I_i \rangle \cap \langle I_j \rangle$. Optimal solutions of (2.41) converge to the optimal solution of the optimization problem $(p, \emptyset, S)$ for increasing monomial bases [60, Theorem 3.6], assuming archimedeanity condition is satisfied at each clique. That is, there are natural numbers N_i such that $N_i - \sum_{x \in I_i} x^2 \in \mathcal{Q}(S_i)$ is a localized sum of squares. However, for a fixed monomial basis the sparsity reduced bound may be weaker than the bound in the dense relaxation because of additional linear constraints on the moment matrix [60, Remark 3.7]. Analogous reductions exist for non-commutative polynomials [96] relying on amalgamation of C^* -algebras (see [60, §6.4]).

Term sparsity provides a different source to reduce polynomial optimization problems exploiting the number of terms appearing in the polynomials. The first example of term sparsity reduction is *Newton method* [100, 101], which prunes unnecessary monomials in the sum of squares decomposition. Thus, we can restrict the monomial basis $\mathcal{I}$ in the semidefinite programming relaxations to smaller subsets enhancing the computational performance.

Example 84 (Newton method). *Consider an unconstrained polynomial optimization problem $(p, \emptyset, \emptyset)$ in $k[X]$ with n generators. Monomials in $[X]$ are fully specified by their exponents, we write $x^a = x_1^{a_1} \dots x_n^{a_n}$ for $a \in \mathbb{N}^n$. The product of monomials translates into addition of their exponents, thus a decomposition $p = \sum_i f_i^2$ limits the exponents that can appear in f_i . In particular, the exponents of all the polynomials in the decomposition must be inside* [60, §7.1]

$$N(p) = \text{conv}\{a/2 \in \mathbb{R}^n : x^a \in \text{supp}(p)\} \cap \mathbb{N}^n, \quad (2.42)$$

which is called the reduced Newton polytope of p .

Newton method can also be extended [102, Algorithm 2.2] to unconstrained polynomial optimization problems in non-commutative variables $k\langle X \rangle$ and commutative problems $(p, G, \emptyset)$ in $k[X]$ with equality constraints [99]. The *reduced Newton polytope* in the latter case is given by [99, §III]

$$N(p, G) = \{b \in \text{supp}(\lceil x^a \rceil) : a \in S/2 \cap \mathbb{N}^n\}, \quad (2.43)$$

where $\lceil x^a \rceil$ denotes the canonical form of the monomial $x^a \in [X]$ with respect to the complete set of replacement rules determined by a Gröbner basis for G and $S = \bigcup_{g \in G} \text{supp}(g)[X] \cup \text{supp}(p)$.

Exploiting term sparsity in commutative and non-commutative constrained optimization problems involves a more sophisticated iterative protocol of support and chordal extensions over the Newton polytope [96]. The procedure produces sequences of graphs $G_0^{(i)}$ and $G_s^{(i)}$ that give rise to the semidefinite programming relaxations [60, §10.1]

$$\begin{aligned} \max \quad & \sum_w p_w y_w \\ \text{s. t.} \quad & y_{\mathbf{1}} = 1, \\ & G_0^{(i)} \circ M \in \text{PSD}(G_0^{(i)}), M \in \mathcal{M} \\ & G_s^{(i)} \circ M^s \in \text{PSD}(G_s^{(i)}), M^s \in \mathcal{M}^s. \end{aligned} \quad (2.44)$$

Here, $G \circ M$ denotes the matrix obtained from M setting $M_{u,v} = 0$ for all pairs $(u, v) \notin E(G)$ and $\text{PSD}(G)$ is the set of matrices with sparsity pattern G that admit a positive semidefinite completion. These relaxations give upper bounds for the dense relaxation at each sublevel i . Moreover, these sparse bounds recover the dense bound at some finite step when the procedure is performed with maximal chordal extensions [60, Theorem 10.1]

Last, *data sparsity* provides another source of reduction that exploits sparsity in the data defining a semidefinite program [94] that is not restricted to polynomial optimization problems. In particular, we can restrict the feasible region of a semidefinite program

$$\begin{aligned} \max \quad & \langle C, X \rangle \\ \text{s. t.} \quad & \langle X, A_i \rangle = b_i, \\ & X \geq 0, \end{aligned} \quad (2.45)$$

to the matrix $*$ -algebra generated by the objective function C and the linear constraints A_i preserving the optimal solution [94, Theorem 2]. Data sparsity can also be combined with term and correlation sparsity reductions.

Symmetry reduction

A problem being invariant under some action manifest the presence of symmetries that can be exploited to reduce both its description and computational

cost. Here, we consider monomial symmetries to reduce the semidefinite programming relaxations of polynomial optimization problems [97, 98].

Let G be a finite group with a *unitary* action on a Hilbert space H , that is, [97, §3] $\langle g(x), y \rangle = \langle x, g^{-1}(y) \rangle$ for all $x, y \in H$ and $g \in G$. A semidefinite program

$$\begin{aligned} \max \quad & \langle C, X \rangle \\ \text{s.t.} \quad & X \in X_0 + \mathcal{L}, \\ & X \geq 0, \end{aligned} \tag{2.46}$$

is *invariant* under the action of G if $g(C) = C$, $g(X_0) = X_0$ and $g(\mathcal{L}) \subset \mathcal{L}$ for all $g \in G$. Since the action is unitary, $\langle g(C), X \rangle = \langle C, g^{-1}(X) \rangle$, the optimal solution of invariant programs is preserved under symmetrization. That is,

$$\begin{aligned} \max \quad & \langle C, \tilde{X} \rangle \\ \text{s.t.} \quad & \tilde{X} \in X_0 + \mathcal{L}, \\ & \tilde{X} \geq 0, \\ & \tilde{X} = g(\tilde{X}) \quad \forall g \in G, \end{aligned} \tag{2.47}$$

preserves the optimal solution [97, Theorem 3.3]. The symmetrized semidefinite program identifies variables in the same *orbit* under the action of G , reducing its size. Moreover, positive semidefinite matrices commuting with the action of a finite group G admit a *Wedderburn* decomposition $\tilde{X} = \bigoplus_i \tilde{X}_i$ into positive semidefinite blocks [97, §4] that corresponds with irreducible representations of G and can often be obtained explicitly with representation theoretical techniques [103, §2.7].

In particular, a unitary action of a finite group G on $k\langle \mathbb{M} \rangle$ induces a unitary action on the Hilbert space of matrices indexed with elements $\mathcal{I} \subset k\langle \mathbb{M} \rangle$. Indeed, assume that $\mathcal{I}$ is closed under the action of G ($g(u) \in \mathcal{I}$ for each $u \in \mathcal{I}$) and take $g(M)_{u,v} = M_{g(u),g(v)}$. The semidefinite programming relaxation for the polynomial optimization problem $(p, \emptyset, S)$ is invariant under the action of G whenever $g(p) = p$ and $g(s) \in \mathcal{Q}(S)$ is a localized sums of squares for each $s \in S$ and $g \in G$. However, we can not restrict a sum of squares decomposition to the symmetric subspace under the action of G [97, Example 5.1].

Jordan algebra reduction

Explicitly finding symmetries of an optimization problem can be challenging. An alternative axiomatic approach was introduced in [104], characterizing those subspaces that preserve optimal solutions. Under mild structural assumptions that are satisfied in semidefinite programs, the *smallest* such subspace can be constructed with a procedure of completing squares. Moreover, there are efficient linear approximations to the optimal invariant subspace.

Consider the primal-dual pair of semidefinite programs

$$\begin{aligned} \max \quad & \langle C, X \rangle \\ \text{s.t.} \quad & X \in X_0 + \mathcal{L} \\ & X \geq 0, \end{aligned} \tag{2.48}$$

$$\begin{aligned} \min \quad & -\langle X_0, Y \rangle \\ \text{s.t.} \quad & Y \in C + \mathcal{L}^\perp \\ & Y \geq 0. \end{aligned} \tag{2.49}$$

Following [104, §2.1], we say a projector P_V projects into an *invariant* subspace V if it preserves (i) positivity, $P_V(A) \geq 0$ for all $A \geq 0$, (ii) primal feasibility, $P_V(X_0 + \mathcal{L}) \subset X_0 + \mathcal{L}$, and (iii) dual feasibility, $P_V(C + \mathcal{L}^\perp) \subset C + \mathcal{L}^\perp$. This is enough to preserve the primal and dual cost functions [104, Proposition 2.1]. That is, we preserve the optimal value when we restrict our problem to an invariant subspace [104, Proposition 2.3],

$$\begin{aligned} \max \quad & \langle P_V(C), X \rangle \\ \text{s.t.} \quad & X \in P_V(X_0) + \mathcal{L} \cap V \\ & X \geq 0, \end{aligned} \tag{2.50}$$

$$\begin{aligned} \min \quad & -\langle P_V(X_0), Y \rangle \\ \text{s.t.} \quad & Y \in P_V(C) + \mathcal{L}^\perp \cap V \\ & Y \geq 0. \end{aligned} \tag{2.51}$$

Conditions (ii) and (iii) are equivalent to $P_{\mathcal{L}}(C), P_{\mathcal{L}^\perp}(X_0) \in V$ and $P_{\mathcal{L}}(V) \subset V$. Moreover, when V is an Euclidean Jordan algebra, condition (i) is equivalent to $\{A^2 : A \in V\} \subset V$ [104, Theorem 3.1]. That is, V contains all its squares.

[105, §II.1] **Definition 85** (Jordan algebra). *A commutative algebra $(A, +, \mathbf{0}, *)$ over k that satisfies Jordan identity, $(a * b) * (a * a) = a * (b * (a * a))$ for all $a, b \in A$, is called Jordan algebra. An associative algebra $(A, +, \mathbf{0}, \circ)$ defines a Jordan algebra with the symmetrized product $a * b = (a \circ b + b \circ a)/2$, these algebras*
 [105, §III.1] *are called special. A Jordan algebra is Euclidean if it has an inner product satisfying $\langle a * b, c \rangle = \langle b, a * c \rangle$ for $a, b, c \in A$. Euclidean Jordan algebras are*
 [105, §III.2] *unital and their squares form a self-dual convex cone, $\mathcal{K} = \{x * x : x \in A\}$.*

In particular, symmetric real matrices or hermitian complex matrices form Euclidean Jordan algebras with respect to the symmetrized matrix product. This gives Algorithm 4 [104, Theorem 3.2] to construct the smallest invariant subspace, which is a Jordan subalgebra V . Moreover, let A be the matrix $*$ -algebra generated by V and $A = \bigoplus_i A_i$ its *Wedderburn decomposition* into simple ideals. If V coincides with the symmetric elements of A , then the symmetric elements of each Wedderburn factor of A decompose V into simple ideals [104, Proposition 2.9].

The outcome of Algorithm 4 is optimal in the sense that it produces a block-diagonal structure that is majorized by any other reduction leaving invariant the optimal solution [104, Theorem 4.1]. Notice that the step of square

Jordan algebra reduction

Completion procedure to obtain the *smallest* Jordan algebra for the semidefinite programs (2.48) and (2.49) [104, Theorem 3.2].

Algorithm 4 Optimal invariant subspace

-
- 1: $V \leftarrow \text{span}\{P_{\mathcal{L}}(C), P_{\mathcal{L}^\perp}(X_0)\}$
 - 2: $V \leftarrow V + P_{\mathcal{L}}(V)$
 - 3: $V \leftarrow V + \{A^2 : A \in V\}$
 - 4: Repeat 2 and 3 until V stabilizes
-

completion is non linear and may be expensive to implement. However, two efficient combinatorial relaxations to this optimal subspace are provided in [104, §5]. One gives the optimal *partition subspace* and other the optimal *relation subspace*, which are particular cases of *0-1 subspaces* and admit efficient descriptions. Moreover, the optimal 0-1 subspace is at least as small as any symmetry reduced space [104, §5] and data sparsity reduced space [104, §6].

Chapter 3

Correlation Settings

This section focusses on characterizing quantum behaviours from measurement correlations. In other words, we aim to identify which physical correlations are compatible with quantum theory. First, we recall how quantum correlations in Bell scenarios are characterized with polynomial optimization. Then, we introduce correlation settings, which can be understood as more general Bell experiments that permit overlapping measurements. We then propose complete hierarchies to characterize quantum behaviours in these settings and investigate novel practical applications. After that, we compare them with the family of measurement protocols known as contextuality scenarios, which also can be understood as generalizations of Bell scenarios but on a different direction. Last, we discuss how correlation settings accommodate several problems in quantum information such as the quantum marginal problem, the existence of absolutely maximal entanglement, or ground energies of local Hamiltonians.

3.1 Bell scenarios

Bell scenarios are local measurement protocols in multipartite systems. Some measurement correlations obtained from these protocols are not compatible with the hypotheses of *locality* (space-like separated systems do no influence each other) and *realism* (physical quantities have well defined values that are revealed in the measurement process) [14], which are termed *nonlocal*. Bell experiments had not only breakthrough consequences in foundations of physics but also finds several technological applications. Nonlocality gives advantages in several informational tasks including conjugate coding [17], key distribution [18] communication complexity [19] or randomness verification [20]. Thus, it is an important problem to determine those resourceful nonclassical behaviours that can be obtained with quantum systems. For this, one needs to consider all possible physical realizations compatible with quantum theory, which requires exploring all possible evaluations on Hilbert spaces.

Characterization from Polynomial Optimization

Quantum observables are represented as self-adjoint bounded operators over a Hilbert space. Thus, to explore all quantum realizations of a measurement protocol one has to consider all possible representations of the measurement effects as self-adjoint bounded operators over some Hilbert space. When the effects are expected to satisfy prescribed polynomial relations, quantum realizations correspond to feasible solutions of a polynomial optimization problem (1.21). The hierarchies of efficient relaxations described in Section 1.6 can be used to characterize quantum behaviours [2] (see also [15, §II.A.3]) and the maximum quantum value of Bell functionals [28, §5] (see also [15, §II.C.1]).

Take the Bell scenario $B(2, 2, 2)$ with two parties performing two dichotomic measurements. Consider the alphabet $X = (A_{a|x}, B_{b|y} : a, b, x, y \in [2])^1$, where letter $A_{a|x}$ corresponds to the effect with outcome a for measurement x performed by the first party, and $B_{b|y}$ the analogue for the second party. The behaviour of projective effects is captured by the polynomial constraints

$$\mathbf{0} = A_{a|x}^2 - A_{a|x}, \quad \mathbf{0} = B_{b|y}^2 - B_{b|y}, \quad (3.1)$$

$$\mathbf{0} = \mathbf{1} - \sum_a A_{a|x}, \quad \mathbf{0} = \mathbf{1} - \sum_b B_{b|y}, \quad (3.2)$$

and the local structure is encoded in the commutativity constraints

$$\mathbf{0} = A_{a|x} B_{b|y} - B_{b|y} A_{a|x}. \quad (3.3)$$

Notice that the constraints above form a Gröbner basis, thus can be efficiently implemented as replacement rules both reducing the number of variables and removing the need to introduce localizing matrices. Moreover, the representation of polynomials in the partially commutative ring $k\langle X; \mathcal{C} \rangle$ automatically takes care of the commutation relations in Eq. (3.3). Thus, the Gröbner basis is given by the projection constraints in Eq. (3.1), which can be efficiently implemented as the level of multiplication.

Given an evaluation (λ, ψ) on some Hilbert space, Born's rule produces a probability model $p \in \mathcal{Q}$,

$$p(a, b|x, y) = \lambda(\psi(A_{a|x} B_{b|y})). \quad (3.4)$$

Thus, it is a feasible point of the semidefinite programming relaxations (1.33)

$$\begin{aligned} [15, \text{§II.A.3}] \quad & \max \quad 0 \\ & \text{s. t.} \quad y_1 = 1, \\ & \quad y_{A_{a|x} B_{b|y}} = p(a, b|x, y), \\ & \quad M \geq 0, \\ & \quad M \in \mathcal{M}. \end{aligned} \quad (3.5)$$

Completeness of the relaxations of polynomial optimization problem implies that the converse is also true. That is, a probability model is quantum, $p \in \mathcal{Q}$, if and only if the relaxation (3.5) is feasible for every indexing sequence $\mathcal{I}$.

¹Given a natural number n , we denote $[n] = \{1, \dots, n\}$.

Characterization from physical principles

Several physical principles have been proposed in the literature to characterize the set of quantum correlations. These include non trivial communication complexity [106], no advantage for nonlocal computation [107], information causality [108], macroscopic locality [109], local orthogonality [110] or many box locality [111]. Despite ruling out certain non-signalling correlations, these principles are not enough to characterize quantum correlations in arbitrary settings. Recently, this goal was achieved [112] combining a principle that generalizes non-signalling to dilations, called *non disturbance*, with a compatibility condition on local time orientations.

3.2 Correlation settings

Next, we introduce more general settings that account for measurements with overlapping supports. Some examples were investigated in [3].

Definition 86 (Correlation setting). *A correlation setting $\mathfrak{S}$ is a collection of triples (s, x, a) . Each triple is called an event and we refer to x as the measurement, a the outcome and s the system support. We denote the collection of supports as $S := \{s : (s, x, a) \in \mathfrak{S}\}$. We say the setting has no overlap if the elements of S are pairwise disjoint, otherwise we say the setting has overlap.*

A setting $\mathfrak{S}$ describes a measurement scenario in which different measurements with different outcomes can be performed on different systems. Definition 86 is a natural generalization of the Bell scenario $B(n, m, k)$, that consider n parties performing m measurements with k outcomes, which corresponds to the setting $\mathfrak{S} = [n] \times [m] \times [k]$. Measurements performed over different systems are compatible, which means can be simultaneously performed. Thus we can combine events whenever their supports do not intersect. This motivates the following definition.

Definition 87 (Composite events). *The composite events on the correlation setting $\mathfrak{S}$ are triples of tuples recursively defined as (i) $\mathfrak{S} \subset \mathfrak{E}$ and (ii) $e_1 \vee e_2 := ((s_1, s_2), (x_1, x_2), (a_1, a_2)) \in \mathfrak{E}$ when $e_1 = (s_1, x_1, a_1), e_2 = (s_2, x_2, a_2) \in \mathfrak{E}$ and $s_1 \cap s_2 = \emptyset$. Here, (s_1, s_2) denotes the concatenation of the tuples s_1 and s_2 , and the intersection is meant pairwise. In such a case, we say e_1 and e_2 are sub-events of $e_1 \vee e_2$ and write $e_1, e_2 \leq e_1 \vee e_2$. This defines a partial order on $\mathfrak{E}$ that has $e_1 \vee e_2$ as the least upper bound, which justifies our choice of notation.*

We identify elements in $\mathfrak{E}$ that differ in a permutation applied to the three tuples. Thus $e_1 \vee e_2 = e_2 \vee e_1$ works as a logical connective for events and $\mathfrak{E}$ can be understood as a bounded *lattice* with minimal elements $\mathfrak{S}$ and maximal elements $\mathfrak{m} := \max \mathfrak{E}$. A physical model for the setting must assign to each event a probability that is compatible with the partial order of the events. That means, the probability of an event can be obtained from any composite

event extending it by forgetting outcomes, which corresponds to taking the marginal of a distribution with more independent variables.

Definition 88 (Probability model). *A probability model p for the setting $\mathfrak{S}$ is a map $p : \mathfrak{E} \rightarrow [0, 1]$ such that (i) $\sum_a p(s, x, a) = 1$ for all $(s, x, a) \in \mathfrak{E}$ and (ii) $p(s, x, a) = \sum_{\bar{a}} p((s, \bar{s}), (x, \bar{x}), (a, \bar{a}))$ for any $((s, \bar{s}), (x, \bar{x}), (a, \bar{a})) \in \mathfrak{E}$ extending $(s, x, a) \in \mathfrak{E}$. We denote the set of probability models as $\mathcal{P}(\mathfrak{S})$.*

Notice that condition (ii) implies that a model is fully specified by its values on the set $\mathfrak{m}$ of maximal events. Models in certain correlation settings convey fundamental properties of the underlying physical theories giving rise to them. For instance, there are models in the Bell scenario $\mathfrak{S} = B(2, 2, 2)$ that are not compatible with local deterministic theories [14] but can be explained within quantum theory. This motivates the definition of the following sets of probability models.

Definition 89 (Local model). *A probability model $p \in \mathcal{P}(\mathfrak{S})$ is deterministic if $\text{ran } p \subset \{0, 1\}$. For each event $(s, x, a) \in \mathfrak{E}$, a deterministic model factorizes as the product of its marginals, $p(s, x, a) = \prod_i p_i(s_i, x_i, a_i)$. Local models, denoted $\mathcal{L}(\mathfrak{S})$, are the convex span of deterministic models.*

Outcome probabilities in quantum theory are obtained with *Born's rule*, and essentially only this way [113], which justifies the following definition.

Definition 90 (Quantum model). *A model $p \in \mathcal{P}(\mathfrak{S})$ is quantum if for each $(s, x, a) \in \mathfrak{E}$ there exist a quantum state ρ_s and an effect $E_{a|x}$ on some Hilbert space H_s such that (i) $p(s, x, a) = \text{tr}(\rho_s E_{a|x})$, (ii) $E_{a|x} = E_{a_1|x_1} \otimes E_{a_2|x_2}$ and $H_s = H_{s_1} \otimes H_{s_2}$ when $(s, x, a) = (s_1, x_1, a_1) \vee (s_2, x_2, a_2)$ and (iii) the marginal states ρ_s are compatible with a global state ρ_S on the global Hilbert space H_S . We denote the set of quantum models as $\mathcal{Q}(\mathfrak{S})$, and $\mathcal{Q}_d(\mathfrak{S})$ those satisfying the additional constraints (iv) $\dim H_s \leq d(s)$ for a given $d : S \rightarrow \mathbb{N} \cup \{\infty\}$.*

Condition (i) is Born's rule, (ii) imposes that composite events are determined by its constituents, (iii) is the *quantum marginal problem*, i.e. each state ρ_s is the reduction to subsystem s of a global state ρ_S , and (iv) are additional constraints on the dimension. Again, it is enough to check the conditions on the set $\mathfrak{m}$ of maximal events. Notice that one can assume the effects to be projectors going to a dilation [114]. Thus, one can replace effects by decompositions of the identity into orthogonal projectors in the definition of $\mathcal{Q}(\mathfrak{S})$. However, a dilation is not possible if the Hilbert space dimension is restricted. The choice for effects in the definition is so that the sets $\mathcal{Q}_d(\mathfrak{S})$ are also convex, thus more amenable to analysis.

Proposition 91. *Let $\mathfrak{S}$ be a correlation setting, we have the chain of inclusions*

$$\mathcal{L}(\mathfrak{S}) \subset \mathcal{Q}_d(\mathfrak{S}) \subset \mathcal{Q}_{d'}(\mathfrak{S}) \subset \mathcal{Q}(\mathfrak{S}) \subset \mathcal{P}(\mathfrak{S}) \quad (3.6)$$

when $1 \leq d \leq d'$ (in the sense $d(s) \leq d'(s)$ for each $s \in \bigcup S$).

Proof. All inclusions but $\mathcal{L}(\mathfrak{S}) \subset \mathcal{Q}_d(\mathfrak{S})$ follow from the definitions. Also, that these sets are convex. Thus, we only need to show that every local deterministic model is already in $\mathcal{Q}_d(\mathfrak{S})$ when $d \geq 1$. Take such a model, p , which is specified by its values on maximal events. The one dimensional Hilbert space with projectors $E_{a|x} = p(x, a)\mathbf{1}$ reproduces the probabilities and fulfills all conditions. $\square$

As mentioned before, probability models $p \in \mathcal{Q}(\mathfrak{S}) - \mathcal{L}(\mathfrak{S})$ can bring significant practical consequences in conjugate coding [17], key distribution [18] communication complexity [19] or randomness verification [20]. Furthermore, models $p \in \mathcal{P}(\mathfrak{S}) - \mathcal{Q}(\mathfrak{S})$ can also bring significant post-quantum advantages in certain information theoretical tasks such as random access codes [108, 115]. It is an important problem to characterise these different subsets. For this, we use standard arguments in convex analysis.

Examples

When $\mathfrak{S}$ has no overlap, the elements of the support $S = \{s_1, \dots, s_t\}$ are pairwise disjoint and events in $\mathfrak{S}$ can be combined into global composite events. That is, maximal events include all elements of S in their support, thus have the form $(a|x) := ((s_1, \dots, s_t), (x_1, \dots, x_t), (a_1, \dots, a_t)) \in \mathfrak{m}$.

Example 92 (Bell scenario). *Consider the Bell scenario $B(2, 2, 2)$ where two parties perform two measurements with two outcomes. The corresponding correlation setting is $\mathfrak{S} = [2] \times [2] \times [2]$. The composite events $\mathfrak{E}$ are all elements $(a_i|x_i) := (i, x_i, a_i)$ in $\mathfrak{S}$ and $(a_1 a_2|x_1 x_2) := ((1, 2), (x_1, x_2), (a_1, a_2))$ for x_1, x_2, a_1 and a_2 in $\{0, 1\}$. Their partial order is illustrated in Fig. 3.1.* [15, §II.A]

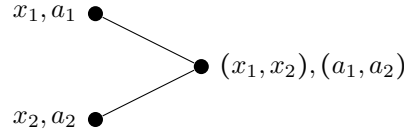


Figure 3.1: **Lattice of events $\mathfrak{E}$** for Bell scenario $\mathfrak{S} = B(2, 2, 2) = [2] \times [2] \times [2]$. Each choice $x_1, x_2, a_1, a_2 \in [2]$ gives an isomorphic section of the lattice.

A probability model is given by an assignment $0 \leq p(a|x) \leq 1$ for each $(a|x) \in \mathfrak{m}$ such that (i) $\sum_a p(a|x) = 1$ for each x and (ii) it is consistent with the order in $\mathfrak{E}$. Namely, $p_{s_2}(a_2|x_2) = \sum_{a_1} p_{s_1 s_2}(a_1, a_2|x_1 x_2)$ for any x_1 . Thus, the consistency conditions reduce to non-signalling in settings with no overlap. Additionally, p is deterministic if $p(a|x) \in \{0, 1\}$ for each $(a|x)$; and p is local if it can be written as the convex sum of deterministic models. Notice that checking this amounts to solving a system of linear equations with as many variables as deterministic models exist in the setting (finite number).

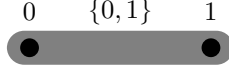
[15, §II.A] cf. Example 92 **Example 93** (Bell scenario). *Consider again the correlation setting $\mathfrak{S} = [2]^{\times 3}$ corresponding to the Bell scenario $B(2, 2, 2)$. There are four maximal measurements, corresponding to $x_1, x_2 \in \{0, 1\}$, with four possible choices for a deterministic output, corresponding to $a_1, a_2 \in \{0, 1\}$. Thus, there are 256 deterministic assignments $p(a|x) \in \{0, 1\}$. However, only sixteen of them satisfy the consistency conditions, giving deterministic models $p \in \mathcal{P}(\mathfrak{S})$ that span all local models $\mathcal{L}(\mathfrak{S})$. For instance, $p(10|00) = 1 = p(00|01)$ is not possible for a model $p \in \mathcal{P}(\mathfrak{S})$, since there is no consistent assignment for $p_1(0|0)$ and $p_1(1|0)$. Moreover, recall from Example 6 that there exist both nonlocal quantum models $p \in \mathcal{Q}(\mathfrak{S}) - \mathcal{L}(\mathfrak{S})$ and post-quantum non-signaling models $p \in \mathcal{P}(\mathfrak{S}) - \mathcal{Q}(\mathfrak{S})$.*

When the correlation setting $\mathfrak{S}$ has overlap, events decompose into events with different supports, $\mathfrak{E} = \bigoplus_{\sigma} \mathfrak{E}_{\sigma}$. A probability model $p \in \mathcal{P}(\mathfrak{S})$ gives by restriction $p_{\sigma} = p|_{\mathfrak{E}_{\sigma}} \in \mathcal{P}(\mathfrak{S}_{\sigma})$ satisfying $p_{\sigma_1}(s, x, a) = p_{\sigma_2}(s, x, a)$ for each common event $(s, x, a) \in \mathfrak{E}_{\sigma_1} \cap \mathfrak{E}_{\sigma_2}$. Conversely, a collection of models $p_{\sigma} \in \mathcal{P}(\mathfrak{S}_{\sigma})$ satisfying those compatibility conditions gives a joint model $p = \bigcup_{\sigma} p_{\sigma} \in \mathcal{P}(\mathfrak{S})$. In particular, let $\mathfrak{S} = \mathfrak{S}_1 \oplus \mathfrak{S}_2$ be the union of settings with no common events², $\mathfrak{E}_1 \cap \mathfrak{E}_2 = \emptyset$. Then, $\mathcal{P}(\mathfrak{S}) = \mathcal{P}(\mathfrak{S}_1) \times \mathcal{P}(\mathfrak{S}_2)$ and since this correspondence preserves deterministic models, also $\mathcal{L}(\mathfrak{S}) = \mathcal{L}(\mathfrak{S}_1) \times \mathcal{L}(\mathfrak{S}_2)$. The situation is similar for quantum models, $\mathcal{Q}(\mathfrak{S}) = \mathcal{Q}(\mathfrak{S}_1) \times \mathcal{Q}(\mathfrak{S}_2)$ since any $p \in \mathcal{Q}(\mathfrak{S})$ gives by restriction $p_i \in \mathcal{Q}(\mathfrak{S}_i)$ and conversely, the extension of a pair of quantum models $p_i \in \mathcal{Q}(\mathfrak{S}_i)$ is again a quantum model $p = p_1 \cup p_2 \in \mathcal{Q}(\mathfrak{S})$. Indeed, let $\rho_{i,s}$ ³ and $E_{i,a|x}$ be the quantum states and effects corresponding with model p_i . The states $\rho_s = \rho_{1,s} \otimes \rho_{2,s}$ and the effects $E_{a|x} = E_{1,a|x} \otimes \mathbf{1}$ if $(s, x, a) \in \mathfrak{E}_1$ and $E_{a|x} = \mathbf{1} \otimes E_{2,a|x}$ if $(s, x, a) \in \mathfrak{E}_2$, produce a quantum representation for the model $p \in \mathcal{Q}(\mathfrak{S})$. Notice that this argument fails when the dimensions are constrained, unless all quantum states $(\rho_{i,s})_{i,s}$ are compatible with a joint state ρ defined on the joint support. Thus, in general, $\mathcal{Q}(\mathfrak{S})_d \subset \mathcal{Q}(\mathfrak{S}_1)_d \times \mathcal{Q}(\mathfrak{S}_2)_d$.

[3, §I] **Example 94** (Overlapping Bell scenario). *Take $\mathfrak{S} = \{0, 1, \{0, 1\}\} \times [2] \times [2]$. This corresponds with two systems performing two measurements with two outcomes, and two additional global measurements with two outcomes. There are two maximal disjoint supports, the tuple $\sigma_1 = (\{0\}, \{1\})$ and the singleton $\sigma_2 = (\{0, 1\})$. Section $\mathfrak{S}_{\sigma_1}$ coincides with the Bell scenario discussed in Examples 92 and 93. Section $\mathfrak{S}_{\sigma_2}$ corresponds with a Bell scenario $B(1, 2, 2)$ with one party performing two measurements with two outcomes. Since σ_1 and σ_2 do not share any element, there are no additional consistency constraints between the sections, thus $\mathcal{P}(\mathfrak{S}) = \mathcal{P}(\mathfrak{S}_{\sigma_1}) \times \mathcal{P}(\mathfrak{S}_{\sigma_2})$ and $\mathcal{L}(\mathfrak{S}) = \mathcal{L}(\mathfrak{S}_{\sigma_1}) \times \mathcal{L}(\mathfrak{S}_{\sigma_2})$. Also, $\mathcal{Q}(\mathfrak{S}) = \mathcal{Q}(\mathfrak{S}_1) \times \mathcal{Q}(\mathfrak{S}_2)$, since the tensor product of models in $\mathfrak{S}_1$ and $\mathfrak{S}_2$ give a model in $\mathfrak{S}$. However, this is not possible when there are dimension constraints. Thus, we only have the inclusion $\mathcal{Q}_d(\mathfrak{S}) \subset \mathcal{Q}_d(\mathfrak{S}_{\sigma_1}) \times \mathcal{Q}_d(\mathfrak{S}_{\sigma_2})$.*

²Notice that the supports $S_i = \{s : (s, x, a) \in \mathfrak{S}_i\}$ may overlap as long as they do not share any measurement. That is, if $s \in S_1 \cap S_2$ and $(s, x, a) \in \mathfrak{E}_1$ then $(s, x, a) \notin \mathfrak{E}_2$.

³We take the one-dimensional state $\rho_{i,s} = 1$ when $s \notin S_i$ for the ease of notation.



3.3 Complete hierarchy

The sets $\mathcal{P}(\mathfrak{S})$, $\mathcal{Q}(\mathfrak{S})$ and $\mathcal{L}(\mathfrak{S})$ are convex and compact (closed and bounded) subsets of some finite dimensional real Euclidean space $V_{\mathfrak{S}}$. Thus, each set can be characterized with linear functionals on $V_{\mathfrak{S}}$. Namely, for each $p \in V_{\mathfrak{S}} - \mathcal{L}(\mathfrak{S})$ there exists a linear functional $\beta : V_{\mathfrak{S}} \rightarrow \mathbb{R}$ such that $\beta|_{\mathcal{L}(\mathfrak{S})} \geq 0$ and $\beta(p) < 0$. Linear witnesses for nonlocal models are typically called *Bell functionals*. The collection of such linear functionals is the dual space of $V_{\mathfrak{S}}$, denoted $V_{\mathfrak{S}}^*$.

Local probability models form a *polytope* and thus can be described with finitely many linear functionals. Characterizing quantum models is more complicated. However, there are complete semidefinite programming hierarchies to approach $\mathcal{Q}(\mathfrak{S})$ [2] and $\mathcal{Q}_d(\mathfrak{S})$ [92] in settings with no overlap. These were adapted in [3] to settings with overlap. Assuming no common events, $\mathcal{Q}(\mathfrak{S}) = \prod_{\sigma} \mathcal{Q}(\mathfrak{S}_{\sigma})$ and one can apply the hierarchies for settings with no overlap to each sector. However, for $\mathcal{Q}_d(\mathfrak{S})$ some extra effort is needed. We make use of the complete hierarchies proposed in Section 2.3.

Consider the setting $\mathfrak{S} = \{0, 1, 2, \{1, 2\}\} \times [2] \times [2]$ corresponding to three local algebras $A_1 \otimes A_2 \otimes A_3$ as in Example 79, with two dichotomic measurements over each factor and two more on the joint factor $A_2 \otimes A_3$. Take the alphabet $X = (A_{a|x}, B_{b|y}, C_{c|z}, D_{d|w})$, where all indices run through $\{0, 1\}$. Letters $A_{a|x}$, $B_{b|y}$ and $C_{c|z}$ correspond to measurement effects on the local factors A_1 , A_2 and A_3 while $D_{d|w}$ to measurement effects on $A_2 \otimes A_3$. The behaviour of projective measurement effects is captured with the polynomial constraints

$$0 = A_{a|x}^2 - A_{a|x}, \quad 0 = B_{b|y}^2 - B_{b|y}, \quad 0 = C_{c|z}^2 - C_{c|z}, \quad 0 = D_{d|w}^2 - D_{d|w}, \quad (3.7)$$

$$0 = 1 - \sum_a A_{a|x}, \quad 0 = 1 - \sum_b B_{b|y}, \quad 0 = 1 - \sum_c C_{c|z}, \quad 0 = 1 - \sum_d D_{d|w}, \quad (3.8)$$

and the local structure is relaxed with the commutativity relations

$$\begin{aligned} 0 &= A_{a|x} B_{b|y} - B_{b|y} A_{a|x}, \\ 0 &= A_{a|x} C_{c|z} - C_{c|z} A_{a|x}, \\ 0 &= B_{b|y} C_{c|z} - C_{c|z} B_{b|y}, \\ 0 &= A_{a|x} D_{d|w} - D_{d|w} A_{a|x}. \end{aligned} \quad (3.9)$$

There are two maximal sections σ_{123} with supports $(\{0\}, \{1\}, \{2\})$ and σ_{14} with supports $(\{0\}, \{1, 2\})$. A pair of models $p_{123} \in \mathcal{P}(\mathfrak{S}_{123})$ and $p_{14} \in \mathcal{P}(\mathfrak{S}_{14})$ is compatible on the intersections if $p_{123}^1(a|x) = p_{14}^1(a|x)$, where

$$\begin{aligned} p_{123}^1(a|x) &= \sum_{b,c} p_{123}(a, b, c|x, y, z), \\ p_{14}^1(a|x) &= \sum_d p_{14}(a, d|x, w). \end{aligned} \quad (3.10)$$

In such a case, $p = p_{123} \cup p_{14} \in \mathcal{P}(\mathfrak{S})$ is a model on the joint setting. Moreover, when $p_{123} \in \mathcal{Q}(\mathfrak{S}_{123})$ and $p_{14} \in \mathcal{Q}(\mathfrak{S}_{14})$, then $p \in \mathcal{Q}(\mathfrak{S})$. However, it is not the case in general that $p \in \mathcal{Q}_d(\mathfrak{S})$ when $p_{123} \in \mathcal{Q}_d(\mathfrak{S}_{123})$ and $p_{14} \in \mathcal{Q}_d(\mathfrak{S}_{14})$ for some dimensional constraints $d : \{0, 1, 2\} \rightarrow \mathbb{N} \cup \{\infty\}$. This is the case, if the dimensionally constrained semidefinite programming relaxations (2.36) are feasible. Namely,

$$\begin{aligned}
& \max && 0 \\
& \text{s. t.} && y_1 = 1, \\
& && y_{A_{a|x}B_{b|y}C_{c|z}} = p_{123}(a, b, c|x, y, z), \\
& && y_{A_{a|x}D_{d|w}} = p_{14}(a, d|x, w), \\
& && M \geq 0, \\
& && M \in \mathcal{M}(d).
\end{aligned} \tag{3.11}$$

Here, $\mathcal{M}(d)$ denotes the linear space spanned by moment matrices obtained sampling in dimension d until stabilized (see Section 2.3). For unconstrained dimension, these hierarchies characterize the algebraic relaxations of quantum behaviours that consider commuting operators instead of tensor factors.

3.4 Settings with shared components

Given the practical implications of special models in certain correlation settings, it is important from a technical side to understand when different applications can be performed while sharing certain components to reduce costs. Namely, if settings $\mathfrak{S}_1$ and $\mathfrak{S}_2$ have models giving advantages in two different tasks, are there models in $\mathfrak{S} = \mathfrak{S}_1 \cup \mathfrak{S}_2$ giving advantages in both tasks when some components are shared? This question was addressed in [3].

[Definition 87] Recall that $\mathfrak{E}_1$ is a sub-lattice of $\mathfrak{E}$, thus a model $p \in \mathcal{P}(\mathfrak{S})$ gives by restriction $p|_{\mathfrak{E}_1} \in \mathcal{P}(\mathfrak{S}_1)$ (assuming the measurements to have the same outcomes in $\mathfrak{S}$ and $\mathfrak{S}_1$). Moreover, when p is local or quantum so is the restriction. Similarly, any linear functional $\beta \in V_{\mathfrak{S}}^*$ gives by restriction $\beta|_{V_{\mathfrak{S}_1}} \in V_{\mathfrak{S}_1}^*$. Conversely, any $\beta \in V_{\mathfrak{S}_1}^*$ gives by extension with zeros $\bar{\beta} \in V_{\mathfrak{S}}^*$, and $\bar{\beta}|_{V_{\mathfrak{S}_1}} = \beta$.

Given $\beta \in V_{\mathfrak{S}_1}^*$ and $\gamma \in V_{\mathfrak{S}_2}^*$, the maximum value of $\bar{\beta} + \bar{\gamma}$ can be attained combining optimal strategies for β and γ . However, this is not possible with constraints on the dimension, and some *frustration* between optimal strategies may appear. Some examples are investigated in [3], using the hierarchies discussed in Section 3.3. This resembles the problem of frustration between local Hamiltonians, in which not all terms can be simultaneously minimized.

[3, §III] **Example 95** (Dimension witnesses). *Consider the setting $\mathfrak{S} = \mathfrak{S}_{12} \sqcup \mathfrak{S}_{13}$ with three local factors $A_1 \otimes A_2 \otimes A_3$, where each $\mathfrak{S}_{1s}$ is a Bell scenario $B(2, 2, 2)$*

between A_1 and A_s . Let β_{1s} be the CHSH functional on $\mathfrak{S}_{1s}$,

$$\begin{aligned} \beta_{1s} = & p_{1s}(00|00) - p_{1s}(01|00) - p_{1s}(10|00) + p_{1s}(11|00) \\ & + p_{1s}(00|01) - p_{1s}(01|01) - p_{1s}(10|01) + p_{1s}(11|01) \\ & + p_{1s}(00|10) - p_{1s}(01|10) - p_{1s}(10|10) + p_{1s}(11|10) \\ & - p_{1s}(00|11) + p_{1s}(01|11) + p_{1s}(10|11) - p_{1s}(11|11). \end{aligned} \quad (3.12)$$

The joint functional $\beta = \bar{\beta}_{12} + \bar{\beta}_{13}$ on $\mathfrak{S}$ can be used to discriminate the local dimension of the factor A_1 [3]. Indeed, optimizing β on the semidefinite programming relaxations 3.11 we obtain different values varying the dimension,

$$\begin{aligned} \beta(p) &\leq 4.8284\dots \quad \forall p \in \mathcal{Q}(\dim A_1 = 2), \\ \beta(p) &\leq 5.1428\dots \quad \forall p \in \mathcal{Q}(\dim A_1 = 3), \\ \beta(p) &\leq 5.6569\dots \quad \forall p \in \mathcal{Q}(\dim A_1 \geq 4). \end{aligned} \quad (3.13) \quad [3, \text{Table I}]$$

When $\dim A_1 = 2$, the maximal quantum value of β corresponds with the maximal quantum value of β_{12} and the maximal local value for β_{13} , which is a consequence of monogamy of entanglement [116]. When $\dim A_1 \geq 4$, the maximal quantum value corresponds with the maximal quantum value of both β_{12} and β_{13} , since A_1 has enough dimension to accommodate one qubit maximally entangled with A_2 and other with A_3 . When $\dim A_1 = 3$, the maximal quantum value is obtained numerically with relaxations (3.11). There is enough dimension to overcome monogamy, but still some frustration between individual optimal strategies for β_{12} and β_{13} .

Many questions about bipartite entanglement can be translated to Bell scenarios. For instance, whether a state is separable, entangled or even maximally entangled [23, 24] (also [15, §III.A]). In the same spirit, problems about multipartite entanglement can be translated to correlation settings with shared components, such as the existence of absolutely maximally entangled states.

Definition 96 (Absolutely maximal entanglement). *A multipartite quantum state, $\rho \in \otimes_{s \in S} \mathcal{B}(H_s)$, is absolutely maximally entangled if it is maximally entangled as a bipartite state across all bipartitions. Namely, if $S = A \sqcup B = A|B$, then ρ is maximally entangled in $\mathcal{B}(H_A) \otimes \mathcal{B}(H_B)$, where $H_X = \otimes_{s \in X} H_s$.* [117, §II.A]

Absolutely maximally entangled states have practical importance in quantum information science, since they are connected with optimal quantum error correction codes [117] and the holographic principle [118]. The existence of absolutely maximally entangled states can be answered from correlation settings. Indeed, maximal entanglement across a bipartition can be self-tested with a linear functional in a Bell scenario [25]. Thus, one just need to take the joint functional over the correlation setting combining all those Bell scenarios.

Example 97 (Four qubits). *Notice that maximal entanglement in bipartitions into size two subsets already imply maximal entanglement in the bipartitions into subsets of size one and three. For each bipartition $S \sqcup T = \{1, 2, 3, 4\}$*

into size two subsets, consider a Bell scenario $\mathfrak{S}_{S|T} = B(2, 2, 4)$ and a linear functional $\beta_{S|T}$ self-testing for maximal entanglement in dimension four using two measurements with four outcomes [25]. Then, the joint functional $\beta = \bar{\beta}_{12|34} + \bar{\beta}_{13|24} + \bar{\beta}_{14|23}$ self-tests for absolutely maximal entanglement in the correlation setting $\mathfrak{S} = \mathfrak{S}_{12|34} \sqcup \mathfrak{S}_{13|24} \sqcup \mathfrak{S}_{14|23}$. That is, the maximum quantum value of $\bar{\beta}_{12|34}$, $\bar{\beta}_{13|24}$ and $\bar{\beta}_{14|23}$ can only be simultaneously attained with an absolutely maximally entangled state of four qubits.

The maximal quantum value is given by the following polynomial optimization problem with overlapping support and dimensional constraints,

$$\begin{aligned} \max \quad & \bar{\beta}_{12|34} + \bar{\beta}_{13|24} + \bar{\beta}_{14|23} \\ \text{s. t.} \quad & S_{a|x}^2 - S_{a|x}^2 = \mathbf{0}, \\ & [S_{a|x}, T_{b|y}] = \mathbf{0} \\ & \mathbf{1} - \sum_a S_{a|x} = \mathbf{0}, \end{aligned} \quad S \sqcup T = \{1, 2, 3, 4\}, \quad (3.14)$$

where $S_{a|x}$ corresponds to the effect with outcome $a \in [4]$ of measurement $x \in [2]$ performed on subsystem $S \in \{\{1, 2\}, \{1, 3\}, \{1, 4\}, \{2, 3\}, \{2, 4\}\}$. The optimal value can be relaxed with the semidefinite programs discussed in Section 2.3.

There exists no absolutely maximally entangled state with four qubits [119, Theorem 1]. Thus, the maximal quantum value of $\bar{\beta}_{12|34}$, $\bar{\beta}_{13|24}$ and $\bar{\beta}_{14|23}$ can not be attained. However, it is not simple to witness such frustration in a semidefinite programming relaxation without a good choice for the subspace. The problem in Eq. (3.14) involves 36 non-commutative operators, so the number of monomials grows fast with the degree even with the projection constraints. Overcoming this computational limitations to exploit the practical applications of correlation settings in quantum information was one of the main motivations to explore the computational reductions discussed in Chapter 2.

3.5 Contextuality settings

A different approach to build measurement protocols can be found in the contextuality scenarios proposed in [120, 4], where the focus is on the compatibility between effects and not the local structure in their supports. A contextuality setting is specified by a graph $G = (V, E)$, whose vertices correspond with *outcomes* and whose edges with *measurements* [4, §2.2]. That is, all vertices in an edge are compatible in the sense that each of them is the possible outcome of a measurement. One can assign probabilities to the outcomes in different ways and obtain probability models. In this section, we recall contextuality settings in [4] to highlight how these differ with respect to correlation settings.

- [4, §2.4] **Definition 98** (Probability model). *A probability model p for the contextuality setting $G = (V, E)$ is a map $p : V \rightarrow [0, 1]$ with the normalization conditions*
- [4, §4.1] *$\sum_{v \in e} p(v) = 1$ for each $e \in E$. The model p is deterministic if $\text{ran } p \subset \{0, 1\}$ and local if it is a convex combination of deterministic models. Denote $\mathcal{P}(G)$ the collection of probability models and $\mathcal{L}(G)$ the collection of local ones.*

Outcome probabilities in quantum theory are obtained with *Born's rule*, and essentially only this way [113]. This justifies the following definition.

Definition 99 (Quantum model). *A model p is quantum if there exist a [4, §5.1] quantum state ρ and projectors P_v on a Hilbert space H such that $p(v) = \text{tr}(\rho P_v)$ for each $v \in V$ and $\sum_{v \in e} P_v = \mathbf{1}$ for each $e \in E$. We denote $\mathcal{Q}(G)$ the collection of quantum models for G .*

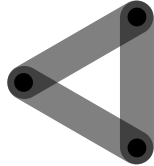
It is clear that $\mathcal{L}(G)$ and $\mathcal{P}(G)$ are convex. This is also the case for $\mathcal{Q}(G)$ if the Hilbert space dimension is unconstrained. Given $p, q \in \mathcal{Q}(G)$ with corresponding states ρ and σ and projectors P_v and Q_v , the state $\lambda\rho \oplus (1-\lambda)\sigma$ and the projectors $P_v \oplus Q_v$ give the model $\lambda p + (1-\lambda)q \in \mathcal{Q}(G)$. Moreover, we have the chain of inclusions $\mathcal{L}(G) \subset \mathcal{Q}(G) \subset \mathcal{P}(G)$. To prove the first inclusion it is enough to show that deterministic models are quantum, which follows from the assignment $P_v = p(v)\mathbf{1} \in \{0, \mathbf{1}\}$ and $\rho = \mathbf{1}$ in $H = \mathbb{C}$.

Example 100 (n -outcome measurement). *The hypergraph $G = ([n], \{[n]\})$ with n vertices and one hyperedge including all vertices corresponds with a measurement with n outcomes. A probability model $p \in \mathcal{P}(G)$ assigns a probability $p(v)$ to each outcome. All models in this setting are local, since we can write $p = \sum_v p(v)\delta_v$, where $\delta_v(u) = \delta_{v,u}$ is the deterministic model corresponding with outcome v . This is not the case when there are vertices shared among different edges.*



Notice that we assign to each vertex $v \in V$ a projector P_v and not a general positive effect E_v . With general effects we can reproduce any probability model $p \in \mathcal{P}(G)$ taking $E_v = p(v)\mathbf{1}$. However, quantum models with projectors are more restricted as the following examples shows.

Example 101 (Triangle setting). *Take $G = (\{1, 2, 3\}, \{\{1, 2\}, \{2, 3\}, \{1, 3\}\})$ [4, Example 2.4.2] with three edges of two vertices each, and one vertex is shared between two edges forming a triangle. The only probability model $p \in \mathcal{P}$ is the constant $p = 1/2$. There are no deterministic models, thus no local models $\mathcal{L} = \emptyset$. Moreover, there are no quantum models since there exist no projector P satisfying $P = \mathbf{1} - P$.*



Several properties of probability models on contextuality settings are captured by combinatorial properties of the graph defining the setting [120, 4].

[4, §2.3] **Definition 102** (Confusability graph). *The confusability or non-orthogonality graph of a contextuality setting $G = (V, E)$ is the graph $\Gamma_G = (V, E_\Gamma)$ where $(v, w) \in E_\Gamma$ when there is no edge $e \in E$ containing both v and w . That is, v and w are not effects of the same measurement and thus can be confused.*

In particular, the *confusability graph* already captures all local models [4, 4.2.1]. Namely, $p \in \mathcal{P}(G)$ is local $p \in \mathcal{L}(G)$ if and only if $\alpha^*(\Gamma_G, p) = 1$, where α^* is the *fractional packing number* of the graph. However, the confusability graph fails to capture all the properties of a contextuality setting needed to characterize quantum models [4, 5.1.3]. Indeed, there exist contextuality settings G and H with the same confusability graph $\Gamma_G = \Gamma_H$ such that $\mathcal{Q}(G) = \emptyset \neq \mathcal{Q}(H)$.

Example 103. *Let $G = ([3], \{\{1, 2\}, \{2, 3\}, \{1, 3\}\})$ be the triangle setting and $H = ([3], \{\{3\}\})$ the setting with one measurement with three effects. In both settings, each pair of vertices is contained in some edge. Thus, the confusability graph has no edges $\Gamma_G = ([3], \emptyset) = \Gamma_H$. Example 101 shows $\mathcal{Q}(G) = \emptyset$. On the other hand, $P_1 = \mathbf{1}$ and $P_2 = P_3 = \mathbf{0}$ gives a quantum model $p \in \mathcal{Q}(H)$.*

Remarkably, contextuality settings also include Bell scenarios [4, §3.4]. The local structure of measurements in a Bell scenario can be encoded in a contextuality setting. Clearly, every choice of local measurements x and y determines a joint measurement xy . Probability models of this contextuality scenario correspond with probability models in the Bell scenario. Moreover, non-signalling conditions can be imposed by a more general composition of local edges [121] that amount to joint measurements in which the choice of input of one party may depend on the output of another.

[4, §3.1] **Definition 104** (Foulis-Randall composition). *Take contextuality settings $G_A = (V_A, E_A)$ and $G_B = (V_B, E_B)$. The Foulis-Randall composition $G = G_A \otimes G_B$ has vertices $V = V_A \times V_B$ and edges $E = E_{A \rightarrow B} \cup E_{B \rightarrow A}$. Here $E_{A \rightarrow B}$ includes all joint measurements for which the input of B depends on the outcome of A .*

The normalization conditions on the measurements $E_{A \rightarrow B}$ give the non-signalling $B \not\rightarrow A$ on the probability models of $G_A \otimes G_B$. This is illustrated in example below.

[4, §3.4] cf. Example 92 **Example 105** (Bell scenario). *The Bell scenario $B(2, 2, 2)$ consists of two parties with two dichotomic measurements each. Probability models $\mathcal{P}(B)$ are bipartite conditional probabilities such that $\sum_{a, b \in \{0, 1\}} p(ab|xy) = 1$ for each $x, y \in \{0, 1\}$. These correspond with probability models in the contextuality setting with edges*

$$\begin{aligned} e_{00} &= \{00|00, 01|00, 10|00, 11|00\}, \\ e_{01} &= \{00|01, 01|01, 10|01, 11|01\}, \\ e_{10} &= \{00|10, 01|10, 10|10, 11|10\}, \\ e_{11} &= \{00|11, 01|11, 10|11, 11|11\}. \end{aligned}$$

This is the cartesian product of the settings $B_A(1, 2, 2)$ and $B_B(1, 2, 2)$ corresponding to the local measurements. That is, $\mathcal{P}(B_A \times B_B) = \mathcal{P}(B)$.

Now consider the joint measurement in which the first party perform local measurement $x = 0$ and obtain output $a \in \{0, 1\}$, then the other party performs measurement $y = a$ (similarly, $y = 1 - a = \bar{a}$). The corresponding edge is

$$e_{0a} = \{00|00, 01|00, 10|01, 11|01\}.$$

Normalization on e_{00} and e_{0a} impose $p(10|00) + p(11|00) = p(10|01) + p(11|01)$, so the marginals $p_1(0|0)$ and $p_1(1|0)$ are well defined. The contextuality setting with Foulis-Randall composition $B_A \otimes B_B$ includes all those edges

$$\begin{aligned} E_{A \rightarrow B} &= \{e_{00}, e_{01}, e_{10}, e_{11}, e_{0a}, e_{0\bar{a}}, e_{1a}, e_{1\bar{a}}\}, \\ E_{B \rightarrow A} &= \{e_{00}, e_{01}, e_{10}, e_{11}, e_{a0}, e_{\bar{a}0}, e_{a1}, e_{\bar{a}1}\}, \end{aligned}$$

giving all marginals $p_1(a|x)$ and $p_2(b|y)$. Thus, $\mathcal{P}(B_A \otimes B_B) = \mathcal{N}(B)$. Local models in Bell scenario are given by convex combinations of deterministic non-signalling models, thus $\mathcal{L}(B_A \otimes B_B) = \mathcal{L}(B)$.

Quantum models for contextuality settings are obtained with Born's rule, [4, §5.2]

$$p(ab|xy) = \text{tr}(\rho A_{a|x} B_{b|y}) \quad (3.15)$$

for some state ρ and projective measurements A_x and B_y on some Hilbert space H , such that $A_{a|x}$ and $B_{b|y}$ commute. Taking projectors $P_{ab|xy} = A_{a|x} B_{b|y}$ and state ρ , we see p is a quantum model for the contextuality setting $B_A \otimes B_B$. Conversely, given a state ρ and projectors $P_{ab|xy}$ determining a quantum model $p \in \mathcal{Q}(B_A \otimes B_B)$, the marginals $A_{a|x} = \sum_b P_{ab|xy}$ and $B_{b|y} = \sum_a P_{ab|xy}$ are well defined. Orthogonality conditions $P_{ab|xy} P_{a'b'|xy} = \delta_{aa'} \delta_{bb'} P_{ab|xy}$ imply $[A_{a|x}, B_{b|y}] = \mathbf{0}$ and $p(ab|xy) = \text{tr}(\rho A_{a|x} B_{b|y})$. Thus, $\mathcal{Q}(B_A \otimes B_B) = \mathcal{Q}(B)$.

Notice that, contrary to what happens in contextuality settings, effects in correlation settings, in particular Bell scenarios, can be taken to be general positive operators. Indeed, going to a Stinespring dilation [114] we can represent general effects as projectors. Thus $\mathcal{Q}_G(B) = \mathcal{Q}(B)$. However, we can get any model in a contextuality setting with general effects, $\mathcal{Q}_G(B_A \otimes B_B) = \mathcal{P}(B_A \otimes B_B) = \mathcal{N}(B)$. This discrepancy occurs because the argument in Example 92 showing that $\mathcal{Q}(B_A \otimes B_B) \subset \mathcal{Q}(B)$ relies on the orthogonality relations for projectors. This manifests a difference between correlation settings, which focus on the local structure of measurements, and contextuality settings, which focus on compatibility of deterministic outcomes assignments.

Notice that Foulis-Randall composition is not associative. That is, in general $G_A \otimes (G_B \otimes G_C) \neq (G_A \otimes G_B) \otimes G_C$. These different ways of composing multipartite settings come from the different ways in which the input of some parties may depend on the output of others. However, all choices lead to the same local, quantum and probability models [4, Theorem C.2.4] thus we denote their Foulis-Randall composition as $G_A \otimes G_B \otimes G_C$. This composition can be adapted to multipartite contextuality settings with shared components.

Example 106 (Composite setting with shared components). *Let G_A , G_{BC} , G_{CD} be contextuality settings on parties A , BC and CD respectively. We can take compositions $G_{A|BC} = G_A \otimes G_{BC}$ and $G_{A|CD} = G_A \otimes G_{CD}$. However, we can not take the composition $G_{BC} \otimes G_{CD}$ since measurements over BC and CD may not be compatible. Thus, we can define the most general composite setting to be $G = G_{A|BC} \cup G_{A|CD}$. Local, quantum and probability models in G are elements of the cartesian product of those models in $G_{A|BC}$ and $G_{A|CD}$.*

Composite contextuality settings corresponding to Bell scenarios with shared components coincide with correlations settings discussed in Section 3.2 when the dimension is unconstrained. In bounded dimension, models in correlation settings invoke the quantum marginal problem through the dimension constraints on the local algebras of operators. In contextuality settings, however, the constraints due to the local structure only appear through particular compositions of edges. Moreover, effects are assumed to be projectors, thus models subject to dimension constraints do not form a convex set.

3.6 Local Hamiltonians

[122, Definition 14.3] **Definition 107** (Local Hamiltonian). *A local Hamiltonian is a self-adjoint operator that decomposes as $h = \sum_{s \in S} h_s$, where h_s has support $H_s = \bigotimes_{i \in s} H_i$, for fixed H_i (typically finite). When $|s| \leq k$ for each $s \in S$, we say h is k -local.*

The lowest eigenvalue of h is called *ground energy*, and can be computed as $\min \sum_{s \in S} \text{tr}(\rho_s h_s)$ where the minimization runs over all marginals $(\rho_s)_{s \in S}$ compatible with a global state ρ . That is, each ρ_s is the reduction of ρ to subsystem s , obtained tracing out the remaining parties. When the elements of the support S are disjoint, all terms commute and can be jointly diagonalized. Thus the ground energy is the sum of the ground energies of each term h_s . In other words, the marginals are always compatible. However, when the supports in S have overlap some *frustration* appear between local ground energies. Finding the ground energy becomes hard, since deciding the compatibility between the marginals is QMA-complete [123]. The compatibility conditions can be fully characterized in some settings [124], or relaxed to obtain efficient approximations, for instance with entropy constraints [125]. Notice that the compatibility conditions between quantum marginals also appear in the consistency conditions on quantum probability models in correlation settings with overlap, characterized with the hierarchies introduced in Section 3.3. This provides necessary conditions for a collection of quantum marginals to be compatible with a joint quantum state.

Claim 108. *Let $(\rho_s)_{s \in S}$ be a collection of quantum marginals, each with dimension d_s , with supports S and compatible with a global state ρ . Let $\mathfrak{S}$ be a correlation setting with supports S . Then, for any choice of measurement effects $E_{s,a|x}$ the assignment $p(s, x, a) = \text{tr}(\rho_s E_{s,a|x})$ gives a model $p \in \mathcal{Q}_d(\mathfrak{S})$.*

Proof. It follows from Definition 90 of $\mathcal{Q}_d(\mathfrak{S})$. □

Most physical Hamiltonians can be characterized algebraically from their defining properties with a system of polynomial constraints. In these cases, the ground state energy corresponds to the solution of a polynomial optimization problem, as noted in [28, §5], and sum of squares decompositions have proven to be a useful technique for analyzing these Hamiltonians [126, 127, 83]. Namely, let $h = \sum_{s \in S} h_s$ be a local Hamiltonian whose terms are characterized by polynomial equality and inequality constraints, G and R , possibly including commutativity constraints that carry the local structure and matrix polynomial identities for given dimensions. Take the alphabet $X = (x_s)_{s \in S}$ and the polynomial $p = -\sum_{s \in S} x_s \in k\langle X \rangle$. The ground energy of h is (changing the sign) the optimal value of the polynomial optimization problem (p, G, R) on $k\langle X \rangle$.

Example 109 (Quantum Max-Cut). *The quantum max-cut Hamiltonian [82, §1.2], [83, §1] for a graph $G = ([n], E)$ acts on $H = (\mathbb{C}^2)^{\otimes n}$ is given by*

$$h(G) = \sum_{(i,j) \in E} (\mathbf{1} - X_i X_j - Y_i Y_j - Z_i Z_j)/2, \quad (3.16)$$

where X_i, Y_i and Z_i are the Pauli matrices acting on the local factor $H_i = \mathbb{C}^2$,

$$X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad Y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \quad (3.17)$$

There is an algebraic formulation in terms of transpositions $(ij) \in \mathcal{S}_n$ from [83, §2.3] the symmetric group of n elements [83, Proposition 2.3],

$$h(G) = \sum_{(i,j) \in E} 2(\mathbf{1} - (ij)), \quad (3.18)$$

since the representation π of $\mathcal{S}_n$ in $(\mathbb{C}^2)^{\otimes n}$ exchanging tensor factors, satisfies $\pi(ij) = (\mathbf{1} + X_i X_j + Y_i Y_j + Z_i Z_j)/2$. Thus, quantum max-cut is a polynomial optimization problem in partially commutative variables $(ij) \in \mathcal{S}_n$. The algebra of transpositions (see Example 75) is characterized with constraints

$$\begin{aligned} (ij)(ij) &= \mathbf{1}, \\ (ij)(jk) &= (ik)(ij), \\ (ij)(kl) &= (kl)(ij). \end{aligned} \quad (3.19)$$

The representation π above satisfies additional quadratic constraints

$$(ij)(jk) + (jk)(ij) = (ij) + (jk) + (ik) - \mathbf{1}, \quad (3.20)$$

which already imply the commutation relations. The ground energy of Eq. (3.16) is thus lower bounded with (changing the sign) the optimal value of the problem

$$\begin{aligned} \max \quad & \sum_{(i,j) \in E} 2((ij) - \mathbf{1}) \\ \text{s. t.} \quad & (ij)(ij) = \mathbf{1}, \\ & (ij)(jk) = (ik)(ij), \\ & (ij)(jk) + (jk)(ij) = (ij) + (jk) + (ik) - \mathbf{1}. \end{aligned} \quad (3.21)$$

Equality constraints can be implemented as replacement rules given a Gröbner basis. Thus, the optimal solution of Eq. (3.21) can be upper bounded with semidefinite programming relaxations of the form

$$\begin{aligned} \max \quad & \sum_{(i,j) \in E} 2(y_{(ij)} - 1) \\ \text{s. t.} \quad & y_{\mathbf{1}} = 1, \\ & M \geq 0, \\ & M \in \mathcal{M}. \end{aligned} \tag{3.22}$$

For the full graph with three elements, the first degree relaxation to the algebraic Hamiltonian in Eq. (3.18) with the constraints in Eq. (3.19) reads

$$\begin{aligned} \max \quad & 2(y_{(12)} + y_{(13)} + y_{(23)} - 3) \\ \text{s. t.} \quad & \begin{pmatrix} 1 & y_{(12)} & y_{(13)} & y_{(23)} \\ y_{(12)} & 1 & y_{(132)} & y_{(123)} \\ y_{(13)} & y_{(123)} & 1 & y_{(132)} \\ y_{(23)} & y_{(132)} & y_{(123)} & 1 \end{pmatrix} \geq 0. \end{aligned} \tag{3.23}$$

This has optimal value 0, which is already tight for Eq. (3.18) [83, Example 4.8]. Similarly, the first degree relaxation for constraints in Eq. (3.20) amounts to the additional constraint $y_{(123)} = -1 + y_{(12)} + y_{(13)} + y_{(23)} - y_{(132)}$, and again has tight optimal value 0 [83, Example 4.11]. This coincides with the ground energy of the max-cut Hamiltonian in Eq. (3.16), which can be obtained analytically for full graphs [83, Lemma 2.12] among others [83, §6].

Chapter 4

Uncertainty Relations

This Chapter, based on [5], investigates the problem of uncertainty relations. Uncertainty relations are arithmetical relations between operators that involve non-linear terms in the expectation values. Thus, we make use of the framework of state polynomial optimization [38] to tackle this problem. First, we introduce a complete hierarchy of semidefinite relaxations to obtain tight uncertainty relations. Next, we focus on the smallest non-trivial example, which are additive uncertainty relations involving second degree correlations. We start recalling known upper bounds for uncertainty relations of this kind [126, 128] based on a combinatorial quantity known as Lovász number [129]. Last, we propose an efficient subhierarchy of relaxations that can be understood as a generalization of the Lovász bound in different directions.

4.1 Tight uncertainty relations

Given self-adjoint bounded operators over a finite dimensional Hilbert space, $A_1, \dots, A_n \in \mathcal{B}(H)$, the tuples $(a_1, \dots, a_n)$ that can be obtained as expectation values $a_i = \langle A_i \rangle = \lambda(A_i)$ for some state $\lambda \in \mathcal{S}(H)$ [Definition 28] are characterized by a semidefinite program of size $\dim H$. In finite dimension, states correspond with density matrices, $\rho \in \mathcal{B}(H)$ satisfying $\rho \geq 0$ and $\text{tr} \rho = 1$, via $\lambda_\rho(A) = \text{tr}(A\rho)$ for $A \in \mathcal{B}(H)$. Then, the semidefinite program reads

$$\begin{aligned} \max \quad & 0 \\ \text{s. t.} \quad & \text{tr}(A_i \rho) = a_i, \\ & \text{tr}(\rho) = 1, \\ & \rho \geq 0. \end{aligned} \tag{4.1}$$

When the operators are not given in a fixed representation $A_1, \dots, A_n \in \mathcal{B}(H)$ but specified by some polynomial constraints $S \subset k\langle X \rangle$, where X is the alphabet with one letter per operator, the tuples of compatible expectation values can be characterized with a polynomial optimization problem $(\mathbf{0}, \emptyset, S)$ in $k\langle X \rangle$.

The semidefinite programming relaxations of moments read

$$\begin{aligned}
 \max \quad & 0 \\
 \text{s. t.} \quad & y_{\mathbf{1}} = 1, \\
 & y_{x_i} = a_i, \\
 & M \geq 0, M \in \mathcal{M}, \\
 & M^s \geq 0, M^s \in \mathcal{M}^s \quad \forall s \in S.
 \end{aligned} \tag{4.2}$$

[Section 1.6] Recall that $\mathcal{M}$ in Eq. (1.28) denotes the linear space of moment matrices and $\mathcal{M}^s$ in Eq. (1.30) the linear space of moment matrices localized at $s \in S$.

Non-commutative polynomials, however, fail to characterize higher order expressions in the expectation values. For instance, expectation values of Pauli matrices form a unit sphere, called *Bloch ball*, that is in correspondence with two dimensional quantum states [35, §5.2]. Thus, the expectations satisfy

$$[\text{35, §5.2}] \quad \langle \sigma_x \rangle^2 + \langle \sigma_y \rangle^2 + \langle \sigma_z \rangle^2 \leq 1 \tag{4.3}$$

and each tuple $(\langle \sigma_x \rangle, \langle \sigma_y \rangle, \langle \sigma_z \rangle)$ uniquely determines a state in $\mathcal{B}(\mathbb{C}^2)$. Moreover, relation (4.3) readily follows from the anti-commutativity between Pauli operators and their normalization (see Example 110) and thus hold for any collection of unitary mutually anti-commuting operators in any dimension.

Another paradigmatic example is given by Robertson relation [130],

$$[\text{7, Box 2.4}] \quad (\langle A^2 \rangle - \langle A \rangle^2)(\langle B^2 \rangle - \langle B \rangle^2) \geq |\langle AB - BA \rangle|^2/4. \tag{4.4}$$

for any pair of (not necessarily bounded) hermitian operators A and B over H . In particular, the *position* and *momentum* operators, X and P , have commutator $XP - PX = i\hbar$ and we obtain Heisenberg uncertainty principle,

$$[\text{7, Box 2.4}] \quad (\langle X^2 \rangle - \langle X \rangle^2)(\langle P^2 \rangle - \langle P \rangle^2) \geq \hbar^2/4. \tag{4.5}$$

Interestingly, bounds in Eqs. (4.3) and (4.4) do not make use of the actual representation of the operators. That is, they exploit an algebraic relation between the operators, anti-commutation for Eq. (4.3) and the commutator for Eq. (4.4), and thus hold for any other representation of operators satisfying those relations. In the same spirit, state polynomial optimization can be used to derive the best possible bound for any polynomial expression involving expectation values that can be obtained with operators just from their algebraic relations. Moreover, GNS construction provides theoretically, or in practice if flatness condition in the state analogue of Lemma 48 is met, bounded operators on a separable Hilbert space attaining the optimal value.

[Definition 40, Definition 34] Recall a state polynomial optimization problem (p, G, S) in $k[\mathbb{S}_\rho(X)]$ is

$$\begin{aligned}
 \max \quad & \lambda \circ \psi(p) \\
 \text{s. t.} \quad & \psi(g) = \mathbf{0} \quad \forall g \in G, \\
 & \psi(s) \geq \mathbf{0} \quad \forall s \in S.
 \end{aligned} \tag{4.6}$$

The maximization runs over all $*$ -representations $\psi : k\langle \mathbb{S}_\varrho(X) \rangle \rightarrow \mathcal{B}(H)$ and states $\lambda \in \mathcal{S}(H)$ compatible with the state symbol. Thus, the sum of squares or moment relaxations give efficient upper bounds for the uncertainty relations described by $p \in k\langle \mathbb{S}_\varrho(X) \rangle$ that hold for any operators satisfying the prescribed relations. These relaxations are semidefinite programs over a finite subspace $V \subset k\langle \mathbb{S}_\varrho(X) \rangle$, and in the limit $V = k\langle \mathbb{S}_\varrho(X) \rangle$, or when the flatness condition in the state analogue of Lemma 48 is met, the optimal solution to the polynomial optimization problem (4.6) is attained.

As we discussed in Section 1.6, there is no available procedure to choose the subspace V efficiently for a general polynomial optimization problem. Polynomial optimization is NP-hard already for quadratic constraints [63]. The general approach is to consider the subspaces V_d spanned by all monomials up to degree d . However, the number of state monomials grows very fast with the degree and number of letters. Thus, in order to obtain good bounds one may need to explore very large subspaces. In general, not all elements in V contribute to the problem at hand. For instance, sparsity techniques discussed in Section 2.4 exploit additional structure in the problem to discard terms that do not contribute, thus reducing the size of the subspace in the relaxation. In next section we design effective subspaces [5] that provide good bounds for a selected class of uncertainty relations.

4.2 Additive uncertainty relations

Given a finite graph $G = (X, E, \zeta, *)$ with edge weights $\zeta_{uv} \in k$ for $(u, v) \in E$ and involution $*$ on the vertices, consider the alphabet X , the state polynomial $p_X = \sum_{u \in X} \varrho(u) \varrho(u^*)$, and the collection I of polynomial constraints $g_u = uu^* - \mathbf{1}$ (unitarity) and $g_{vw} = vw - \zeta_{vw} wv$ (commutation rules) for $u \in X$ and $(v, w) \in E$. The polynomial optimization problem $(p, I, \emptyset)$ on $k\langle \mathbb{S}_\varrho(X) \rangle$ reads

$$\begin{aligned} \beta(G) = \max \quad & \sum_{u \in X} \varrho(u) \varrho(u^*) \\ \text{s. t.} \quad & uu^* - \mathbf{1} = \mathbf{0} \quad \forall u \in X, \\ & uv - \zeta_{uv} wv = \mathbf{0} \quad \forall (u, v) \in E. \end{aligned} \tag{4.7} \quad [5, \text{Equation 3}]$$

These uncertainty relations, studied in [126, 128, 131], include the relation for Pauli matrices in Eq. (4.3) and already find several applications in spectroscopy and atomic clocks [36], and quantum metrology [37]. Considerable effort has been dedicated to derive state-independent bounds [132, 133]. Here, we present the approach based on state polynomial optimization proposed in [5].

Notice that unitarity constraints the weights to satisfy $|\zeta_{uv}| = 1$ for each $(u, v) \in E$, since for any representation of u, v with operators $U, V \in \mathcal{B}(H)$ we have $1 = \|U\| = |\zeta_{uv}| \|VUV^*\| \leq |\zeta_{uv}|$ and $1 \geq \|UVU^*\| = |\zeta_{uv}| \|V\| = |\zeta_{uv}|$. In *hermitian* settings, i.e. the involution is the identity, this means $\zeta_{uv} \in \{-1, 1\}$. When G is a full graph, $k\langle X \rangle / \langle I \rangle$ is a 2^n -dimensional vector space over k spanned by all subsets of letters, where $n = |X|$ is the number of letters.

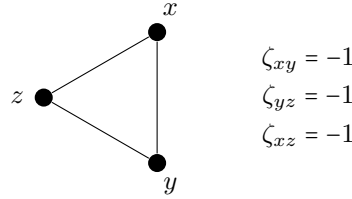
Example 110 (Pauli matrices). *The graph G corresponding to three hermitian, unitary anti-commuting operators has three vertices $X = (x, y, z)$ with $\zeta_{uv} = -1$ and $v^* = v$ for all $u, v \in X$. Problem (4.7) reads*

$$\begin{aligned} \max \quad & \varrho(x)^2 + \varrho(y)^2 + \varrho(z)^2 \\ \text{s.t.} \quad & x^2 = \mathbf{1}, \quad yz = -zy, \\ & y^2 = \mathbf{1}, \quad zx = -xz, \\ & z^2 = \mathbf{1}, \quad xy = -yx, \end{aligned} \tag{4.8}$$

and has optimal value 1, which is certified by the sum of squares decomposition

$$1 - \varrho(x)^2 - \varrho(y)^2 - \varrho(z)^2 = \varrho([\mathbf{1} - x\varrho(x) - y\varrho(y) - z\varrho(z)]^2). \tag{4.9}$$

To see this, first notice that the moments can be taken to be real, since the problem is invariant under involution. Then, $\varrho(xy) = \varrho(xy)^* = \varrho(yx) = -\varrho(xy) = 0$. This recovers a representation-free version of relation $\langle \sigma_x \rangle^2 + \langle \sigma_y \rangle^2 + \langle \sigma_z \rangle^2 \leq 1$, since Pauli matrices satisfy all constraints in Eq. (4.8). The bound is tight since it matches the lower bound obtained with the one dimensional representation, $\lambda \circ \psi : k\langle \mathbb{S}_\varrho(X) \rangle \rightarrow \mathbb{C}$, $(x, y, z) \mapsto (1, 0, 0)$. Alternatively, for the two dimensional representation with Pauli matrices, it matches the value obtained with an eigenstate of σ_x , which have zero expectation value for σ_y and σ_z .



Example 110 shows that we can obtain lower bounds from the one dimensional representation assigning the value one to an independent set (totally disconnected) and zero to the other vertices, already noticed in [126, 131, 5]. Moreover, sum of squares decompositions give upper bounds. For the graph with three unitary anti-commuting operators it is enough to consider the four dimensional subspace $V \subset k\langle \mathbb{S}_\varrho(X) \rangle$ spanned by $(\mathbf{1}, x\varrho(x), y\varrho(y), z\varrho(z))$. The dual of the sum of squares decomposition, the moment relaxation, reads

$$1 = \max \quad a + b + c \tag{4.10}$$

$$\text{s.t.} \quad \begin{pmatrix} 1 & a & b & c \\ a & a & 0 & 0 \\ b & 0 & b & 0 \\ c & 0 & 0 & c \end{pmatrix} \geq 0. \tag{4.11}$$

This is the *Lovász number* of the graph $G = (\{x, y, z\}, \{(x, y), (x, z), (y, z)\})$.

Definition 111 (Lovász number). *The Lovász number of a graph $G = (X, E)$ is the solution to any of the following equivalent semidefinite programs*

$$\begin{aligned} \vartheta(G) = \min_M \quad & \lambda_{\max}(M) \\ \text{s.t.} \quad & M_{v,v} = 1, \\ & M_{u,v} = 1 \quad (u, v) \notin E \\ & M \geq 0. \end{aligned} \tag{4.12} \quad [129, \text{Theorem 3}]$$

$$\begin{aligned} \vartheta(G) = \max_M \quad & \sum_{u,v \in X} M_{u,v} \\ \text{s.t.} \quad & \text{tr } M = 1, \\ & M_{u,v} = 0 \quad (u, v) \in E \\ & M \geq 0, \end{aligned} \tag{4.13} \quad [129, \text{Theorem 4}]$$

$$\begin{aligned} \vartheta(G) = \max_M \quad & \sum_{v \in X} M_{v,v} \\ \text{s.t.} \quad & M_{v,v} = a_v, \\ & M_{u,v} = 0 \quad (u, v) \in E \\ & \begin{pmatrix} 1 & a^T \\ a & M \end{pmatrix} \geq 0, \end{aligned} \tag{4.14} \quad [134, \S 2.2, \text{SDP2}]$$

Lovász number is sandwiched by two important graph invariants [135],

$$\alpha(G) \leq \vartheta(G) \leq \chi(\bar{G}). \tag{4.15}$$

The *independence number*, $\alpha(G)$, is the maximum number of pairwise independent vertices of G . The *chromatic number*, $\chi(G)$, is the minimum number of different colors needed to assign to the vertices of G , such that no connected vertices share the same color. These quantities, called *graph invariants*, encode key combinatorial properties of G and are NP-complete [136]. Here, $\bar{G} = (X, \bar{E})$ is the *complement* of $G = (X, E)$, with $(u, v) \in \bar{E}$ if and only if $(u, v) \notin E$.

Lovász number has already been linked to nonlocality, contextuality, and quantum zero-error communication [120, 4, 137, 138]. Remarkably, it also gives upper bounds for the uncertainty relation $\beta(G)$ in Eq.(4.7) for hermitian full graphs, as already noticed in [126, 128, 131]. We can recover Lovász bound from state polynomial optimization with a sum of squares relaxation [5].

Lemma 112 (Lovász bound). *Let $G = (X, E, \zeta, \mathbf{1})$ be the fully connected hermitian graph with edge weights $\zeta_{uv} \in \{-1, 1\}$. The moment relaxation corresponding to the indexing sequence $\mathcal{I} = (\mathbf{1}, v\rho(v) : v \in X)$ is the Lovász number of the anti-commutativity graph $G' = (X, E')$, where $(u, v) \in E'$ when $\zeta_{uv} = -1$.*

Proof. Take the matrix of moments M indexed with $(\mathbf{1}, v\rho(v) : v \in X)$. First, $M_{v\rho(v), v\rho(v)} = \rho(v)^2 = M_{\mathbf{1}, v\rho(v)}$ since $vv = \mathbf{1}$. We can assume M to be real,

since the problem is invariant under involution. Then, when $\zeta_{uv} = -1$ we find that $M_{u\varrho(u),v\varrho(v)} = uv\varrho(u)\varrho(v) = -vu\varrho(u)\varrho(v) = -M_{v\varrho(v),u\varrho(u)} = 0$, because M is symmetric. Those are precisely the edges in the anti-commutativity graph, thus we can write the moment relaxation as

$$\begin{aligned} \max_M \quad & \sum_{v \in X} M_{v\varrho(v),v\varrho(v)} \\ \text{s. t.} \quad & M_{\mathbf{1},\mathbf{1}} = 1, \\ & M_{v\varrho(v),v\varrho(v)} = M_{\mathbf{1},v\varrho(v)}, \\ & M_{u\varrho(u),v\varrho(v)} = 0 \quad (u,v) \in E' \\ & M \geq 0, \end{aligned} \tag{4.16}$$

which coincides with the Lovász number of $G' = (X, E')$ in Eq. (4.14). $\square$

Lemma 112 gives a new algebraic definition for the Lovász number. Let $G = (X, E)$ be a graph over the alphabet X and I the collection of polynomials

$$\begin{aligned} g_u &= uu - \mathbf{1}, \\ g_{uv} &= uv - vu \quad (u,v) \notin E, \\ g_{uv} &= uv + vu \quad (u,v) \in E. \end{aligned} \tag{4.17}$$

The *Lovász number* ϑ is the infimum of the real numbers such that

$$\vartheta - \sum_v \varrho(v)\varrho(v) = \sum_i \varrho(f_i^2), \tag{4.18}$$

is a sum of squares in $k\langle \mathbb{S}_\varrho(X) \rangle / \langle I \rangle$ for some polynomials f_i in the linear span of $(\mathbf{1}, v\varrho(v) : v \in X)$. Moreover, the sum of squares relaxation is not limited to hermitian graphs or restricted to commutation and anti-commutation relations. Similar bounds can be obtained for arbitrary graphs. This motivates an analogue for the Lovász number in more general settings.

Definition 113 (Algebraic Lovász number). *Let $G = (X, E, \zeta, *)$ be a graph over the alphabet X with edge weights $\zeta_{uv} \in k$ for $(u,v) \in E$ and involution. Take the state polynomial $p_X = \sum_{v \in X} \varrho(v)\varrho(v^*)$, and the collection I of the polynomials $g_u = uu^* - \mathbf{1}$ and $g_{vw} = vw - \zeta_{vw}wv$ for $u \in X$ and $(v,w) \in E$. The Lovász number is the solution to the semidefinite programming relaxation to the polynomial optimization problem $(p, I, \emptyset)$ in $k\langle \mathbb{S}_\varrho(X) \rangle$ to the subspace spanned by $(\mathbf{1}, v\varrho(v^*) : v \in X)$. That is,*

$$\begin{aligned} \vartheta(G) &= \max \sum_{v \in X} M_{v\varrho(v^*),v\varrho(v^*)} \\ \text{s. t.} \quad & M_{\mathbf{1},\mathbf{1}} = 1, \\ & M_{u^*\varrho(u),v\varrho(v^*)} = \zeta_{uv}M_{v^*\varrho(v),u\varrho(u^*)}, \\ & M_{v\varrho(v^*),v\varrho(v^*)} = M_{\mathbf{1},v\varrho(v^*)}, \\ & M \geq 0. \end{aligned} \tag{4.19}$$

Equivalently, the smallest real number such that

$$\vartheta - p_X = \sum_i \varrho(f_i f_i^*) \quad (4.20)$$

in $k\langle \mathbb{S}_\varrho(X) \rangle / \langle I \rangle$, for some f_i in the linear span of $(\mathbf{1}, v\varrho(v^*) : v \in X)$.

Example 114 (Non-hermitian Lovász). Take $\omega = e^{i\pi\alpha}$ with $\alpha \in [0, 1]$ and unitary operators X and Y with commutation relation $\omega XY = YX$, which already implies $\omega^* X^* Y = Y X^*$, $\omega^* X Y^* = Y^* X$ and $\omega X^* Y^* = Y^* X^*$. The algebraic Lovász number takes the form

$$\begin{aligned} \vartheta(\omega) = \max \quad & a + b + a + b \\ \text{s. t.} \quad & \begin{pmatrix} 1 & a & b & a & b \\ a & a & c & d & e \\ b & c^* & b & \omega^* e & f \\ a & d^* & \omega e^* & a & \omega^* c^* \\ b & e^* & f^* & \omega c & b \end{pmatrix} \geq 0. \end{aligned} \quad (4.21)$$

The optimal values in Figure 4.1 show how the algebraic definition of Lovász number is sensitive to complex phases in the commutation relations. Indeed, for $1/2 \leq \alpha \leq 1$, the algebraic Lovász number gives the tight uncertainty relation

$$|\langle X \rangle|^2 + |\langle Y \rangle|^2 \leq 1. \quad (4.22)$$

This value can be obtained with the one dimensional representation, assigning value one to X and X^* and value zero to Y and Y^* . It is analogous to the Lovász bound for anti-commuting self-adjoint operators, $\langle X \rangle^2 + \langle Y \rangle^2 \leq 1$. For $\alpha = 0$, we also have the tight uncertainty relation

$$|\langle X \rangle|^2 + |\langle Y \rangle|^2 \leq 2. \quad (4.23)$$

This value can be again obtained with the one dimensional representation, assigning the value one to all letters. It is analogous to the Lovász bound for commuting self adjoint operators, $\langle X \rangle^2 + \langle Y \rangle^2 \leq 2$. However, for $0 < \alpha < 1/2$, the algebraic Lovász number interpolates between these two cases.

Lovász bound gives the first level of the hierarchies proposed in [5].

4.3 Lovász hierarchy

The sum of squares decomposition corresponding to the Lovász bound is restricted to the subspace $V = \text{span}(\mathbf{1}, v\varrho(v^*) : v \in X)$. This bound was tight for the Pauli matrices in Example 110, however this is not always the case, as already noticed in [126, 128]. We can make these bounds stronger expanding this subspace, and in the limit $V = k\langle \mathbb{S}_\varrho(X) \rangle$, or when the flatness condition in the state analogue of Lemma 48 is met, we attain the tight value $\beta(G)$ for the uncertainty relation in the polynomial optimization problem 4.7.

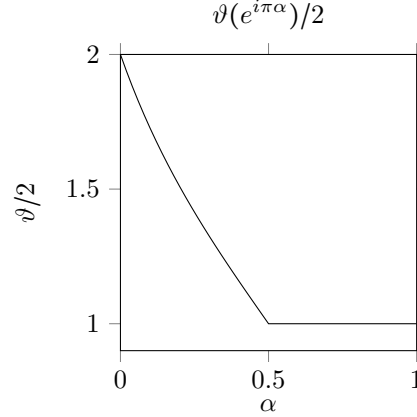
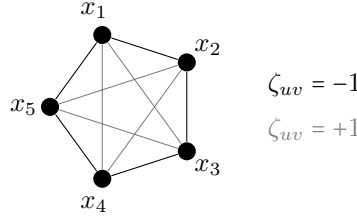


Figure 4.1: **Algebraic Lovász** number in Definition 113 for non-hermitian graphs with two vertices and their adjoints, with commutation $e^{i\pi\alpha}XY = YX$.

[126, Remark 4.9] **Example 115** (Pentagon graph). *Let $G = (X, E, \zeta, \mathbf{1})$ be a full graph with edge weights $\zeta_{uv} \in \{-1, 1\}$ and whose anti-commutativity graph is a pentagon. By Lemma 112, the algebraic Lovász number is the Lovász number of the anti-commutativity graph, giving the bound $\beta(G) \leq \vartheta(G) = \sqrt{5}$. Enlarging the space in the sum of squares decomposition with the products $(uv\rho(u)\rho(v) : u, v \in X)$, we obtain a stronger bound, $\beta(G) \leq 2$. This value is tight, since it matches the independence number lower bound obtained with the one dimensional representation assigning one to a pair of commuting letters and zero to all others.*

[5, Table II]



In particular, taking products of the monomials in the Lovász indexing sequence we obtain good efficient bounds in practice [5]. This motivates the following definition.

[5, §III] **Definition 116** (Lovász hierarchy). *Given a subset $S = \{v_1, \dots, v_d\} \subset X$ denote $w_S = v_1 \dots v_d \rho(v_1^*) \dots \rho(v_d^*)$ and set $w_\emptyset = \mathbf{1}$. For each $d \leq |X|$, let V_d be the subspace spanned by monomials $\mathcal{I}_d = (w_S : S \subset X, |S| \leq d)$. We call the sum of squares relaxation corresponding to the subspaces V_d for $d = 1, \dots, |X|$, with optimal values $\vartheta_d(G)$, the level d of the Lovász hierarchy.*

The moment relaxation for d -th level of Lovász hierarchy reads

$$\begin{aligned} \vartheta_d(G) = \max_M \quad & \sum_{v \in X} M_{1,v\varrho(v)} \\ \text{s.t.} \quad & M_{u,v} = cM_{x,y} \quad \forall \varrho(u^*v) = c\varrho(x^*y), \\ & M_{1,1} = 1, \\ & M \geq 0. \end{aligned} \quad (4.24)$$

For an alphabet with $n = |X|$ letters, the size of the d -th level of Lovász hierarchy is the number of subsets of X with up to d elements. Thus, first level has size $1 + n$, second level has size $1 + n + n(n-1)/2$ and last level has size 2^n . Lovász hierarchy improves upon upper bounds for all 1292 additive uncertainty relations on up to nine operators for which a tight bound is not known [5]. Moreover, if the flatness condition in Lemma 48 is met, we can guarantee the hierarchy attained the optimal value. However, it is an open question whether Lovász hierarchy is complete for the uncertainty relation $\beta(G)$ in problem 4.7.

4.4 Applications in quantum information

Last, we illustrate some interesting applications in quantum information for the uncertainty relation β in problem 4.7. In particular, entanglement and non-locality characterization, as already proposed in [128, 5].

Example 117 (Entanglement witness). *The linear operator for two qubits* [5, Appendix A]

$$W = \mathbf{1} \otimes \mathbf{1} - \sigma_x \otimes \sigma_x - \sigma_y \otimes \sigma_y + \sigma_z \otimes \sigma_z, \quad (4.25)$$

gives an entanglement witness. It follows from Cauchy-Schwarz inequality and the uncertainty bound in Example 110 that $\langle W \rangle \geq 0$ for all separable states,

$$\langle W \rangle \geq 1 - \langle \sigma_x \rangle^2 - \langle \sigma_y \rangle^2 - \langle \sigma_z \rangle^2 \geq 0. \quad (4.26)$$

Whereas a joint eigenstate of $\sigma_x \otimes \sigma_x$, $\sigma_y \otimes \sigma_y$, and $\sigma_z \otimes \sigma_z$ can achieve $\langle W \rangle = -2$.

Similar constraints involving Pauli strings bound the quantum value of Bell inequalities and Bell monogamy for dichotomic measurements [139].

Example 118 (Bell inequalities). *Let $\mathcal{L}$ be a Bell inequality with two measurements with two outcomes between two parties. Normalizing the value over classical correlations as $\mathcal{L} \leq 1$, the optimal quantum value is obtained by the value in problem 4.7 for the following Pauli strings* [5, Appendix A]

$$\mathcal{L}^2 \leq \langle \sigma_x \otimes \sigma_x \rangle^2 + \langle \sigma_x \otimes \sigma_y \rangle^2 + \langle \sigma_y \otimes \sigma_x \rangle^2 + \langle \sigma_y \otimes \sigma_y \rangle^2 \leq 2. \quad (4.27)$$

Already the first level of the hierarchy, Lovász bound in Lemma (112), recovers Tsirelson's bound for CHSH inequality. Similarly, the tight uncertainty bound

$$\begin{aligned} & \langle \sigma_x \otimes \sigma_x \otimes \mathbf{1} \rangle^2 + \langle \sigma_x \otimes \sigma_y \otimes \mathbf{1} \rangle^2 + \langle \sigma_y \otimes \sigma_x \otimes \mathbf{1} \rangle^2 + \langle \sigma_y \otimes \sigma_y \otimes \mathbf{1} \rangle^2 \\ & + \langle \sigma_x \otimes \mathbf{1} \otimes \sigma_x \rangle^2 + \langle \sigma_x \otimes \mathbf{1} \otimes \sigma_y \rangle^2 + \langle \sigma_y \otimes \mathbf{1} \otimes \sigma_x \rangle^2 + \langle \sigma_y \otimes \mathbf{1} \otimes \sigma_y \rangle^2 \leq 2 \end{aligned} \quad (4.28)$$

recovers strong Bell monogamy $\mathcal{L}_{AB}^2 + \mathcal{L}_{AC}^2 \leq 2$ for two CHSH inequalities [140].

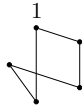
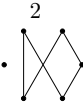
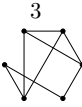
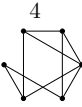
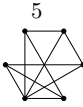
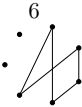
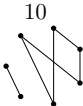
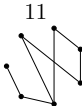
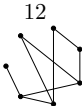
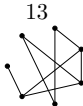
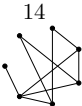
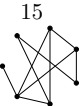
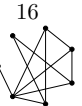
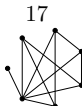
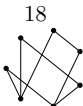
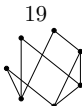
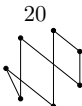
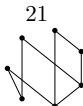
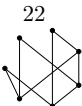
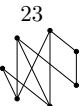
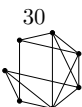
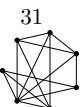
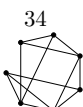
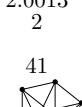
								
ϑ_1	2.2361	3.2361	2.2361	2.2361	2.2361	4.2361	3.2361	3.2361
ϑ_2	2.0000	3.0000	2.0000	2.0000	2.0000	4.0000	3.0000	3.0000
α	2	3	2	2	2	4	3	3
								
ϑ_1	3.2361	3.2361	3.2361	3.2361	3.2361	3.2361	3.2361	3.2361
ϑ_2	3.0000	3.0000	3.0000	3.0000	3.0000	3.0000	3.0000	3.0000
α	3	3	3	3	3	3	3	3
								
ϑ_1	3.2361	3.1966	3.0642	3.3177	3.2361	3.1966	3.2361	3.1966
ϑ_2	3.0000	3.0000	3.0000	3.0000	3.0000	3.0000	3.0000	3.0000
α	3	3	3	3	3	3	3	3
								
ϑ_1	2.2361	2.2361	2.2361	2.2361	2.2361	2.2361	2.2361	2.2361
ϑ_2	2.0000	2.0000	2.0000	2.0000	2.0000	2.0000	2.0000	2.0000
α	2	2	2	2	2	2	2	2
								
ϑ_1	2.2361	2.2361	2.2361	2.2361	2.2361	2.2361	2.2361	2.2361
ϑ_2	2.0363	2.0056	2.0085	2.0033	2.0249	2.0392	2.0121	2.0024
ϑ_3	2.0067	2.0004	2.0017	2.0000	2.0047	2.0052	2.0006	2.0000
ϑ_7	2.0013	2.0000	2.0003	2.0000	2.0014	2.0011	2.0002	2.0000
α	2	2	2	2	2	2	2	2
								
ϑ_1	2.2361	2.2361	2.1099					
ϑ_2	2.0910	2.0000	2.0950					
ϑ_3	2.0076	2.0000	2.0938					
ϑ_7	2.0024	2.0000	2.0938					
α	2	2	2					

Table 4.1: **Hierarchy of upper bounds** for all non-isomorphic graphs up to seven vertices for which $\beta(G) < \vartheta_1(G)$ [5, Table IV]. Lovász hierarchy $\vartheta_d(G)$ in Eq. (4.24) improves on Lovász bound closing the gap (up to numerical precision) for all but 7 graphs: 33, 35, 37, 38, 39, 41 and 43.

The bound with Pauli matrices in Eq. (4.27) relies on the equivalence between *anti-commutativity* ($AB = -BA$) and *complementarity* (for any state, $\langle A \rangle = 1$ implies $\langle B \rangle = 0$ and viceversa $\langle B \rangle = 1$ implies $\langle A \rangle = 0$) [139]. It is an interesting open question whether Example 118 extends to measurements with more than two outcomes. For measurements with d outcomes one could replace Pauli matrices by non-hermitian operators with outcomes $\{1, \omega, \dots, \omega^{d-1}\}$, where $d = e^{2\pi i/d}$ is the d -th root of unity. Examples of these, are the Heisenberg-Weyl operators.

Definition 119 (Heisenberg-Weyl). *Heisenberg-Weyl operators generalize [35, §12.1] Pauli matrices to Hilbert spaces with dimension d . These are defined as*

$$\sigma(k, l) |i\rangle = X^k Z^l |i\rangle = \omega^{il} |i + k\rangle, \quad (4.29)$$

where $X |j\rangle = |j + 1\rangle$ and $Z |j\rangle = \omega^j |j\rangle$ are called shift and clock. These form a basis of the space of $d \times d$ matrices and verify $\sigma(k, l)^d = \mathbf{1}$ and $\langle \mathbf{1}, \sigma(k, l) \rangle = 0$.

Commutation relations and adjoints for Heisenberg-Weyl operators are

$$\sigma(k, l) \sigma(m, n) = \omega^{lm - kn} \sigma(m, n) \sigma(k, l), \quad (4.30) \quad [5, \text{Appendix E}]$$

$$\sigma(k, l)^* = \omega^{kl} \sigma(d - k, d - l). \quad (4.31)$$

These matrices are a d dimensional representation of the generalized Clifford algebra [141] with generators satisfying (i) $x_i^d = \mathbf{1}$ and (ii) $x_j * x_i = \omega x_i * x_j$. First considered by Sylvester to solve the matrix equation $PX = XQ$ [142], these matrices were later rediscovered by Weyl as generators of finite dimensional quantum dynamics [143]. Since then, they found several applications in finite dimensional quantum mechanics [144]. Lovász hierarchy also produces efficient upper bounds for these uncertainty relations involving non-hermitian operators.

Example 120 (Heisenberg-Weyl). *Heisenberg-Weyl operators are Fourier conjugates. For instance, in dimension three the discrete Fourier transform F , and Heisenberg-Weyl operators X and Z introduced in Eq. (4.29) look like*

$$F = \begin{pmatrix} 1 & 1 & 1 \\ 1 & \omega & \omega^2 \\ 1 & \omega^2 & \omega \end{pmatrix}, \quad X = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}, \quad Z = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \omega & 0 \\ 0 & 0 & \omega^2 \end{pmatrix}, \quad [35, \text{§12.1 \& §12.3}]$$

where $\omega = e^{2\pi i/3}$ is the third root of unity. These are related via $X = FZF^*$. [144, §3.2] Indeed, this relation holds in any dimension d , which readily follows from the relation $1 + \omega + \dots + \omega^{d-1} = 0$. The commutation relations between these Heisenberg-Weyl operators are $\omega XZ = ZX$. Thus, these define precisely the non-hermitian graph discussed in Example 114. Recall that the algebraic Lovász number yields [Definition 113] the tight uncertainty relation $|\langle X \rangle|^2 + |\langle Y \rangle|^2 \leq 1$ when $\omega = e^{i\pi\alpha}$ and $1/2 \leq \alpha \leq 1$. This is the case for dimensions $d = 2, 3, 4$. However, for $d = 5$, the algebraic Lovász number yields the bound $|\langle X \rangle|^2 + |\langle Y \rangle|^2 \leq 1.15838\dots$ Increasing the dimension we approach the limit of commuting operators, $\omega = 1$, for which the algebraic Lovász gives the tight bound $|\langle X \rangle|^2 + |\langle Y \rangle|^2 \leq 2$.

Similar bounds can be obtained for more than two operators. For instance, let $X = \sigma(1,0)$, $Y = \sigma(1,1)$ and $Z = (0,1)$ be Heisenberg-Weyl operators in dimension d and $\omega = e^{2\pi i/d}$ the d -th root of unity. Relations in Eq. (4.30) are

$$\omega XY = YX, \quad \omega XZ = ZX, \quad \omega YZ = ZY. \quad (4.32)$$

Again, algebraic Lovász number gives the tight uncertainty relation

$$|\langle X \rangle|^2 + |\langle Y \rangle|^2 + |\langle Z \rangle|^2 \leq 1, \quad (4.33)$$

for dimensions $d = 2, 3$. However, in dimension $d = 4$ the third operator gives a higher value in the algebraic Lovász,

$$|\langle X \rangle|^2 + |\langle Y \rangle|^2 + |\langle Z \rangle|^2 \leq 1.26795... \quad (4.34)$$

Increasing the dimension, one approaches the limit of commuting operators for which the algebraic Lovász also gives a tight uncertainty relation

$$|\langle X \rangle|^2 + |\langle Y \rangle|^2 + |\langle Z \rangle|^2 \leq 3. \quad (4.35)$$

This value is obtained with the trivial one dimensional representation.

Notice that the dimension of the Heisenberg-Weyl operators enters the problem through the commutation relations in Eq. (4.30), which depend on the d -th root of unity. However, the bounds obtained with Lovász hierarchy hold for any representation of such relations without other assumptions on the dimension.

Chapter 5

Conclusions

The framework of polynomial optimization finds several applications, among many others, in the field of quantum information, both to address fundamental questions and to design protocols with practical consequences. This thesis summarizes techniques from polynomial optimization and discuss applications that can be beneficial for researchers in the field.

On the one hand, we made use of non-commutative polynomial optimization to characterize quantum correlations in *correlation settings* [Definition 86], which generalize Bell scenarios to account for measurements with overlapping supports. These scenarios allow to translate several problems in quantum information to the language of nonlocality (see Example 97). In particular, correlations in simple settings can be used to discriminate the physical dimension of subsystems [Example 95].

On the other hand, we make use of the recently developed framework of state polynomial optimization to characterize uncertainty relations between operators satisfying prescribed algebraic constraints [Section 4.1]. For additive uncertainty relations we provide efficient approximations that both improve on and extend the bounds given by the Lovász number of the anti-commutativity graph. These techniques can be used to design witnesses for entanglement [Example 117] and Bell monogamy [Example 118] and provide a novel algebraic interpretation of the Lovász number in terms of sum of squares decompositions [Definition 113] that is applicable to non-hermitian settings [Example 120].

In both chapters, techniques from polynomial optimization based on sum of squares decompositions produce efficient approximations to quantum behaviours with semidefinite programs [Section 1.6]. The main inconvenience is that there exist no general criteria to choose small linear subspaces of polynomials giving good approximations. Thus, more elaborate applications may not be readily available since obtaining good bounds may require exploring very large subspaces that soon become intractable. This highlights the practical importance of exploiting additional structure to reduce the computational cost of the relaxations [Chapter 2]. In particular, we propose computational tools tailored to exploit the tantamount structure of partial commutations [Section 2.2].

Bibliography

- [1] Moisés Bermejo Morán, Abhishek Mishra, and Stefano Pironio. Polynomial optimization with partially commutative variables. *In preparation*, 2024.
- [2] Miguel Navascués, Stefano Pironio, and Antonio Acín. A convergent hierarchy of semidefinite programs characterizing the set of quantum correlations. *New Journal of Physics*, 10(7):073013, 2008.
- [3] Moisés Bermejo Morán, Alejandro Pozas-Kerstjens, and Felix Huber. Bell inequalities with overlapping measurements. *Physical Review Letters*, 131:080201, Aug 2023.
- [4] Antonio Acín, Tobias Fritz, Anthony Leverrier, and Ana Belén Sainz. A combinatorial approach to nonlocality and contextuality. *Communications in Mathematical Physics*, 334:533–628, 2015.
- [5] Moisés Bermejo Morán and Felix Huber. Uncertainty relations from state polynomial optimization. *Physical Review Letters*, 132:200202, May 2024.
- [6] Ola Bratteli and Derek William Robinson. *Operator algebras and quantum statistical mechanics: Volume 1: C^* - and W^* -Algebras. Symmetry Groups. Decomposition of States*. Springer Science & Business Media, 2012.
- [7] Michael A Nielsen and Isaac L Chuang. *Quantum computation and quantum information*. Cambridge university press, 2010.
- [8] Giacomo Mauro D’Ariano, Giulio Chiribella, and Paolo Perinotti. *Quantum theory from first principles: an informational approach*. Cambridge University Press, 2017.
- [9] Hans F de Groote. Observables II: Quantum observables. *arXiv preprint math-ph/0509075*, 2005.
- [10] Chris J. Isham. Topos methods in the foundations of physics. *Deep beauty*, pages 187–206, 2011.

- [11] Daniel Gottesman. Fault-tolerant quantum computation with higher-dimensional systems. In *NASA International Conference on Quantum Computing and Quantum Communications*, pages 302–313. Springer, 1998.
- [12] John F Clauser, Michael A Horne, Abner Shimony, and Richard A Holt. Proposed experiment to test local hidden-variable theories. *Physical Review Letters*, 23(15):880, 1969.
- [13] Simon Gröblacher, Tomasz Paterek, Rainer Kaltenbaek, Časlav Brukner, Marek Żukowski, Markus Aspelmeyer, and Anton Zeilinger. An experimental test of non-local realism. *Nature*, 446(7138):871–875, 2007.
- [14] John S Bell. On the Einstein-Podolsky-Rosen paradox. *Physics Physique Fizika*, 1(3):195, 1964.
- [15] Nicolas Brunner, Daniel Cavalcanti, Stefano Pironio, Valerio Scarani, and Stephanie Wehner. Bell nonlocality. *Reviews of modern physics*, 86(2):419, 2014.
- [16] Andrew C Doherty, Yeong-Cherng Liang, Ben Toner, and Stephanie Wehner. The quantum moment problem and bounds on entangled multi-prover games. In *2008 23rd Annual IEEE Conference on Computational Complexity*, pages 199–210. IEEE, 2008.
- [17] Stephen Wiesner. Conjugate coding. *ACM Sigact News*, 15(1):78–88, 1983.
- [18] Charles H Bennett and Gilles Brassard. Quantum cryptography: Public key distribution and coin tossing. *Theoretical computer science*, 560:7–11, 2014.
- [19] Richard Cleve and Harry Buhrman. Substituting quantum entanglement for communication. *Physical Review A*, 56(2):1201, 1997.
- [20] Stefano Pironio, Antonio Acín, Serge Massar, A Boyer de La Giroday, Dzmitry N Matsukevich, Peter Maunz, Steven Olmschenk, David Hayes, Le Luo, T Andrew Manning, et al. Random numbers certified by Bell’s theorem. *Nature*, 464(7291):1021–1024, 2010.
- [21] Ryszard Horodecki, Paweł Horodecki, Michał Horodecki, and Karol Horodecki. Quantum entanglement. *Reviews of Modern Physics*, 81(2):865, 2009.
- [22] Reinhard F Werner. Quantum states with Einstein-Podolsky-Rosen correlations admitting a hidden-variable model. *Physical Review A*, 40(8):4277, 1989.
- [23] Francesco Buscemi. All entangled quantum states are nonlocal. *Physical Review Letters*, 108(20):200401, 2012.

- [24] Dominic Mayers and Andrew Yao. Self testing quantum apparatus. *Quantum Information & Computation*, 4(4):273–286, 2004.
- [25] Alexia Salavrakos, Remigiusz Augusiak, Jordi Tura, Peter Wittek, Antonio Acín, and Stefano Pironio. Bell inequalities tailored to maximally entangled states. *Physical Review Letters*, 119(4):040402, 2017.
- [26] Sandu Popescu and Daniel Rohrlich. Quantum nonlocality as an axiom. *Foundations of Physics*, 24:379–385, 1994.
- [27] Boris S Cirel’son. Quantum generalizations of Bell’s inequality. *Letters in Mathematical Physics*, 4:93–100, 1980.
- [28] Stefano Pironio, Miguel Navascués, and Antonio Acín. Convergent relaxations of polynomial optimization problems with noncommuting variables. *SIAM Journal on Optimization*, 20(5):2157–2180, 2010.
- [29] Miguel Navascués, Adrien Feix, Mateus Araújo, and Tamás Vértesi. Characterizing finite-dimensional quantum behavior. *Physical Review A*, 92(4):042117, 2015.
- [30] Christopher A Fuchs and Asher Peres. Quantum-state disturbance versus information gain: Uncertainty relations for quantum information. *Physical Review A*, 53(4):2038, 1996.
- [31] Otfried Gühne. Characterizing entanglement via uncertainty relations. *Physical Review Letters*, 92(11):117903, 2004.
- [32] Eric G Cavalcanti and Margaret D Reid. Uncertainty relations for the realization of macroscopic quantum superpositions and EPR paradoxes. *Journal of Modern Optics*, 54(16-17):2373–2380, 2007.
- [33] Joseph M Renes and Jean-Christian Boileau. Conjectured strong complementary information tradeoff. *Physical Review Letters*, 103(2):020402, 2009.
- [34] Daniel Gottesman. An introduction to quantum error correction and fault-tolerant quantum computation. In *Quantum information science and its contributions to mathematics, Proceedings of Symposia in Applied Mathematics*, volume 68, pages 13–58, 2010.
- [35] Ingemar Bengtsson and Karol Życzkowski. *Geometry of quantum states: an introduction to quantum entanglement*. Cambridge university press, 2017.
- [36] Anders S. Sørensen and Klaus Mølmer. Entanglement and extreme spin squeezing. *Physical Review Letters*, 86:4431–4434, May 2001.
- [37] Géza Tóth and Florian Fröwis. Uncertainty relations with the variance and the quantum Fisher information based on convex decompositions of density matrices. *Physical Review Research*, 4:013075, Jan 2022.

- [38] Igor Klep, Victor Magron, Jurij Volčič, and Jie Wang. State polynomials: positivity, optimization and nonlinear Bell inequalities. *Mathematical Programming*, pages 1–47, 2023.
- [39] Serge Lang. *Algebra*, volume 211. Springer Science & Business Media, 2012.
- [40] Ryszard Engelking. General topology. *Sigma series in pure mathematics*, 6, 1989.
- [41] John B Conway. *A course in functional analysis*, volume 96. Springer, 2019.
- [42] Israel Gelfand and Mark Neumark. On the imbedding of normed rings into the ring of operators in Hilbert space. *Математический сборник*, 12(2):197–217, 1943.
- [43] Irving Ezra Segal. Irreducible representations of operator algebras. *Bulletin of the American Mathematical Society*, 53:73–88, 1947.
- [44] Raymond A Ryan. *Introduction to tensor products of Banach spaces*, volume 73. Springer, 2002.
- [45] Nicolas Bourbaki. *Topological vector spaces*. Springer Science & Business Media, 2013.
- [46] Nathaniel Patrick Brown and Narutaka Ozawa. *C^* -Algebras and Finite-Dimensional Approximations*, volume 88. American Mathematical Society, 2008.
- [47] Bruce Blackadar. *Operator algebras: theory of C^* -algebras and von Neumann algebras*, volume 122. Springer Science & Business Media, 2006.
- [48] Igor Klep, Victor Magron, and Jurij Volčič. Optimization over trace polynomials. In *Annales Henri Poincaré*, volume 23, pages 67–100. Springer, 2022.
- [49] Masamichi Takesaki et al. *Theory of operator algebras I*. Springer, 1979.
- [50] Yurii Nesterov and Arkadii Nemirovskii. *Interior-point polynomial algorithms in convex programming*. SIAM, 1994.
- [51] Stephen P Boyd and Lieven Vandenberghe. *Convex optimization*. Cambridge university press, 2004.
- [52] Narendra Karmarkar. A new polynomial-time algorithm for linear programming. In *Proceedings of the sixteenth annual ACM symposium on Theory of computing*, pages 302–311, 1984.
- [53] Lieven Vandenberghe and Stephen Boyd. Semidefinite programming. *SIAM Review*, 38(1):49–95, 1996.

- [54] Jean B Lasserre. Global optimization with polynomials and the problem of moments. *SIAM Journal on optimization*, 11(3):796–817, 2001.
- [55] Ronald V Book and Friedrich Otto. *String-rewriting systems*. Springer, 1993.
- [56] Yang Dandan and Victoria Gould. On graph products of monoids. *Journal of Algebra*, 620:113–156, 2023.
- [57] Volker Diekert and Yves Métivier. *Partial Commutation and Traces*, pages 457–533. Springer Berlin Heidelberg, Berlin, Heidelberg, 1997.
- [58] Claudio Procesi. The invariant theory of $n \times n$ matrices. *Advances in Mathematics*, 19(3):306–381, 1976.
- [59] Birgit Reinert. *On Gröbner Bases in Monoid and Group Rings*. PhD thesis, Technische Universität Kaiserslautern, 1995.
- [60] Victor Magron and Jie Wang. *Sparse polynomial optimization: theory and practice*. World Scientific, 2023.
- [61] Mihai Putinar. Positive polynomials on compact semi-algebraic sets. *Indiana University Mathematics Journal*, 42(3):969–984, 1993.
- [62] J William Helton and Scott McCullough. A Positivstellensatz for non-commutative polynomials. *Transactions of the American Mathematical Society*, 356(9):3721–3737, 2004.
- [63] Sartaj Sahni. Computationally related problems. *SIAM Journal on computing*, 3(4):262–279, 1974.
- [64] Pablo A Parrilo. Exploiting algebraic structure in sum of squares programs. In *Positive polynomials in control*, pages 181–194. Springer, 2005.
- [65] Bruno Buchberger. Bruno Buchberger’s PhD thesis 1965: An algorithm for finding the basis elements of the residue class ring of a zero dimensional polynomial ideal. *Journal of symbolic computation*, 41(3-4):475–511, 2006.
- [66] Teo Mora. An introduction to commutative and noncommutative Gröbner bases. *Theoretical Computer Science*, 134(1):131–173, 1994.
- [67] Cédric Bamps and Stefano Pironio. Sum-of-squares decompositions for a family of clausser-horne-shimony-holt-like inequalities and their application to self-testing. *Phys. Rev. A*, 91:052111, May 2015.
- [68] Klaus Madlener and Birgit Reinert. Computing Gröbner bases in monoid and group rings. In *Proceedings of the 1993 international symposium on Symbolic and algebraic computation*, pages 254–263, 1993.

- [69] Arjeh M Cohen. Non-commutative polynomial computations. *Verfügbar unter <http://www.win.tue.nl/~amc/pub/grobner/gbnp.pdf>* (Zugriff am 22.08. 2014), 2007.
- [70] Jean-Charles Faugère. A new efficient algorithm for computing Gröbner bases (F4). *Journal of pure and applied algebra*, 139(1-3):61–88, 1999.
- [71] Jean-Charles Faugère. A new efficient algorithm for computing Gröbner bases without reduction to zero (F5). In *Proceedings of the 2002 international symposium on Symbolic and algebraic computation*, pages 75–83, 2002.
- [72] Francis Sowerby Macaulay. Some formulae in elimination. *Proceedings of the London Mathematical Society*, 1(1):3–27, 1902.
- [73] Ana Maria Mandache. The Gröbner basis algorithm and subresultant theory. In *Proceedings of the international symposium on Symbolic and algebraic computation*, pages 123–128, 1994.
- [74] Alexander Maletsky. Gröbner bases and Macaulay matrices in Isabelle/HOL. *Formal Aspects of Computing*, 2024. Under consideration for publication.
- [75] Manuela Wiesinger-Widi. *Gröbner bases and generalized Sylvester matrices/ingereicht von: Manuela Wiesinger-Widi*. PhD thesis, Linz, 2015.
- [76] Matthias Aschenbrenner and Anton Leykin. Degree bounds for Gröbner bases in algebras of solvable type. *Journal of Pure and Applied Algebra*, 213(8):1578–1605, 2009.
- [77] Xingqiang Xiu. *Non-commutative Gröbner bases and applications*. PhD thesis, Universität Passau, 2012.
- [78] Pierre Cartier and Dominique Foata. Applications probabilistes. *Problèmes combinatoires de commutation et réarrangements*, pages 72–83, 1969.
- [79] Antoni Mazurkiewicz. Concurrent program schemes and their interpretations. *DAIMI Report Series*, (78), 1977.
- [80] Volker Diekert and Grzegorz Rozenberg. *The book of traces*. World scientific, 1995.
- [81] Christine Duboc. On some equations in free partially commutative monoids. *Theoretical Computer Science*, 46:159–174, 1986.
- [82] Sevag Gharibian and Ojas Parekh. Almost optimal classical approximation algorithms for a quantum generalization of max-cut. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2019)*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2019.

- [83] Adam Bene Watts, Anirban Chowdhury, Aidan Epperly, J William Helton, and Igor Klep. Relaxations and exact solutions to quantum max cut via the algebraic structure of swap operators. *Quantum*, 8:1352, May 2024.
- [84] António Veloso da Costa. Graph products of monoids. In *Semigroup Forum*, volume 63, pages 247–277. Springer, 2001.
- [85] Konrad Schmüdgen. *Unbounded self-adjoint operators on Hilbert space*, volume 265. Springer Science & Business Media, 2012.
- [86] Zhengfeng Ji, Anand Natarajan, Thomas Vidick, John Wright, and Henry Yuen. $MIP^* = RE$. *Communications of the ACM*, 64(11):131–138, 2021.
- [87] Isaac Goldbring. The Connes embedding problem: A guided tour. *Bulletin of the American Mathematical Society*, 59(4):503–560, 2022.
- [88] Volkher B Scholz and Reinhard F Werner. Tsirelson’s problem. *arXiv preprint math-ph/0812.4305*, 2008.
- [89] Vesselin Drensky. Polynomial identities for 2×2 matrices. *Topics in Computational Algebra*, pages 137–161, 1990.
- [90] Edward Formanek. *The polynomial identities and invariants of $n \times n$ matrices*. Number 78. American Mathematical Society, 1991.
- [91] Vesselin Drensky and Edward Formanek. *Polynomial identity rings*. Birkhäuser, 2012.
- [92] Miguel Navascués and Tamás Vértesi. Bounding the set of finite dimensional quantum correlations. *Physical Review Letters*, 115(2):020501, 2015.
- [93] Gilbert W Stewart. *Matrix algorithms: volume 1: basic decompositions*. SIAM, 1998.
- [94] Etienne De Klerk. Exploiting special structure in semidefinite programming: A survey of theory and applications. *European Journal of Operational Research*, 201(1):1–10, 2010.
- [95] Jie Wang and Victor Magron. Exploiting term sparsity in noncommutative polynomial optimization. *Computational Optimization and Applications*, 80(2):483–521, 2021.
- [96] Igor Klep, Victor Magron, and Janez Povh. Sparse noncommutative polynomial optimization. *Mathematical Programming*, pages 1–41, 2022.
- [97] Karin Gatermann and Pablo A Parrilo. Symmetry groups, semidefinite programs, and sums of squares. *Journal of Pure and Applied Algebra*, 192(1-3):95–128, 2004.

- [98] Christine Bachoc, Dion C Gijswijt, Alexander Schrijver, and Frank Vallentin. *Invariant semidefinite programs*, pages 219–269. Springer, 2012.
- [99] Frank Permenter and Pablo A Parrilo. Selecting a monomial basis for sums of squares programming over a quotient ring. In *2012 IEEE 51st IEEE Conference on Decision and Control (CDC)*, pages 1871–1876. IEEE, 2012.
- [100] Bruce Reznick. Extremal PSD forms with few terms. *Duke Mathematical Journal*, 45:363–374, 1978.
- [101] Bernd Sturmfels. Polynomial equations and convex polytopes. *The American mathematical monthly*, 105(10):907–922, 1998.
- [102] Sabine Burgdorf, Igor Klep, and Janez Povh. *Optimization of polynomials in non-commuting variables*, volume 2. Springer, 2016.
- [103] Jean-Pierre Serre. *Linear representations of finite groups*, volume 42. Springer, 1977.
- [104] Frank Permenter and Pablo A Parrilo. Dimension reduction for semidefinite programs via Jordan algebras. *Mathematical Programming*, 181:51–84, 2020.
- [105] Jacques Faraut and Adam Korányi. *Analysis on symmetric cones*. Oxford university press, 1994.
- [106] Gilles Brassard, Harry Buhrman, Noah Linden, André Allan Méthot, Alain Tapp, and Falk Unger. Limit on nonlocality in any world in which communication complexity is not trivial. *Physical Review Letters*, 96(25):250401, 2006.
- [107] Noah Linden, Sandu Popescu, Anthony J Short, and Andreas Winter. Quantum nonlocality and beyond: limits from nonlocal computation. *Physical Review Letters*, 99(18):180502, 2007.
- [108] Marcin Pawłowski, Tomasz Paterek, Dagomir Kaszlikowski, Valerio Scarani, Andreas Winter, and Marek Żukowski. Information causality as a physical principle. *Nature*, 461(7267):1101–1104, 2009.
- [109] Miguel Navascués and Harald Wunderlich. A glance beyond the quantum model. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 466(2115):881–890, 2010.
- [110] Tobias Fritz, Ana Belén Sainz, Remigiusz Augusiak, J Bohr Brask, Rafael Chaves, Anthony Leverrier, and Antonio Acín. Local orthogonality as a multipartite principle for quantum correlations. *Nature communications*, 4(1):2263, 2013.
- [111] Yuqian Zhou, Yu Cai, Jean-Daniel Bancal, Fei Gao, and Valerio Scarani. Many-box locality. *Physical Review A*, 96(5):052108, 2017.

- [112] Markus Frembs and Andreas Döring. Characterization of nonsignaling bipartite correlations corresponding to quantum states. *Physical Review A*, 106(6):062420, 2022.
- [113] Andrew M Gleason. Measures on the closed subspaces of a Hilbert space. In *The Logico-Algebraic Approach to Quantum Mechanics: Volume I: Historical Evolution*, pages 123–133. Springer, 1975.
- [114] W Forrest Stinespring. Positive functions on C^* -algebras. *Proceedings of the American Mathematical Society*, 6(2):211–216, 1955.
- [115] Sandu Popescu. Nonlocality beyond quantum mechanics. *Nature Physics*, 10(4):264–270, 2014.
- [116] Valerie Coffman, Joydip Kundu, and William K Wootters. Distributed entanglement. *Physical Review A*, 61(5):052306, 2000.
- [117] Dardo Goyeneche, Daniel Alsina, José I Latorre, Arnau Riera, and Karol Życzkowski. Absolutely maximally entangled states, combinatorial designs, and multiunitary matrices. *Physical Review A*, 92(3):032316, 2015.
- [118] Fernando Pastawski, Beni Yoshida, Daniel Harlow, and John Preskill. Holographic quantum error-correcting codes: Toy models for the bulk/boundary correspondence. *Journal of High Energy Physics*, 2015(6):1–55, 2015.
- [119] Atsushi Higuchi and Anthony Sudbery. How entangled can two couples get? *Physics Letters A*, 273(4):213–217, 2000.
- [120] Adán Cabello, Simone Severini, and Andreas Winter. Graph-theoretic approach to quantum correlations. *Physical Review Letters*, 112(4):040401, 2014.
- [121] David J Foulis and Charles H Randall. Empirical logic and tensor products. In *Interpretations and foundations of quantum theory*. Bibliographisches Institut, 1981.
- [122] Alexei Yu Kitaev, Alexander Shen, and Mikhail N Vyalyi. *Classical and quantum computation*. Number 47. American Mathematical Society, 2002.
- [123] Yi-Kai Liu. *Consistency of Local Density Matrices Is QMA-Complete*, page 438. Springer Berlin Heidelberg, 2006.
- [124] David A Mazziotti. Structure of fermionic density matrices: Complete N -representability conditions. *Physical Review Letters*, 108(26):263002, 2012.
- [125] Hamza Fawzi, Omar Fawzi, and Samuel O Scalet. Entropy constraints for ground energy optimization. *Journal of Mathematical Physics*, 65(3), 2024.

- [126] 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.
- [127] Matthew B Hastings. Perturbation theory and the sum of squares. *arXiv preprint arXiv:2205.12325*, 2022.
- [128] Carlos de Gois, Kiara Hansenne, and Otfried Gühne. Uncertainty relations from graph theory. *Physical Review A*, 107(6):062211, 2023.
- [129] László Lovász. On the Shannon capacity of a graph. *IEEE Transactions on Information Theory*, 25(1):1–7, 1979.
- [130] Howard Percy Robertson. The uncertainty principle. *Physical Review*, 34(1):163, 1929.
- [131] Zhen-Peng Xu, René Schwonnek, and Andreas Winter. Bounding the joint numerical range of Pauli strings by graph parameters. *PRX Quantum*, 5(2):020318, 2024.
- [132] Lars Dammeier, René Schwonnek, and Reinhard F Werner. Uncertainty relations for angular momentum. *New Journal of Physics*, 17(9):093046, 2015.
- [133] Konrad Szymański and Karol Życzkowski. Geometric and algebraic origins of additive uncertainty relations. *Journal of Physics A: Mathematical and Theoretical*, 53(1):015302, 2019.
- [134] Laura Galli and Adam N Letchford. On the Lovász theta function and some variants. *Discrete Optimization*, 25:159–174, 2017.
- [135] Donald E Knuth. The sandwich theorem. *The Electronic Journal of Combinatorics*, 1(1), 1994.
- [136] Richard M Karp. *Reducibility among Combinatorial Problems*, pages 85–103. Springer US, Boston, MA, 1972.
- [137] Antonio Acín, Runyao Duan, David E. Roberson, Ana Belén Sainz, and Andreas Winter. A new property of the Lovász number and duality relations between graph parameters. *Discrete Applied Mathematics*, 216:489–501, jan 2017.
- [138] Runyao Duan, Simone Severini, and Andreas Winter. Zero-error communication via quantum channels, noncommutative graphs, and a quantum Lovász number. *IEEE Transactions on Information Theory*, 59(2):1164–1174, 2013.
- [139] Paweł Kurzyński, Tomasz Paterek, Ravishankar Ramanathan, Wiesław Laskowski, and Dagomir Kaszlikowski. Correlation complementarity yields Bell monogamy relations. *Physical Review Letters*, 106(18):180402, 2011.

- [140] Benjamin Toner and Frank Verstraete. Monogamy of Bell correlations and Tsirelson's bound. *arXiv preprint quant-ph/0611001*, 2006.
- [141] Ashley O Morris. On a generalized Clifford algebra. *The Quarterly Journal of Mathematics*, 18(1):7–12, 1967.
- [142] James Joseph Sylvester. Thoughts on inverse orthogonal matrices, simultaneous signsuccessions, and tessellated pavements in two or more colours, with applications to Newton's rule, ornamental tile-work, and the theory of numbers. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 34(232):461–475, 1867.
- [143] Hermann Weyl. Quantenmechanik und Gruppentheorie. *Zeitschrift für Physik*, 46(1):1–46, 1927.
- [144] Apostolos Vourdas. Quantum systems with finite Hilbert space. *Reports on Progress in Physics*, 67(3):267, 2004.