First-order optimality conditions for non-commutative optimization problems

Mateus Araújo,[1,*] Igor Klep,[2,†] Andrew J. P. Garner,[3,‡] Tamás Vértesi,[4,§] and Miguel Navascués[3,¶]

[1] *Departamento de Física Teórica, Atómica y Óptica, Laboratory for Disruptive Interdisciplinary Science (LaDIS), Universidad de Valladolid, 47011 Valladolid, Spain*
[2] *Faculty of Mathematics and Physics, University of Ljubljana & Famnit, University of Primorska, Koper & Institute of Mathematics, Physics and Mechanics, Ljubljana, Slovenia*
[3] *Institute for Quantum Optics and Quantum Information (IQOQI) Vienna Austrian Academy of Sciences, Boltzmanngasse 3, Wien 1090, Austria*
[4] *HUN-REN Institute for Nuclear Research, P.O. Box 51, H-4001 Debrecen, Hungary*

We consider the problem of optimizing the state average of a polynomial of non-commuting variables, over all states and operators satisfying a number of polynomial constraints, and over all Hilbert spaces where such states and operators are defined. Such non-commutative polynomial optimization (NPO) problems are routinely solved through hierarchies of semidefinite programming (SDP) relaxations. By formulating the general NPO problem in Lagrangian terms, we heuristically derive first-order optimality conditions via small variations in the problem variables. Although the derivation is not rigorous, it gives rise to two types of optimality conditions—state and operator—which are rigorously analyzed in the paper. Both types of conditions can be enforced through additional positive semidefinite constraints in the SDP hierarchies. State optimality conditions are shown to be satisfied by all NPO problems. For NPO problems with optimal solutions (such as, e.g., Archimedean ones) they allow enforcing a new type of constraints: namely, restricting the optimization over states to the set of common ground states of an arbitrary number of operators. Operator optimality conditions are the non-commutative analogs of the Karush–Kuhn–Tucker (KKT) conditions, which are known to hold in many classical optimization problems. In this regard, we prove that a weak form of operator optimality holds for all NPO problems; stronger versions require the problem constraints to satisfy some qualification criterion, just like in the classical case (e.g.: Mangasarian–Fromovitz constraint qualification). We test the power of the new optimality conditions by computing local properties of ground states of many-body spin systems and the maximum quantum violation of Bell inequalities.

AMS classification: 14A22, 47A62, 65K05, 90C23, 90C22, 90C46.
Keywords: noncommutative polynomial optimization, optimality conditions, semidefinite programming, Bell nonlocality, statistical physics.
Communicated by JM Landsberg.

CONTENTS

1 Introduction 2
2 Background on polynomial optimization theory 4
2.1 Classical polynomial optimization 4
2.2 Non-commutative polynomial optimization 5
3 First-order optimality conditions 8
3.1 State optimality conditions 9
3.2 Operator optimality conditions 14
3.3 A numerical example 18

* mateus.araujo@uva.es
† igor.klep@fmf.uni-lj.si
‡ ajp.garner@gmail.com
§ tvertesi@atomki.hu
¶ Corresponding author. miguel.navascues@oeaw.ac.at

4 Non-commutative constraint qualification . . . 19
4.1 Essential KKT conditions . . . 20
4.1.1 Setting ϵ_i to zero . . . 25
4.2 Normed KKT conditions . . . 29
4.2.1 Linearly independent gradients . . . 29
4.2.2 Non-commutative Mangasarian–Fromovitz Constraint Qualification . . . 33
4.3 Strong KKT conditions . . . 35
4.3.1 Equality constraints with a faithful finite-dimensional $*$-representation . . . 37
4.3.2 Convexity . . . 37
5 Partial operator optimality conditions . . . 39
5.1 Partial normed ncKKT conditions . . . 39
5.2 Partial strong ncKKT conditions . . . 41
6 Applications . . . 44
6.1 Many-body quantum systems . . . 44
6.2 The curious case of quantum Bell inequalities . . . 49
6.2.1 Only two outcomes . . . 50
6.2.2 Numerical implementation . . . 51
7 Conclusion . . . 54
Acknowledgments . . . 54
References . . . 55
A Heuristic derivation of the first-order optimality conditions . . . 59
1 State optimality conditions . . . 60
2 Strong ncKKT . . . 60
B Re-derivation of the first-order optimality conditions of [1] through Theorem 4.1 . . . 61
C NPO for quantum nonlocality . . . 61

1. INTRODUCTION

Non-commutative polynomial optimization (NPO) studies the problem of minimizing the bottom of the spectrum of a polynomial of non-commuting variables, over all operator representations of these variables satisfying a number of polynomial equations and inequalities. As it turns out, in quantum mechanics many interesting physical quantities such as energy, spin and momentum are represented by operators satisfying polynomial constraints. Consequently, natural applications of NPO have been found in quantum information theory, quantum chemistry and condensed matter physics in the last decades. Examples of practical NPO problems include computing the maximal quantum violation of a Bell inequality [2, 3], the electronic energy of atoms and molecules [4–6], the ground state energies of spin systems [7–10], or the ground state behavior of fermions at finite density [11].

From the work of Pironio *et al.* [12], (see also Refs. [2, 3, 13–15]), we know that all NPO problems involving bounded operators can be solved through hierarchies of semidefinite programs (SDPs) [16, 17] of increasing complexity. While the first levels of said hierarchies provide very good approximations for many NPO problems, sometimes there are considerable gaps between the lower bound provided by the SDP solver and the conjectured solution of the problem. That is, even though the SDP hierarchies converge for any problem, for some NPO problems they seem to converge too slowly. This leaves many important problems in quantum nonlocality and many-body physics unsolved, due to a lack of computational resources.

In this paper, we introduce an improved and stronger method to tackle NPO problems. The main idea is that NPOs, like classical optimization problems, very often obey a number of optimality relations, which in the classical case are dubbed the Karush–Kuhn–Tucker (KKT) conditions [18, 19]. In the present work, we adapt and generalize these conditions to the non-commutative setting. We find that they come in two flavors: state and operator optimality

conditions, both of which take the form of positive semidefinite constraints on top of the original SDP hierarchies [12]. Optimality constraints on the solutions of specific NPO problems stemming from first-order differentiation have already been considered in the literature [1], although not in the context of deriving new or improving existing numerical methods. Motivation aside, our contribution differs from this earlier work in the generality and scope of our optimality conditions, which we believe exhaust the set of first-order optimality constraints and can be applied to a large variety of NPO problems.

Our new optimality conditions come with two benefits: on one hand, they boost the speed of convergence of the original SDP hierarchy, often yielding convergence at a finite level. On the other hand, they allow us to enforce new types of constraints on NPO problems, such as demanding that the states over which the optimization takes place are the ground states of certain operators. We exploit this feature in Sections 3.1 and 6.1, where we extract certified lower and upper bounds on local properties of the ground state of many-body spin systems. Remarkably, the ground state condition can be enforced in translation-invariant quantum systems featuring infinitely many particles. This allows us to make rigorous claims about the physics of quantum spin chains in the thermodynamic limit, thus solving an important open problem in condensed matter physics.

As in the classical, commutative case, careful study is required to justify exactly when the new non-commutative optimality conditions hold. While the state optimality conditions are easily seen to hold in all NPO problems, justifying the corresponding operator optimality conditions (the non-commutative Karush-Kuhn-Tucker conditions, or ncKKT conditions) requires more work. In this context, we show that: (a) essential ncKKT, the most relaxed variant of the operator optimality conditions, holds in all NPO problems; (b) normed ncKKT, a substantially strengthened version, holds in all "bounded" NPO problems that satisfy the non-commutative analog of the Mangasarian–Fromovitz conditions [20]; (c) the even more restrictive strong ncKKT conditions hold if either the NPO problem is convex or its solution is achieved at a finite level of the original SDP hierarchy.

Since the NPO formulation of quantum nonlocality only seems to satisfy essential ncKKT conditions, we also provide sufficient conditions that guarantee that either normed or strong ncKKT condition partially holds in a given NPO problem. Namely, in scenarios where the set of all variables can be partitioned into subsets that commute with each other and the remaining constraints and objective function for each of these parts are convex (satisfy the non-commutative Mangasarian–Fromovitz conditions), then a relaxed form of strong (normed) ncKKT holds. This result allows us to enforce more powerful optimality conditions on quantum nonlocality problems, with the resulting boost in convergence.

Guide to the paper. The structure of this paper is as follows: In Section 2.1 we recall optimization of commutative polynomials and then in Section 2.2 we move to the class of non-commutative optimization problems that we consider in this paper, and present their corresponding hierarchies of SDP relaxations. In Section 3, we propose several generalizations of the first-order conditions for the non-commutative framework, which will allow us to incorporate extra constraints into our optimization problems. The necessity of the state optimality conditions will be already proven in Section 3.1. Sufficient criteria for the validity of the different forms of operator optimality are presented in Section 4. Section 5 investigates when it is legitimate to enforce the new optimality conditions partially. In Section 6, we will conduct numerical tests to see how the optimality conditions perform in practical problems. In this regard, we present two applications: the computation of the local properties of many-body quantum systems at zero temperature (Section 6.1) and the maximum violation of bipartite Bell inequalities (Section 6.2). We then present our conclusions.

While conducting this research, we found that Fawzi *et al.* [21] had independently arrived at the state optimality conditions (3.2). In their interesting paper, the authors provide a sequence of convex optimization relaxations of the set of local averages of condensed matter systems at

finite temperature. When the temperature parameter is set to zero, their convex optimization hierarchy turns into an SDP hierarchy, which coincides with the one presented in Section 6.1 of this paper.

2. BACKGROUND ON POLYNOMIAL OPTIMIZATION THEORY

2.1. Classical polynomial optimization

In this section, we informally introduce the state-of-the-art on optimization of commutative polynomials through hierarchies of semidefinite programs. For a more detailed account, we refer the reader to [22–24].

Consider a classical polynomial optimization problem, i.e., a problem of the form:

$$\begin{aligned} p^\star := &\min f(x) \\ &\text{s.t. } g_i(x) \geq 0, \quad i = 1, \dots, m, \\ &\quad\; h_j(x) = 0, \quad j = 1, \dots, m', \end{aligned} \tag{2.1}$$

where $x = (x_1, \dots, x_n)$ is a vector of real variables, and f, g_i, h_j are real-valued polynomials thereof. Given a function $s(x)$, call $\partial_x s$ its gradient, i.e., $\partial_x s = \left(\frac{\partial s(x)}{\partial x_1}, \dots, \frac{\partial s(x)}{\partial x_n}\right)$. Let $\mathbb{A}(x)$ denote the set of *active* inequality constraints, i.e., the set of indices $i \in \{1, \dots, m\}$ for which $g_i(x) = 0$. In this commutative scenario, the Karush–Kuhn–Tucker (KKT) conditions read:

$$\begin{aligned} &\exists \{\mu_i\}_i \subset \mathbb{R}_{\geq 0},\ \{\lambda_j\}_j \subset \mathbb{R}, \\ &\text{such that } \partial_x f\Big|_{x=x^\star} = \sum_{i \in \mathbb{A}(x^\star)} \mu_i \partial_x g_i\Big|_{x=x^\star} + \sum_j \lambda_j \partial_x h_j\Big|_{x=x^\star}, \\ &\mu_i g_i(x^\star) = 0, \quad \forall i, \end{aligned} \tag{2.2}$$

where $x^\star$ is an optimizer of (2.1).

The optimal solutions of many classical optimization problems, even those involving non-polynomial functions, are known to satisfy the KKT conditions, also known as first-order optimality conditions [18, 19]. Sufficient criteria to ensure that the KKT conditions hold are dubbed 'constraint qualifications'. A well-known instance of the latter is Mangasarian–Fromovitz constraint qualification MFCQ [20], which stipulates that any problem of the form (2.1) satisfies the (commutative) KKT conditions if

$$\{\partial_x h_j(x^\star)\}_j \text{ linearly independent}, \tag{2.3}$$

and there exists $|z\rangle \in \mathbb{R}^n$ such that[1]

$$\begin{aligned} &\langle \partial_x g_i(x^\star) | z \rangle > 0, \quad \forall i \in \mathbb{A}(x^\star), \\ &\langle \partial_x h_j(x^\star) | z \rangle = 0, \quad j = 1, \dots, m'. \end{aligned} \tag{2.4}$$

Polynomial optimization problems can be tackled via the Lasserre-Parrilo moment-SOS hierarchy of semidefinite programming (SDP) relaxations [25, 26]. This is a sequence of SDPs, with solutions $(p^k)_k$, such that $p^1 \leq p^2 \leq \cdots \leq p^\star$. Moreover, if the constraints of Problem (2.1) imply that any feasible point x is bounded (more precisely, the constraints in (2.1) satisfy a condition known as Archimedeanity, see below), then it holds that $\lim_{k\to\infty} p^k = p^\star$.

In [27], Nie shows that generically, if we explicitly add the KKT constraints (2.2) to Problem (2.1) and then we apply the Lasserre-Parrilo construction, we end up with a hierarchy of SDP

[1] In the following, we use the physicist's "braket" notation, whereby a column vector ψ is denoted by the "ket" $|\psi\rangle$; and its conjugate transpose, by the "bra" $\langle\psi|$. In this notation, the scalar product between the vectors ϕ, ψ is represented by the "braket" $\langle\phi|\psi\rangle$, which is anti-linear in its first argument and linear in the second.

relaxations that collapses at a finite order. Namely, there exists $\bar{k} \in \mathbb{N}$ such that $p^{\bar{k}} = p^\star$. Harrow, Natarajan and Wu go one step beyond in [28], by upper bounding $\bar{k}$ for a large class of polynomial optimization problems with applications in quantum information theory.

The overarching goal of this paper is to generalize these methods to non-commutative polynomial optimization problems, which are the subject of the next subsection.

2.2. Non-commutative polynomial optimization

In this section, we summarize current notions and methods in non-commutative polynomial optimization (NPO). For a deeper treatment of these matters, we recommend the reviews [29], [15].

First, let us set the notation to be used throughout the paper. In this work, we will be interested in polynomials with complex coefficients of n non-commuting variables $x = (x_1, \ldots, x_n)$. We will denote by $\mathbb{C}\langle x\rangle$ the set of all such polynomials; the set of monomials or words of the alphabet x will be called $\langle x\rangle$. By $\mathbb{C}\langle x\rangle_d$ and $\langle x\rangle_d$ we will respectively mean the set of polynomials or words of degree smaller than or equal to d.

The set $\mathbb{C}\langle x\rangle$ comes with an involution $*$, which has the effect of conjugating the coefficients of the polynomial and inverting the order of the monomials. A polynomial $p(x)$ is called symmetric or Hermitian if $p(x) = p(x)^*$. Note that, according to this definition, the variables $x_1, \ldots, x_n$ themselves are Hermitian.

The variables $x_1, \ldots, x_n$ are meant to model Hermitian elements of a complex C^*-algebra $\mathcal{A}$, namely, a Banach algebra with an involution $*$ such that $\|aa^*\| = \|a\|^2$, for all $a \in \mathcal{A}$. Note that, for any Hilbert space $\mathcal{H}$, the set of bounded operators $B(\mathcal{H})$ with the operator norm is a C^*-algebra. In fact, for every C^*-algebra, there exists a Hilbert space $\mathcal{H}$ such that $\mathcal{A}$ is isomorphic to some closed $*$-subalgebra of $B(\mathcal{H})$. A *state* ψ in $\mathcal{A}$ is a linear functional $\psi : \mathcal{A} \to \mathbb{C}$ satisfying: $\psi(a^*) = \psi(a)^*$ for all $a \in \mathcal{A}$ (Hermiticity), $\psi(1) = 1$ (normalization) and $\psi(aa^*) \geq 0$ (positivity). Given an algebra $\mathcal{A}$ with an operator representation in $B(\mathcal{H})$, any normalized vector $|\psi\rangle \in \mathcal{H}$ defines a state ψ through $\psi(\bullet) = \langle\psi|\pi(\bullet)|\psi\rangle$, where π is the $*$-homomorphism that maps elements of $\mathcal{A}$ into $B(\mathcal{H})$. A state ψ is *normal* if there exist vectors $\{|\psi_k\rangle\}_k \subset \mathcal{H}$ such that $\psi(\bullet) = \sum_k \langle\psi_k|\pi(\bullet)|\psi_k\rangle$.

For a tuple of Hermitian operators $X \in B(\mathcal{H})^n$, $\mathcal{A}(X)$ will denote the unital $*$-subalgebra of $B(\mathcal{H})$ generated by X, i.e., linear combinations of finite products of the X_i. Its closure in the norm is a C^*-subalgebra of $B(\mathcal{H})$, denoted $C^*(X)$. Finally, its bicommutant (that is, the set of all elements of $B(\mathcal{H})$ that commute with all elements that commute with all X_i) is the von Neumann algebra generated by X, denoted $W^*(X)$. By von Neumann's bicommutant theorem [30, Theorem II.3.9], $\mathcal{A}(X)$ is dense in $W^*(X)$ in the weak or strong operator topology.

We are ready to define NPO problems.

Definition 2.1. *Let $x = (x_1, \ldots, x_n)$ be a tuple of non-commuting variables, and let f, $\{g_i : i = 1, \ldots, m\}$, $\{h_j : j = 1, \ldots, m'\}$ be symmetric polynomials on those variables. Then, the following program is a non-commutative polynomial optimization (NPO) problem:*

$$\begin{aligned} p^\star := \inf_{\mathcal{H}, X, \psi}\ & \psi(f(X)) \\ s.t.\ \ & g_i(X) \geq 0, \quad i = 1, \ldots, m, \\ & h_j(X) = 0, \quad j = 1, \ldots, m', \end{aligned} \tag{2.5}$$

where the minimization takes place over all Hilbert spaces $\mathcal{H}$, states $\psi : B(\mathcal{H}) \to \mathbb{C}$ and Hermitian operators $(X_1, \ldots, X_n) \in B(\mathcal{H})^n$.

A non-commutative polynomial optimization (NPO) problem [12, 15] is the natural analog of a polynomial optimization problem [22, 25, 31]. Computing the maximal quantum violation

of a Bell inequality [32–34] or the energy of a many-body quantum system [35] are examples of NPO problems.

Remark 2.2. One might think of extending the definition of NPO problem to also allow equality constraints of the form $h(X) = 0$, with h not necessarily a Hermitian polynomial. However, such problems are also encompassed in the definition above. Indeed, any constraint of the form $h(X) = 0$ is equivalent to the double constraint $\{h^1(X) = 0, h^2(X) = 0\}$, where

$$h^1 := h + h^*, h^2 := i(h - h^*) \tag{2.6}$$

are Hermitian polynomials.

Remark 2.3. Equality constraints of the form $h(x) = 0$, with $h \in \mathbb{C}\langle x\rangle$ Hermitian, are equivalent to the inequality constraints $\{h(x) \geq 0, -h(x) \geq 0\}$. However, the relaxations that we next present to tackle Problem (2.5) do make a distinction between both ways of modeling equality constraints. Ditto for the operator optimality conditions that we introduce in Sections 3.2 and 5. In both cases, separately handling equality constraints using ideals as we do here is more efficient than handling two inequality constraints.

The formulation (2.5) of the NPO problem is very difficult to work with: since the optimal Hilbert space $\mathcal{H}^\star$ (if it exists) might well be infinite dimensional, it requires conducting an optimization over infinitely many variables! Luckily, Problem (2.5) can be relaxed to an (as we next see, more tractable) optimization over linear functionals on $\mathbb{C}\langle x\rangle$, i.e., elements of $\mathbb{C}\langle x\rangle^*$. In this regard, the following two definitions will be useful.

Definition 2.4. *A linear functional $\sigma : \mathbb{C}\langle x\rangle \to \mathbb{C}$ is Hermitian if $\sigma(p) \in \mathbb{R}$ for all Hermitian $p \in \mathbb{C}\langle x\rangle$. A (Hermitian) linear functional $\sigma : \mathbb{C}\langle x\rangle \to \mathbb{C}$ is positive if $\sigma(pp^*) \geq 0$, for all $p \in \mathbb{C}\langle x\rangle$. These notions naturally extend to linear functionals $\mathcal{A} \to \mathbb{C}$ for any $*$-algebra $\mathcal{A}$.*

Given a feasible point $(\mathcal{H}, \psi, X)$ of Problem (2.5), the linear functional $\sigma : \mathbb{C}\langle x\rangle \to \mathbb{C}$, defined through

$$\sigma(p) := \psi(p(X)) \tag{2.7}$$

is obviously Hermitian and positive. Moreover, it satisfies $\sigma(1) = 1$.

To model the problem constraints, we will resort to the following two notions from algebraic geometry.

Definition 2.5. *Given a vector of Hermitian polynomials $\mathbf{h} = (h_1, \ldots, h_{m'})$, the* ideal generated by $\mathbf{h}$ *is the set of polynomials of the form:*

$$I(\mathbf{h}) = \left\{ \sum_{j,k} p_{j,k}^+ h_j p_{j,k}^- : \{p_{jk}^+\}_{jk} \cup \{p_{jk}^-\}_{jk} \subset \mathbb{C}\langle x\rangle \right\}. \tag{2.8}$$

We will denote by $I(\mathbf{h})_d$ the set of polynomials admitting a decomposition as above with $\deg(p_{j,k}^+) + \deg(h_j) + \deg(p_{j,k}^-) \leq d$, for all j, k.

Definition 2.6. *Given a vector of Hermitian polynomials $\mathbf{g} = (g_1, \ldots, g_m)$, the* quadratic module generated by $\mathbf{g}$ *is the set of polynomials of the form:*

$$M(\mathbf{g}) = \left\{ \sum_k p_k p_k^* + \sum_{i,k} p_{ik} g_i p_{ik}^* : \{p_j\}_j \cup \{p_{ik}\}_{ik} \subset \mathbb{C}\langle x\rangle \right\}. \tag{2.9}$$

We will denote by $M(\mathbf{g})_d$ the set of polynomials admitting a decomposition as above with $2\deg(p_k), 2\deg(p_{ik}) + \deg(g_i) \leq d$, for all i, k.

Note that both ideals and quadratic modules are convex cones of polynomials. Some such cones satisfy an important algebraic property that will appear over and over in this paper: Archimedeanity.

Definition 2.7. *Let C be a nonempty convex cone of polynomials.The cone C is Archimedean if, for any $p \in \mathbb{C}\langle x\rangle$, there exists $K \in \mathbb{R}^+$ such that $K - pp^* \in C$.*

Remark 2.8. For $\mathbf{g} \in \mathbb{C}\langle x\rangle^m$, $\mathbf{h} \in \mathbb{C}\langle x\rangle^{m'}$, the set of polynomials $M(\mathbf{g}) + I(\mathbf{h})$ defines a cone containing the identity. As proven in [36, Lemma 4] (or see [37, Section 2], [38, Section 2]), any such set is Archimedean iff there exists $K \in \mathbb{R}^+$ such that

$$K - \sum_i x_i^2 \in M(\mathbf{g}) + I(\mathbf{h}). \tag{2.10}$$

Hermitian elements s of $M(\mathbf{g}) + I(\mathbf{h})$ are also called *weighted sums of squares (SOS) polynomials*. The right-hand side of expressions of the form

$$s = \sum_k p_k p_k^* + \sum_{i,k} p_{ik} g_i p_{ik}^* + \sum_{j,k} p_{j,k}^+ h_j p_{j,k}^- \tag{2.11}$$

is called an *SOS decomposition for s*. Note that, since s is Hermitian, we can rewrite the above as

$$s = \sum_k p_k p_k^* + \sum_{i,k} p_{ik} g_i p_{ik}^* + \frac{1}{2}\sum_{j,k} p_{j,k}^+ h_j p_{j,k}^- + (p_{j,k}^-)^* h_j (p_{j,k}^+)^*. \tag{2.12}$$

Proposition 2.9. *If $s \in M(\mathbf{g}) + I(\mathbf{h})$, i.e., s admits an SOS decomposition as in (2.11), and the tuple of operators $\bar{X} \in B(\mathcal{H})^n$ satisfies the constraints of Problem (2.5), then $s(\bar{X})$ is a positive semidefinite operator.*

It is immediate that any linear functional σ, defined through Eq. (2.7) will satisfy the constraints $\sigma(s) \geq 0$, for all $s \in M(\mathbf{g}) + I(\mathbf{h})$, a condition which we will denote by $\sigma(M(\mathbf{g})) \geq 0$, and $\sigma(I(\mathbf{h})) = 0$. We arrive at the following relaxation of Problem (2.5):

$$\begin{aligned} q^\star := \inf_{\sigma:\mathbb{C}\langle x\rangle\to\mathbb{C}} & \sigma(f) \\ \text{s.t. } & \sigma(1) = 1, \\ & \sigma(M(\mathbf{g})) \geq 0, \\ & \sigma(I(\mathbf{h})) = 0, \end{aligned} \tag{2.13}$$

with $\mathbf{g} = (g_1, \ldots, g_m)$, $\mathbf{h} = (h_1, \ldots, h_{m'})$. Clearly, $q^\star \leq p^\star$. Moreover, if the set $M(\mathbf{g}) + I(\mathbf{h})$ is Archimedean, then $q^\star = p^\star$ and both problems admit an optimal solution [39]. This is partly a consequence of the Gelfand–Naimark–Segal (GNS) construction [40, 41]; see [30, Theorem I.9.14] for a modern treatment.

Given a positive linear functional σ satisfying the conditions of Problem (2.13), the GNS construction proceeds as follows: σ defines a sesquilinear form $(a, b) := \sigma(a^* b)$ on $\mathbb{C}\langle x\rangle$. By modding out nullvectors $\mathcal{N}$, this form induces a scalar product, whence upon completion we arrive at a Hilbert space $\mathcal{H}$. Left multiplication by the variables x_j on $\mathbb{C}\langle x\rangle$ induces a linear operator X_j on $\mathbb{C}\langle x\rangle/\mathcal{N}$, and then bounded operators X_j on $\mathcal{H}$ (at this step the Archimedean condition enters). These operators satisfy the constraints of Problem (2.5). Finally, the quotient image modulo $\mathcal{N}$ of the polynomial $1 \in \mathbb{C}\langle x\rangle$ is a unit vector $\phi \in \mathcal{H}$ such that

$$\langle\phi|p(X)|\phi\rangle = \sigma(p(x)), \quad \forall p \in \mathbb{C}\langle x\rangle. \tag{2.14}$$

Defining $\psi(\bullet) := \langle\phi| \bullet |\phi\rangle$, we thus have that $(\mathcal{H}, X, \psi)$ is a feasible point of Problem (2.5) with the same objective value as σ. This observation allows us to define a notion of boundedness for positive linear functionals in $\mathbb{C}\langle x\rangle$.

Definition 2.10. *Let $\sigma : \mathbb{C}\langle x\rangle \to \mathbb{C}$ be Hermitian and positive. We say that σ is* bounded *if its GNS construction generates bounded operators $X_1, \ldots, X_n$.*

Very conveniently, Problem (2.13) can be relaxed through hierarchies of semidefinite programs (SDP) [2, 3, 12]. Call $\mathbb{H}_k$ the set of Hermitian linear functionals defined in $\mathbb{C}\langle x\rangle_k$, and, for $k \in \mathbb{N}$ large enough, consider the finite-dimensional optimization problem

$$\begin{aligned} p^k := \inf_{\sigma^k \in \mathbb{H}_{2k}} \ & \sigma^k(f) \\ \text{s.t. } & \sigma^k(1) = 1, \\ & \sigma^k(M(\mathbf{g})_{2k}) \geq 0, \quad \sigma^k(I(\mathbf{h})_{2k}) = 0. \end{aligned} \tag{2.15}$$

The relaxation (2.15) can be cast as a semidefinite program [16, 42] with $\dim(\mathbb{C}\langle x\rangle_{2k})$ free complex variables. To implement the constraints, it suffices to consider bases of monomials. Let $\{o_a\}_a$ ($\{o_a^l\}_a$) be monomial bases of polynomials of degree at most k or smaller (degree at most $k - \left\lceil \frac{\deg(g_l)}{2} \right\rceil$). Then, the matrices

$$\begin{aligned} \left(M^k(\sigma^k)\right)_{ab} &:= \sigma^k(o_a^* o_b), \\ \left(M_l^k(\sigma^k)\right)_{ab} &:= \sigma^k((o_a^l)^* g_l o_b^l), \end{aligned} \tag{2.16}$$

are respectively called the *k^{th}-order moment matrix of σ^k* and the *k^{th}-order localizing matrix of σ^k for constraint g_l* [12]. Enforcing the second line of constraints in Problem (2.15) boils down to demanding that the moment matrix $M^k(\sigma^k)$ and the localizing matrices $\{M_l^k(\sigma^k)\}_l$ are positive semidefinite. The last line can be similarly dealt with: it suffices to make sure, for $j = 1, \ldots, m'$, that $\sigma^k(s h_j s') = 0$ for all monomials s, s' with $\deg(s) + \deg(s') + \deg(h_j) \leq 2k$.

The SDP dual of (2.15) can be formulated as:

$$q^k := \sup\{\theta : f - \theta \in M(\mathbf{g})_{2k} + I(\mathbf{h})_{2k}\}, \tag{2.17}$$

with $\mathbf{g} = (g_1, \ldots, g_m)$, $\mathbf{h} = (h_1, \ldots, h_{m'})$. This problem is also an SDP [13]. Problem (2.17) can be interpreted as finding the maximum real number θ such that the operator $f(X) - \theta$ can be proven (with an SOS certificate whose degree is bounded by $2k$) positive semidefinite for all tuples of operators X satisfying the constraints of Problem (2.5).

As the degree k of the available polynomials grows, one would expect the sequences of lower bounds $(q^k)_k$, $(p^k)_k$ to better approximate the solution $q^\star$ of Problem (2.13). Clearly, $p^1 \leq p^2 \leq \cdots \leq q^\star$, and similarly for the q's. In fact, a sufficient condition for these hierarchies to be complete, in the sense that $\lim_{n\to\infty} p^n = \lim_{n\to\infty} q^n = q^\star$, is that the cone of polynomials $M(\mathbf{g}) + I(\mathbf{h})$ satisfies Archimedeanity, in which case we will say that Problems (2.5) and (2.13) are Archimedean [12]. The Archimedean property further implies that Problems (2.5) and (2.13) are equivalent (and so $q^\star = p^\star$) and have optimal solutions (hence we can replace the inf's by min's in the problem definitions). Note that, if the Archimedean condition holds, then all the feasible points X are bounded. Conversely, if the set of feasible operators is bounded, then a relation of the form $K - \sum_i x_i^2 \geq 0$ can be added to the problem inequality constraints essentially without changing the problem.

3. FIRST-ORDER OPTIMALITY CONDITIONS

In this section, we present first-order optimality conditions for Problems (2.5) or (2.13). We arrived at them through a heuristic (namely, non-rigorous) argument, which we present in Appendix A. The argument involves mathematically ill-defined objects and it is not needed to understand the optimality conditions, or to find sufficient conditions to ensure that the conditions

apply. We decided to include it in the paper nonetheless, as it could be a useful starting point to generalize our results (e.g: to NPO problems with state constraints [43]).

This said, first-order optimality constraints come in two types: state and operator optimality conditions. Let us start with the former.

3.1. State optimality conditions

We present two definitions, one for each formulation of the NPO problem.

Definition 3.1. *We say that Problem (2.5) satisfies* state optimality conditions *if, for any solution $(\mathcal{H}^\star, \psi^\star, X^\star)$, it holds that*

$$\psi^\star([f(X^\star), s(X^\star)]) = 0, \quad \forall s \in \mathbb{C}\langle x\rangle, \tag{3.1a}$$

$$\psi^\star\left(s^*(X^\star)f(X^\star)s(X^\star) - \frac{1}{2}\{f(X^\star), s^*(X^\star)s(X^\star)\}\right) \geq 0, \quad \forall s \in \mathbb{C}\langle x\rangle, \tag{3.1b}$$

where the expression $\{a, b\}$ denotes the anti-commutator of a, b, i.e., $\{a, b\} := ab + ba$.

Definition 3.2. *We say that Problem (2.13) satisfies* state optimality conditions *if, for any bounded solution $\sigma^\star$, it holds that*

$$\sigma^\star([f, s]) = 0, \quad \forall s \in \mathbb{C}\langle x\rangle, \tag{3.2a}$$

$$\sigma^\star\left(s^* f s - \frac{1}{2}\{f, s^* s\}\right) \geq 0, \quad \forall s \in \mathbb{C}\langle x\rangle. \tag{3.2b}$$

Remark 3.3. Note that, according to the definitions above, if Problem (2.5) (Problem (2.13)) happens not to have optimal solutions (bounded optimal solutions), then it will satisfy the state optimality conditions. The same will hold true for the coming operator optimality conditions, introduced in Section 3.2.

As we will soon see, state optimality conditions are easily shown to hold for all NPO problems. Before we provide the proof, though, let us introduce a few useful definitions.

Definition 3.4. *Let $\mathcal{H}$ be a Hilbert space and let $H \in B(\mathcal{H})$ be a Hermitian operator.* The ground state energy of H*, denoted by $E_0(H)$, is the bottom of the spectrum of H. A* ground state of H *is a state $\sigma \in B(\mathcal{H})^*$ such that $\sigma(H) = E_0(H)$. We will denote by* $\mathrm{Gr}(H)$ *the set of ground states of H.*

The following proposition will play a role in the derivations to follow.

Proposition 3.5. *Let σ be a ground state of H. Then, for any operator Q it holds that*

$$\sigma(HQ) = E_0(H)\sigma(Q) = \sigma(QH). \tag{3.3}$$

Proof. Define $\bar{H} := H - E_0(H)$. Then we have that $\bar{H} \geq 0$ and that $\sigma(\bar{H}) = 0$. Since σ is a positive functional, it holds that the matrix

$$\begin{pmatrix} \sigma(Q^*\bar{H}Q) & \sigma(Q^*\bar{H}) \\ \sigma(\bar{H}Q) & \sigma(\bar{H}) \end{pmatrix} \tag{3.4}$$

is positive semidefinite: otherwise, there would exist a vector (a, b) such that $\sigma((aQ+b)^*\bar{H}(aQ+b)) < 0$. Since one of the diagonals in (3.4) is zero, it follows that the off-diagonal terms are zero too. Thus, $\sigma(\bar{H}Q) = \sigma(Q^*\bar{H}) = 0$, which implies $\sigma(HQ) = E_0(H)\sigma(Q)$, $\sigma(Q^*H) = E_0(H)\sigma(Q^*)$. The second equality in Eq. (3.3) follows by replacing Q^* by Q in the last relation. □

If σ is the ground state of some operator H, then σ can be shown to satisfy an analog of Eq. (3.1).

Proposition 3.6. *Let $\sigma \in B(\mathcal{H})^*$ be the ground state of the Hermitian operator $H \in B(\mathcal{H})$. Then, for any $Q \in B(\mathcal{H})$, it holds that*

$$\sigma([H,Q]) = 0, \quad \sigma\left(QHQ^* - \frac{1}{2}\{H, QQ^*\}\right) \geq 0. \tag{3.5}$$

Proof. Let $Q \in B(\mathcal{H})$. By Proposition 3.5, we have, on one hand, that

$$\sigma([H,Q]) = \sigma(HQ) - \sigma(QH) = E_0(H)(\sigma(Q) - \sigma(Q)) = 0. \tag{3.6}$$

On the other hand,

$$\begin{aligned} \sigma\left(QHQ^* - \frac{1}{2}\{H, QQ^*\}\right) &= \sigma\left(QHQ^* - E_0(H)QQ^*\right) && \text{(3.7a)} \\ &= \sigma\left(Q(H - E_0(H))Q^*\right) && \text{(3.7b)} \\ &\geq 0, && \text{(3.7c)} \end{aligned}$$

where the last inequality stems from the relation $H - E_0(H) \geq 0$. □

We are ready to prove the universal validity of the state optimality conditions.

Theorem 3.7. *Both Problem (2.5) and Problem (2.13) satisfy the state optimality conditions.*

Proof. We will first prove the statement for Problem (2.5). Let $(\mathcal{H}^\star, \Psi^\star, X^\star)$ be a solution of (2.5). Then, $p^\star \geq E_0(f(X^\star))$, since $f(X^\star) - E_0(f(X^\star)) \geq 0$. Further, $E_0(f(X^\star)) \geq p^\star$: otherwise, $\psi(f(X^\star)) < p^\star$, for any $\psi \in \mathrm{Gr}(f(X^\star))$, so $\Psi^\star$ would not be an optimal solution of (2.5).

Thus, $E_0(f(X^\star)) = p^\star$, and so $\Psi^\star$ is a ground state of the operator $f(X^\star)$. It follows, by Proposition 3.6, that $\Psi^\star$ satisfies Eqs. (3.5) for $H = f(X^\star)$ and any operator of the form $Q = s(X^\star)$, with $s \in \mathbb{C}\langle x\rangle$.

We next prove that Problem (2.13) also satisfies the state optimality conditions. Let $\sigma^\star$ be a bounded solution. Then we can use the GNS construction (see the paragraph leading up to (2.14)) to construct a representation $(\mathcal{H}^\star, \Psi^\star, X^\star)$ with $\sigma^\star(s) = \Psi^\star(s(X^\star))$ for all $s \in \mathbb{C}\langle x\rangle$. From here on, the proof goes exactly as for Problem (2.5). □

Remark 3.8. Note that, if Problem (2.13) is Archimedean, then: (a) it has optimal solutions; and (b) all of them are bounded. Thus, one can add the extra SDP constraints

$$\begin{aligned} \sigma^k([f,s]) = 0, \quad \forall s \in \mathbb{C}\langle x\rangle,\, \deg(s) + \deg(f) \leq 2k, && \text{(3.8a)} \\ \sigma^k\left(p^* f p - \frac{1}{2}\{f, p^* p\}\right) \geq 0, \quad \forall p \in \mathbb{C}\langle x\rangle,\, 2\deg(p) + \deg(f) \leq 2k. && \text{(3.8b)} \end{aligned}$$

to the SDP hierarchy (2.15). This has sometimes the effect of boosting its speed of convergence, as we show in Sections 3.3, 6.1.

By Proposition 3.6, conditions analog to (3.2) allow us to enforce a new type of constraints in non-commutative optimization problems: namely, restricting the state optimization to the set of ground states of a number of operators defined through Hermitian polynomials of the problem

variables. Indeed, given any set of Hermitian polynomials $b_1(x), \ldots, b_{m''}(x)$, consider a problem of the form

$$\begin{aligned} p_G^\star := \min_{\mathcal{H}, X, \psi} \ & \psi(f(X)) \\ \text{s.t. } & g_i(X) \geq 0, \qquad i = 1, \ldots, m, \\ & h_j(X) = 0, \qquad j = 1, \ldots, m', \\ & \psi \in \mathrm{Gr}(b_l(X)), \quad l = 1, \ldots, m''. \end{aligned} \tag{3.9}$$

By Proposition 3.6, one can relax this problem by adding to the relaxation (2.15) the SDP constraints:

$$\sigma^k([b_l, s]) = 0, \quad \forall s \in \mathbb{C}\langle x \rangle,\, \deg(s) + \deg(b_l) \leq 2k, \text{ for } l = 1, \ldots, m'', \tag{3.10a}$$

$$\sigma^k\left(p^* b_l p - \frac{1}{2}\{b_l, p^* p\}\right) \geq 0, \quad \forall p \in \mathbb{C}\langle x \rangle,\, 2\deg(p) + \deg(b_l) \leq 2k, \text{ for } l = 1, \ldots, m''. \tag{3.10b}$$

Call p_G^k the result of running program (2.15) together with the SDP constraints (3.10). Then, we trivially have that $p_G^1 \leq p_G^2 \leq \cdots \leq p_G^\star$. The following theorem provides a sufficient condition for the completeness of this hierarchy of relaxations.

Theorem 3.9. *Let $M(\mathbf{g}) + I(\mathbf{h})$ be Archimedean, and suppose that all tuples of Hermitian operators $X \in B(\mathcal{H})^n$ satisfying the problem constraints $\{g_i(X) \geq 0\}_i$ and $\{h_j(X) = 0\}_j$ have these two properties:*

(1) *For some unitary $U \in B(\mathcal{H})$,*

$$U\mathcal{A}(X)U^* = \bigoplus_l \mathcal{A}_l \otimes \mathbb{I}_{\bar{H}_l}, \tag{3.11}$$

where the direct sum is taken to be countable and, for each l, $\mathcal{A}_l$ is an irreducible $*$-algebra.

(2) *There exist projectors $\{P_l \in W^*(X)\}_l$ such that, for all l,*

$$UP_l\mathcal{A}(X)P_lU^* = 0 \oplus (\mathcal{A}_l \otimes \mathbb{I}_{\bar{H}_l}) \oplus 0. \tag{3.12}$$

Then, $\lim_{k\to\infty} p_G^k = p_G^\star$.

Remark 3.10. The conditions of the theorem are satisfied if either of the conditions below hold:

(1) For any X such that $\{g_i(X) \geq 0\}_i$ and $\{h_j(X) = 0\}_j$, $W^*(X)$ is *purely atomic*, i.e., it decomposes as a (at most countable) direct sum of irreducible von Neumann algebras. This could be the case, for instance, if the problem's equality constraints are such that, for some Hermitian polynomials $\{p_k(x)\}_k$, it holds that

$$\begin{aligned} &[x_i, p_k], p_k p_j - \delta_{jk} p_k \in I(\mathbf{h}),\ \forall j, k, \\ &1 - \sum_k p_k \in I(\mathbf{h}), \end{aligned}$$

and, in addition, for each k, the ideal I_k generated by $\{p_k x_l\}_l$ is such that $I_k / \big(I(\mathbf{h}) \cap I_k\big)$ admits a single irreducible representation π with $\pi(p_k) = 1$. For any feasible operator tuple $X \in B(\mathcal{H})^n$, the operators $\{p_k(X)\}_k$ would thus be central projections in $W^*(X)$, yielding a block decomposition

$$\mathcal{H} = \bigoplus_{k=1}^r p_k(X)\mathcal{H}, \qquad W^*(X) = \bigoplus_{k=1}^r W^*\Big(X\big|_{p_k(X)\mathcal{H}}\Big),$$

with each summand $W^*(X|_{p_k(X)\mathcal{H}})$ irreducible.

(2) For any X such that $\{g_i(X) \geq 0\}_i$ and $\{h_j(X) = 0\}_j$, $\mathcal{A}(X)$ is finite dimensional. This is often the most straightforward condition to verify in applications, and is of importance since it covers, e.g., the Pauli constraints, see Section 6.1. If the equality constraints $h_j(x) = 0$ imply that the quotient $*$-algebra $\mathbb{C}\langle x\rangle/I(\mathbf{h})$ is finite-dimensional (equivalently: for every feasible tuple X, the $*$-algebra $\mathcal{A}(X)$ generated by X is finite-dimensional), then $\mathcal{A}(X)$ is isomorphic to a finite direct sum of full matrix algebras,

$$\mathcal{A}(X) \cong \bigoplus_{\ell=1}^{r} M_{d_\ell}(\mathbb{C}),$$

and the decompositions in Theorem 3.9 are obtained by taking P_ℓ to be the (central) block projections corresponding to these simple summands.

To prove Theorem 3.9, we need the following lemma.

Lemma 3.11. *Let $\mathcal{H}$ be a Hilbert space, let ψ be a normal state and let $X \in B(\mathcal{H})^n$ be Hermitian operators. Let X have no non-trivial invariant subspaces, i.e., for any non-zero $|\phi\rangle \in \mathcal{H}$, it holds that $\mathcal{A}(X)|\phi\rangle$ is dense in $\mathcal{H}$. Then, for any Hermitian $b \in W^*(X)$, the conditions*

$$\psi\left([b, s]\right) = 0, \quad \forall s \in \mathcal{A}(X) \tag{3.13a}$$

$$\psi\left(sbs^* - \frac{1}{2}\{ss^*, b\}\right) \geq 0, \quad \forall s \in \mathcal{A}(X) \tag{3.13b}$$

imply that ψ is a ground state of b.

Proof. Note that the first condition of (3.13) implies that $\psi([b^n, s]) = 0$ for all $n \in \mathbb{N}$, $s \in \mathcal{A}(X)$. Indeed, consider the following chain of equalities:

$$\psi(b^n s) = \psi(b(b^{n-1}s)) = \psi((b^{n-1}s)b) = \psi(b(b^{n-2}sb)) = \psi(b^{n-2}sb^2) = \cdots = \psi(sb^n).$$

Since $\mathcal{A}(X)$ is dense in $W^*(X)$ in the weak topology, the relation (3.13a) also holds for all $s \in W^*(X)$. Invoking again the density of $\mathcal{A}(X)$ and the equation above, it follows that, for any measurable function $r : \mathbb{R} \to \mathbb{R}$,

$$\psi([r(b), s]) = 0, \quad \forall s \in W^*(X). \tag{3.14}$$

For any measurable subset A of $[E_0(b), -E_0(-b)]$, define $\Pi(A) := \chi_A(b)$, i.e., $\Pi(A)$ is the projector onto the generalized eigenstates of b with eigenvalues within A. Note that, for any set of disjoint intervals $\{A_j\} \subset [E_0(b), -E_0(-b)]$ with $\cup_j A_j = [E_0(b), -E_0(-b)]$ it holds that

$$\psi(\bullet) = \psi(\sum_j \Pi(A_j)^2 \bullet) = \psi(\sum_j \Pi(A_j) \bullet \Pi(A_j)), \tag{3.15}$$

where the first equality follows from $\sum_j \Pi(A_j) = 1$ and $\Pi(A_j)^2 = \Pi(A_j)$, and the second one is a consequence of Eq. (3.14), with $r(b) = \Pi(A_j)$, $s = \Pi(A_j)\bullet$.

For $\epsilon > 0$ sufficiently small consider the projectors $\Pi_G := \Pi([E_0(b), E_0(b) + \epsilon])$, $\Pi_E := \Pi([E_0(b) + 2\epsilon, -E_0(-b)])$. Let $N \in \mathbb{N}$ and, for $j = 0, \ldots, N-1$, define $\Pi_j := \Pi([E_j, E_j + \delta)$, with $E_j := E_0(h) + 2\epsilon + j\delta$ and $\delta := \frac{-E_0(-b) - E_0(b) - 2\epsilon}{N}$. Note that $\{\Pi_j\}_j$ are orthogonal projectors and they add up to Π_E.

Setting $s = \Pi_E \tilde{s} \Pi_G$, in Eq. (3.13), we have that

$$\begin{aligned}
&\psi\left(sbs^* - \frac{1}{2}\{ss^*, b\}\right) \\
&= \psi\left(\Pi_E \tilde{s} \Pi_G b \Pi_G \tilde{s}^* \Pi_E\right) - \psi\left(\Pi_E \tilde{s} \Pi_G \tilde{s}^* \Pi_E b\right)
\end{aligned}$$

$$
\begin{aligned}
&\leq (E_0(b)+\epsilon)\psi\left(\Pi_E\tilde{s}\Pi_G\tilde{s}^*\Pi_E\right) - \sum_j \psi\left(\Pi_j\Pi_E\tilde{s}\Pi_G\tilde{s}^*\Pi_E b\Pi_j\right)\\
&= (E_0(b)+\epsilon)\psi\left(\Pi_E\tilde{s}\Pi_G\tilde{s}^*\Pi_E\right) - \sum_j E_j\psi\left(\Pi_j\tilde{s}\Pi_G\tilde{s}^*\Pi_j\right) - \sum_j \psi\left(\Pi_j\tilde{s}\Pi_G\tilde{s}^*\Pi_E(b\Pi_j - E_j\Pi_j)\right)\\
&\leq \sum_j (E_0(b)+\epsilon-E_j)\psi\left(\Pi_j\tilde{s}\Pi_G\tilde{s}^*\Pi_j\right) + \sum_j \psi\left(\Pi_j\right)\|\tilde{s}\|^2\delta\\
&\leq -\epsilon\sum_j \psi\left(\Pi_j\tilde{s}\Pi_G\tilde{s}^*\Pi_j\right) + \|\tilde{s}\|^2\delta\\
&= -\epsilon\psi\left(\Pi_E\tilde{s}\Pi_G\tilde{s}^*\Pi_E\right) + \|\tilde{s}\|^2\delta. \qquad (3.16)
\end{aligned}
$$

We can make δ as small as we wish while keeping ϵ the same. Since the first expression of Eq. (3.16) above is non-negative, it follows that

$$\psi\left(\Pi_E\tilde{s}\Pi_G\tilde{s}^*\Pi_E\right) = 0. \qquad (3.17)$$

Now, ψ is a normal state, and thus $\psi(\bullet) = \sum_k \langle\psi_k| \bullet |\psi_k\rangle$, for some non-zero vectors $|\psi_k\rangle$. Eq. (3.17) implies that

$$\Pi_G\tilde{s}^*\Pi_E|\psi_k\rangle = 0, \qquad (3.18)$$

for all $\tilde{s} \in \mathcal{A}(X)$. Since by assumption there are no invariant subspaces, this implies that $\Pi_E|\psi_k\rangle = 0$ for all k. Hence, $\psi(\Pi_E) = 0$, which implies that

$$\psi(b) = \psi(\Pi_E b\Pi_E) + \psi((1-\Pi_E)b(1-\Pi_E)) = \psi((1-\Pi_E)b(1-\Pi_E)) \leq E_0(b) + 2\epsilon. \qquad (3.19)$$

Now let $\epsilon \searrow 0$ to deduce $\psi(b) \leq E_0(b)$. On the other hand, $\psi(b) \geq E_0(b)$, since ψ is a state. It follows that $\psi(b) = E_0(b)$, and so ψ is a ground state of b. □

Proof of Theorem 3.9. If Problem (3.9) is feasible and Archimedean, then one can follow the proof of Theorem 1 in [12] to conclude that there exist $(\mathcal{H}^\star, \psi^\star, X^\star)$ such that:

1. $\psi^\star(\bullet) = \langle\phi^\star| \bullet |\phi^\star\rangle$, for some unit vector $|\phi^\star\rangle \in \mathcal{H}^\star$.
2. $X^\star$ satisfies $g_i(X^\star) \geq 0$, for $i = 1, \ldots, m$ and $h_j(X^\star) = 0$, for $j = 1, \ldots, m'$.
3. $\psi^\star(f(X^\star)) = \lim_{k\to\infty} p_G^k$.
4. For $l = 1, \ldots, m''$, it holds that

$$\psi^\star\left([b_l(X^\star), s]\right) = 0, \quad \forall s \in \mathcal{A}(X^\star) \qquad (3.20a)$$

$$\psi^\star\left(sb_l(X^\star)s^* - \frac{1}{2}\{ss^*, b_l(X^\star)\}\right) \geq 0, \quad \forall s \in \mathcal{A}(X^\star). \qquad (3.20b)$$

Since the operators $X^\star$ satisfy the problem constraints, by assumption we have that there exists a unitary $U \in B(\mathcal{H})$ such that

$$UX_i^\star U^* = \bigoplus_j X_{i,j}^\star \otimes \mathbb{I}_{\bar{\mathcal{H}}_j}, \qquad (3.21)$$

where $\{X_{i,j}^\star : i\} \subset B(\mathcal{H}_j)$ generate an irreducible $*$-algebra in $B(\mathcal{H}_j)$.

By assumption, for any j, there exists a projector $P_j \in W^*(X^\star)$ onto the j^{th} block of (3.21). Then, $\pi_j(\bullet) = P_j \bullet P_j$ is a homomorphism on $\mathcal{A}(X^*)$, and thus $g_i(\pi_j(X^\star)) \geq 0$, for all i and $h_l(\pi_j(X^\star)) = 0$, for all l.

Let $W_l(X^\star)$ be the algebra generated by $\{X^\star_{i,l}\}_i$, and let $\psi^\star_j : W_j(X^\star) \to \mathbb{C}$ be defined as

$$\psi^\star_j(\bullet) := \langle \phi^\star_j | (0 \oplus \bullet \otimes \mathbb{I}_{\bar{H}_j} \oplus 0) | \phi^\star_j \rangle, \tag{3.22}$$

with $|\phi^\star_j\rangle = U^* P_j |\phi^\star\rangle$. Then, for $\psi^\star_j \neq 0$, $\frac{1}{\psi^\star_j(1)}\psi^\star_j$ is a normal state. In addition,

$$\psi^\star(\bullet) = \sum_j \psi^\star(P_j \bullet P_j) = \sum_j \psi^\star_j(\pi_j(\bullet)). \tag{3.23}$$

Now, for all j such that $\psi^\star_j \neq 0$, Eq. (3.20) is satisfied if we replace $\psi^\star$ by $\psi^\star_j$ and all operators $\bullet$ by $\pi_j(\bullet)$. Indeed, it suffices to take $s = P_j s P_j$ in Eq. (3.20). By Lemma 3.11, it thus follows that $\psi^\star_j$ is a ground state of $\pi_j(b_l(X^\star))$ for $l = 1, \ldots, m''$.

Hence, $\psi^\star(f(X^\star))$ can be expressed as a convex combination of the objective values of the feasible points $(\mathcal{H}_j, \frac{1}{\psi^\star_j(1)}\psi^\star_j, \pi_j(X^\star))$ of Problem (3.9) (for j such that $\psi^\star_j \neq 0$). Thus, $\lim_{k\to\infty} p^k_G \geq p^\star_G$. Since $p^k_G \leq p^\star_G$ for all k, we arrive at the statement of the theorem. □

As we will see in Section 6.1, the ability to conduct optimizations over ground states has important applications in many-body quantum physics.

3.2. Operator optimality conditions

Operator optimality conditions are the non-commutative analogs of the Karush–Kuhn–Tucker (KKT) conditions [20]. Such non-commutative Karush–Kuhn–Tucker (ncKKT) conditions come in several variants of different strength. To formulate them, we must first introduce an analog of the gradient for non-commutative polynomials. Closely related notions have been studied in free function theory [44] and non-commutative real algebraic geometry [45].

Definition 3.12. *The gradient of a (non-commutative) polynomial $p(x)$ is a polynomial of the original non-commuting variables x and their 'variations' $y = (y_1, \ldots, y_n)$, linear in y. This polynomial, denoted as $\nabla_x p(y)$, is obtained from $p(x)$ by evaluating $p(x + \epsilon y)$ and keeping only the terms linear in ϵ. Informally:*

$$\nabla_x p(y) := \lim_{\epsilon \to 0} \frac{p(x + \epsilon y) - p(x)}{\epsilon}. \tag{3.24}$$

The evaluation of the x variables in the expression $\nabla_x p(y)$, i.e., any replacement of the form $x \to z$ will be denoted as $\nabla_x p(y)\Big|_{x=z}$.

Note that this definition of a gradient is the non-commutative analog of directional derivatives.

Example 3.13. Let $p(x) = x_1^3$. Then, $\nabla_x p(y) = y_1 x_1^2 + x_1 y_1 x_1 + x_1^2 y_1$.

Armed with this definition, we can state the strongest of the ncKKT conditions:

Definition 3.14 (Strong ncKKT)**.** *We say that the NPO Problem (2.5) satisfies the* strong non-commutative Karush–Kuhn–Tucker conditions *if, for any solution $(\mathcal{H}^\star, X^\star, \psi^\star)$ of Problem (2.5), there exist positive linear functionals $\{\mu_i : \mathcal{A}(X^\star) \to \mathbb{C}\}_{i=1}^m$ and Hermitian linear functionals $\{\lambda_j : \mathcal{A}(X^\star) \to \mathbb{C}\}_{j=1}^{m'}$ such that*

$$\mu_i(g_i(X^\star)) = 0, \quad i = 1, \ldots, m, \tag{3.25a}$$

$$\psi^\star\left(\nabla_x f(p)\Big|_{x=X^\star}\right) - \sum_i \mu_i\left(\nabla_x g_i(p)\Big|_{x=X^\star}\right) - \sum_j \lambda_j\left(\nabla_x h_j(p)\Big|_{x=X^\star}\right) = 0, \quad \forall p \in \mathcal{A}(X^\star)^n. \tag{3.25b}$$

Note that, if Problem (2.5) does not have optimal solutions, then it satisfies strong ncKKT, see Remark 3.3.

The problem with the above definition is that we do not know how to enforce conditions (3.25a) fully. Indeed, current NPO techniques do not allow optimizing over both $\star$-algebras $\mathcal{A}(X^\star)$ and several (more than one) independent linear functionals defined on $\mathcal{A}(X^\star)$, unless $\mathcal{A}(X^\star)$ is uniquely defined, modulo isometries, by the problem constraints. Replacing $\mathcal{A}(X^\star)$ by $\mathbb{C}\langle x\rangle$ in the definition above, we arrive at a much more convenient set of constraints, phrased in terms of the solutions of Problem (2.13), rather than Problem (2.5).

Definition 3.15 (Normed ncKKT)**.** *We say that the NPO Problem* (2.13) *satisfies the* normed non-commutative Karush–Kuhn–Tucker conditions *if, for any bounded solution $\sigma^\star$ of Problem* (2.13)*, it holds that*

$$\forall i,j,\ \exists \text{ Hermitian } \mu_i, \lambda_j^{\pm} : \mathbb{C}\langle x\rangle \to \mathbb{C}, \quad \mu_i(M(\mathbf{g}) + I(\mathbf{h})),\ \lambda_j^{\pm}(M(\mathbf{g}) + I(\mathbf{h})) \geq 0, \tag{3.26a}$$

$$\text{s.t. } \mu_i(g_i) = 0, \quad i = 1,\dots,m, \tag{3.26b}$$

$$\sigma^\star\left(\nabla_x f(p)\right) - \sum_i \mu_i\left(\nabla_x g_i(p)\right) - \sum_j \lambda_j\left(\nabla_x h_j(p)\right) = 0, \quad \forall p \in \mathbb{C}\langle x\rangle^n, \tag{3.26c}$$

$$\text{with } \lambda_j := \lambda_j^+ - \lambda_j^-, \quad j = 1,\dots,m'. \tag{3.26d}$$

Note that we demand each functional λ_j to be the difference of two positive functionals $\lambda_j^{\pm}$. This is a relaxation of the property that the functional $\lambda_j : \mathcal{A}(X^\star) \to \mathbb{C}$ appearing in the definition of strong ncKKT is bounded (or normed). Indeed, the Jordan decomposition [30, Proposition III.2.1] shows that any Hermitian linear functional $\lambda : \mathcal{A}(X^\star) \to \mathbb{C}$ of bounded norm can be expressed as the difference between two positive linear functionals.

If we are willing to drop the boundedness assumption on $\{\lambda_j\}_j$, then we can simply define the latter to be Hermitian functionals of $\mathbb{C}\langle x\rangle$. Without the variables $\lambda_j^{\pm}$, though, it is no longer possible to enforce that λ_j is compatible with the inequality constraints $\{g_i(x) \geq 0\}_i$ in any meaningful way. We arrive at a weaker variant of the ncKKT conditions.

Definition 3.16 (Weak ncKKT)**.** *We say that the NPO Problem* (2.13) *satisfies the* weak non-commutative Karush–Kuhn–Tucker conditions *if, for any bounded solution $\sigma^\star$ of Problem* (2.13)*, it holds that*

$$\forall i,j,\ \exists \text{ Hermitian } \mu_i, \lambda_j : \mathbb{C}\langle x\rangle \to \mathbb{C}, \quad \mu_i(M(\mathbf{g}) + I(\mathbf{h})) \geq 0,\ \lambda_j(I(\mathbf{h})) = 0, \tag{3.27a}$$

$$\text{s.t. } \mu_i(g_i) = 0, \quad i = 1,\dots,m, \tag{3.27b}$$

$$\sigma^\star\left(\nabla_x f(p)\right) - \sum_i \mu_i\left(\nabla_x g_i(p)\right) - \sum_j \lambda_j\left(\nabla_x h_j(p)\right) = 0, \quad \forall p \in \mathbb{C}\langle x\rangle^n. \tag{3.27c}$$

If we were guaranteed that Problem (2.13) had bounded solutions and that the weak ncKKT conditions held, then we could add some further constraints to our SDP relaxation (2.15), namely:

$$\begin{aligned} &\text{For each } i = 1,\dots,m: \\ &\exists \text{ Hermitian } \mu_i^k : \mathbb{C}\langle x\rangle_{2k} \to \mathbb{C}, \quad \mu_i^k(M(\mathbf{g})_{2k} + I(\mathbf{h})_{2k}) \geq 0, \qquad \mu_i^k(g_i) = 0, \end{aligned} \tag{3.28a}$$

$$\begin{aligned} &\text{For each } j = 1,\dots,m': \\ &\exists \text{ Hermitian } \lambda_j^k : \mathbb{C}\langle x\rangle_{2k} \to \mathbb{C}, \quad \lambda_j^k(I(\mathbf{h})_{2k}) = 0, \end{aligned} \tag{3.28b}$$

and

$$\sigma^k(\nabla_x f(p(x))) - \sum_i \mu_i^k(\nabla_x g_i(p(x))) + \sum_j \lambda_j^k(\nabla_x h_j(p(x))) = 0,$$

$$\forall p \in \mathbb{C}\langle x\rangle^n,\ \deg\left(\nabla_x f\right),\ \deg\left(\nabla_x g_i\right),\ \deg\left(\nabla_x h_j\right) \leq 2k - \deg(p), \quad (3.28c)$$

where, for any polynomial s with $\nabla_x s(p) = \sum_{i,k} s_{ik}^+ p_k s_{ik}^-$, the expression $\deg(\nabla_x s)$ denotes $\max_{i,k} \deg(s_{ik}^+) + \deg(s_{ik}^-)$, and, for any n-tuple of polynomials p, $\deg(p) = \max_k \deg(p_k)$. Note that, in order to enforce the last condition (3.28c), it is enough to consider tuples of monomials p with just one non-zero entry.

Similarly, to enforce normed ncKKT, one just needs to add to Eq. (3.28), for each $j = 1, \dots m'$, the positive semidefinite constraints:

$$\exists \text{ Hermitian } \lambda_j^{\pm,k} : \mathbb{C}\langle x\rangle_{2k} \to \mathbb{C}, \quad \lambda_j^{\pm,k}(M(\mathbf{g})_{2k} + I(\mathbf{h})_{2k}) \geq 0, \quad (3.29a)$$

$$\lambda_j^k = \lambda_j^{+,k} - \lambda_j^{-,k}. \quad (3.29b)$$

As we will see, there are Archimedean problems that do not satisfy weak ncKKT conditions. This led us to introduce an even weaker variant of the ncKKT conditions, the so-called essential ncKKT conditions. In Theorem 4.1 below we show that these conditions are satisfied by all NPO problems.

Definition 3.17 (Essential ncKKT)**.** *An NPO Problem* (2.13) *satisfies the* essential ncKKT conditions *if, for any bounded minimizer $\sigma^\star$ of Problem* (2.13) *and for all $k \in \mathbb{N}$, $\{\epsilon_i\}_i \subset \mathbb{R}^+$ there exist Hermitian linear functionals $\mu_i^k : \mathbb{C}\langle x\rangle_{2k} \to \mathbb{C}$, $i = 1, \dots, m$, $\lambda_j^k : \mathbb{C}\langle x\rangle_{2k} \to \mathbb{C}$, $j = 1, \dots, m'$, such that*

For each $i = 1, \dots, m$:

$$\mu_i^k(pp^*) + \epsilon_i \|p\|_2^2 \geq 0, \qquad \forall p \in \mathbb{C}\langle x\rangle,\ \deg(p) \leq k, \quad (3.30a)$$

$$\mu_i^k(pg_l p^*) + \epsilon_i \|p\|_2^2 \geq 0, \qquad \forall p \in \mathbb{C}\langle x\rangle,\ \deg(p) \leq k - \left\lceil \frac{\deg(g_l)}{2} \right\rceil,\ l = 1, \dots, m, \quad (3.30b)$$

$$\mu_i^k(I(\mathbf{h})_{2k}) = 0, \quad (3.30c)$$

$$\mu_i^k(sg_i) = \mu_i^k(g_i s) = 0, \qquad \forall s \in \mathbb{C}\langle x\rangle,\ \deg(s) \leq 2k - \deg(g_i), \quad (3.30d)$$

For each $j = 1, \dots, m'$:

$$\lambda_j^k(I(\mathbf{h})_{2k}) = 0, \quad (3.30e)$$

and

$$\sigma^\star\left(\nabla_x f(p)\right) - \sum_i \mu_i^k\left(\nabla_x g_i(p)\right) - \sum_j \lambda_j^k\left(\nabla_x h_j(p)\right) = 0,$$
$$\forall p \in \mathbb{C}\langle x\rangle^n, \deg\left(\nabla_x f\right), \deg\left(\nabla_x g_i\right), \deg\left(\nabla_x h_j\right) \leq 2k - \deg(p), \quad (3.30f)$$

where $\|p\|_2$ denotes the 2-norm of the vector of coefficients of p.

Remark 3.18. Essential ncKKT is a relaxation of weak ncKKT, whereby we relax each positive semidefiniteness condition in the SDP (3.28), for each k. Indeed, enforcing the first two lines of Eq. (3.30) amounts to demanding that the k^{th}-order moment and localizing matrices (see Eq. (2.16) for a definition) $M^k(\mu_i), M_l^k(\mu_i)$, rather than being positive semidefinite, satisfy

$$M^k(\mu_i) + \epsilon_i \mathbb{I} \geq 0,\ M_l^k(\mu_i) + \epsilon_i \mathbb{I} \geq 0, \quad l = 1, \dots, m. \quad (3.31)$$

The essential ncKKT conditions thus translate to non-trivial positive semidefinite constraints, which one can use to boost the speed of convergence of the SDP relaxation (2.15).

Remark 3.19. Rather than demanding $\mu_i(g_i) = 0$ (as we did in Eq. (3.28a)), in Eq. (3.30d) we demand the stronger condition $\mu_i(sg_i) = \mu_i(g_i s) = 0$ for all $s \in \mathbb{C}\langle x\rangle$. The explanation is as follows: if ϵ_i were to vanish, then this condition would follow from $\mu_i(g_i) = 0$ and the positivity of μ_i and g_i. For $\epsilon_i \in \mathbb{R}^+$, this is no longer the case, so we impose the condition by hand.

Proposition 3.20. *Call* ***Essential***, ***Weak***, ***Normed***, ***Strong*** *the set of NPO problems respectively satisfying the essential, weak, normed and strong ncKKT conditions. Then we have that*

$$\textbf{\textit{Essential}} \supsetneq \textbf{\textit{Weak}} \supsetneq \textbf{\textit{Normed}} \supseteq \textbf{\textit{Strong}}. \tag{3.32}$$

The above relations also hold if we restrict to Archimedean NPO problems.

Proof. **Essential** is a relaxation of **Weak**, which is a relaxation of **Normed**, which is a relaxation of **Strong**. Thus,

$$\textbf{Essential} \supseteq \textbf{Weak} \supseteq \textbf{Normed} \supseteq \textbf{Strong}. \tag{3.33}$$

To prove the proposition, we need to show that there exists an Archimedean problem that satisfies essential ncKKT, but not weak ncKKT, and another one that satisfies weak ncKKT, but not normed ncKKT.

Consider thus the following instance of Problem (2.13):

$$\begin{aligned} &\min\ \sigma(x) \\ &\text{s.t. } \sigma(1) = 1,\ \sigma(M(-x^2)) \geq 0. \end{aligned} \tag{3.34}$$

By Remark 2.8, this problem is clearly Archimedean. The solution is, obviously, $\sigma^\star(x^n) = 0$, for $n \geq 1$. The weak ncKKT conditions would demand the existence of a positive linear functional μ, with $\mu(M(-x^2)) \geq 0$, and such that

$$-\mu(\{x, p\}) = \sigma^\star(p), \quad \forall p \in \mathbb{C}\langle x \rangle. \tag{3.35}$$

Taking $p = x^n$, this implies that

$$\mu(x) = -\frac{1}{2},\ \mu(x^n) = 0,\ \forall n \geq 2. \tag{3.36}$$

The value of $\mu(1)$ is not fixed by relation (3.35). This means that, for every order $k \geq 2$, the moment matrix $M_k(\mu)$ is of the form

$$M_k(\mu) = \begin{pmatrix} \mu(1) & -\frac{1}{2} & 0 & \cdots \\ -\frac{1}{2} & 0 & 0 & \cdots \\ 0 & 0 & 0 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix} \tag{3.37}$$

This matrix cannot be made positive semidefinite, no matter the choice of $\mu(1)$. Thus, Problem (3.34) does not satisfy weak ncKKT. Note, however, that, for any $\epsilon > 0$, $M_k(\mu) + \epsilon\mathbb{I}$ is positive semidefinite if one takes $\mu(1) = \frac{1}{\epsilon}$. In addition, $-\mu(p^* x^2 p) = 0$, for all $p \in \mathbb{C}\langle x \rangle$, and so the localizing matrix associated to the constraint $-x^2 \geq 0$ is positive semidefinite. Thus, Problem (3.34) satisfies essential ncKKT and the first inclusion relation of (3.33) is strict.

Next, consider the NPO problem

$$\begin{aligned} &\min\ \sigma(x) \\ &\text{s.t. } \sigma(1) = 1,\ \sigma(I(x^2)) = 0. \end{aligned} \tag{3.38}$$

This problem is also Archimedean. Its only solution is $\sigma^\star(x^n) = 0$, for all $n \geq 1$, which is bounded. Weak ncKKT demands that there exists a Hermitian $\lambda : \mathbb{C}\langle x \rangle \to \mathbb{C}$ such that

$$\begin{aligned} &\lambda(\{x, p\}) = \sigma^\star(p), \quad \forall p \in \mathbb{C}\langle x \rangle, \\ &\lambda(I(x^2)) = 0. \end{aligned} \tag{3.39}$$

The above is equivalent to

$$\begin{aligned} \lambda(x) &= \frac{1}{2}, \\ \lambda(x^n) &= 0, \quad \forall n \geq 2. \end{aligned} \tag{3.40}$$

These relations clearly give rise to a Hermitian linear functional $\lambda : \mathbb{C}\langle x \rangle \to \mathbb{C}$ satisfying (3.39) so Problem (3.38) satisfies weak ncKKT. However, Problem (3.38) does not satisfy normed ncKKT. Indeed, the requirement that the 2×2 matrices $M_{kl}^{\pm} := \lambda^{\pm}(p_k p_l^*)$, with $p_1 = 1, p_2 = x$, be positive semidefinite implies that the off-diagonal elements $\lambda^{\pm}(x)$ vanish, and so $\lambda(x) = 0$. □

In Section 4 we derive constraint qualification criteria which, similarly to the classical case, ensure that each of the ncKKT conditions holds for a given NPO problem. These are the main conclusions:

1. All NPO problems satisfy the essential ncKKT conditions (Theorem 4.1).

2. Under Archimedeanity and the (verifiable) algebraic condition (4.34), all the SDP relaxations derived from weak ncKKT hold (Theorem 4.5). That is, for any $k \in \mathbb{N}$, it is sound to add the SDP constraints (3.28) to the SDP relaxation (2.15).

3. Normed ncKKT holds for Archimedean problems if the operator constraints of the NPO problem satisfy a non-commutative, algebraic version of Mangasarian–Fromovitz constraint qualification [20] (Theorem 4.17).

4. NPO problems admitting an exact sum-of-squares resolution satisfy strong ncKKT (Theorem 4.19).

As it turns out, some interesting NPO problems in quantum information theory seem to satisfy neither normed nor strong ncKKT. Such NPO problems have a very peculiar structure, namely, the set of variables is partitioned in such a way that: (a) all variables within a partition commute with all the variables within any other partition; (b) there are no further operator equalities or inequalities relating variables from different partitions. In view of this, in Section 5 we further define variants of strong and normed ncKKT where Eqs. (3.25b) or (3.26c) only apply to polynomials of non-commuting variables within a partition. We identify sufficient (and very mild) criteria for those 'partial' operator optimality conditions to hold (Theorems 5.1 and 5.2). In Section 6.2, we show that the strongest of them, partial strong ncKKT, is satisfied by the NPO formulation of the quantum nonlocality problem.

3.3. A numerical example

We next apply the numerical tools presented so far to a problem adapted from Nie's [27, Example 5.4]. Let $x = (x_1, x_2, x_3)$, and, for $\delta, r_{\min}, r_{\max} \in \mathbb{R}^+$, with $r_{\min} < r_{\max}$, consider the following NPO problem:

$$\begin{aligned} &\min \ \sigma\left(f(x)\right), \\ &\text{s.t. } r_{\min} \leq x_1^2 + x_2^2 + x_3^2 \leq r_{\max}, \\ &\qquad \delta \pm i[x_k, x_l] \geq 0, \quad (k,l) = (1,2),\ (2,3),\ (3,1), \end{aligned} \tag{3.41}$$

where the objective function

$$f(x) := \frac{1}{2}\{x_1^2, x_2^4\} + \frac{1}{2}\{x_1^4, x_2^2\} + x_3^6 - \frac{3}{2}\left(x_1^2 x_2^2 x_3^2 + x_3^2 x_2^2 x_1^2\right) \tag{3.42}$$

is a symmetrized version of the Motzkin polynomial [46].

The constraint involving $r_{\max}$ ensures that this problem is Archimedean. Hence, it satisfies both essential ncKKT (see Theorem 4.1 below) and the state optimality conditions (3.2).

Let us see what enforcing the essential ncKKT conditions entails. To begin with, we need to invoke the existence of eight Lagrange multipliers: $\mu^{\max}, \mu^{\min}, \{\mu^{\pm}_{kl}\}_{kl}$, one for each inequality constraint. Each such multiplier is required to have moment and localizing matrices ϵ-away from positive semidefinite matrices, see the first two lines of Eq. (3.30).

Complementary slackness –Eq. (3.30d)– translates to

$$\mu^{\max}\left(\left(r_{\max}-\sum_k x_k^2\right)s\right)=\mu^{\max}\left(s\left(r_{\max}-\sum_k x_k^2\right)\right)=0, \tag{3.43a}$$

$$\mu^{\min}\left(\left(\sum_k x_k^2-r_{\min}\right)s\right)=\mu^{\min}\left(s\left(\sum_k x_k^2-r_{\min}\right)\right)=0, \tag{3.43b}$$

$$\mu^{\pm}_{k,k+1}\left(\left(\delta\pm i[x_k,x_{k+1}]\right)s\right)=\mu^{\pm}_{k,k+1}\left(s\left(\delta\pm i[x_k,x_{k+1}]\right)\right)=0,\quad k=1,2,3, \tag{3.43c}$$

for all monomials $s\in\mathbb{C}\langle x\rangle$. In the equation above, we respectively identify $\mu^{\pm}_{3,4}, x_4$ with $\mu^{\pm}_{3,1}, x_1$.

Finally, Eq. (3.30f) reduces to:

$$\begin{aligned}\sigma^\star\left(\nabla_x f(p)\right)=&-\mu^{\max}\left(\sum_k\{x_k,p_k\}\right)+\mu^{\min}\left(\sum_k\{x_k,p_k\}\right)\\&+i\sum_k\left\{\mu^{+}_{kl}\left([x_k,p_{k+1}]+[p_k,x_{k+1}]\right)-\mu^{-}_{kl}\left([x_k,p_{k+1}]+[p_k,x_{k+1}]\right)\right\},\end{aligned} \tag{3.44}$$

for any tuple of polynomials $p=(p_1,p_2,p_3)$ (again, x_4,p_4 are to be understood as x_1,p_1).

Setting $\delta=0.01$, $r_{\min}=1$, and $r_{\max}=7.5$, we ran the SDP relaxation (2.15) for $k=4$, with and without essential ncKKT and state optimality conditions. Table 3.1 shows the resulting lower bounds on the solution of Problem (3.41)[2].

	without ncKKT	essential ncKKT
without state optimality	−1.9774	−1.8179
with state optimality	−1.8522	−1.4266

Table 3.1. Lower bounds on the solution of Problem (3.41), computed at level $k=4$ of the hierarchy of SDP relaxations (2.15), under different optimality constraints. Essential ncKKT was enforced with $\epsilon_i=0.0001$ for all i.

This example illustrates that even the very general essential ncKKT conditions can greatly boost the speed of convergence of the hierarchy of SDPs (2.15), especially when combined with the state optimality conditions (3.2).

4. NON-COMMUTATIVE CONSTRAINT QUALIFICATION

In this section, we provide sufficient criteria for the different variants of the ncKKT conditions to hold. The roadmap from essential ncKKT to strong ncKKT is as follows:

- In section 4.1, we prove that all NPO problems satisfy the essential ncKKT conditions. Moreover, we will further identify, in subsection 4.1.1, 'boundedness conditions' that allow replacing the variables μ_i^k in Definition 3.17 by bounded, positive linear functionals $\{\mu_i\}_i$ independent of k and satisfying (3.27a) and (3.27b). This is not the same as having weak ncKKT conditions: those would in addition demand the existence of k-independent linear functionals $\{\lambda_j\}_j$. NPO problems satisfying the boundedness condition admit, however, the same SDP relaxations as problems satisfying weak ncKKT.

[2] A MATLAB implementation of this is available at https://github.com/ajpgarner/nckkt-nie-motzkin.

- To prove the existence of k-independent $\{\lambda_j\}_j$, we introduce in section 4.2.1 the notion of linear independence of non-commutative gradients. If both the boundedness condition and gradient linear independence hold, then one can prove that there exist bounded, k-independent linear functionals $\{\mu_j\}_i$, $\{\lambda_j\}_j$ complying with the conditions of normed ncKKT, see section 4.2.2.

- We were not able to carry this argument any further to find even more stringent conditions that would guarantee strong ncKKT. Thus, we switched to a completely different route: in section 4.3, we prove that NPO problems admitting an exact SOS resolution satisfy strong ncKKT.

4.1. Essential KKT conditions

In this section we will prove the universal validity of the essential ncKKT conditions.

Theorem 4.1. *Any* bounded *solution* $\sigma^\star$ *of the NPO problem* (2.13) *satisfies Eq.* (3.30)*. That is, all NPO problems satisfy the essential ncKKT conditions.*

Remark 4.2. Setting $\epsilon_i = 0$ for all i, essential ncKKT reduces to the SDP constraints conditions (3.28) implied by weak ncKKT on the SDP relaxation (2.15). The reason why Theorem 4.1 demands non-zero ϵ_i is to prevent that $\mu_i^k(p) = \infty$ for some polynomials $p \in \mathbb{C}\langle x\rangle_{2k}$. See, e.g., the NPO problem (3.34): the requirement that the moment and localizing matrices of μ^k be ϵ-close to positive implies that $\mu(1) \geq O\left(\frac{1}{\epsilon}\right)$.

To prove Theorem 4.1, we will need the following technical result.

Lemma 4.3. *Let the* n*-tuple of symmetric polynomials* $q(x)$ *satisfy*

$$\nabla_x h_j(q(x)) \in I(\mathbf{h}), \quad j = 1, \ldots, m', \tag{4.1}$$

and let $X^\star \in B(\mathcal{H}^\star)^n$ *be a tuple of Hermitian operators acting on the Hilbert space* $\mathcal{H}^\star$*, such that*

$$h_j(X^\star) = 0, \quad j = 1, \ldots, m'. \tag{4.2}$$

Then, there exists $\epsilon > 0$ *and an analytic trajectory* $\{X(t) : t \in [-\epsilon, \epsilon]\} \subset B(\mathcal{H}^\star)^n$ *of Hermitian operators such that*

$$h_j(X(t)) = 0, \quad j = 1, \ldots, m',\ t \in [-\epsilon, \epsilon], \tag{4.3a}$$

$$X(0) = X^\star, \tag{4.3b}$$

$$\frac{dX(t)}{dt}\Big|_{t=0} = q(X^\star). \tag{4.3c}$$

If, in addition, $g_i(X^\star) \geq 0$ *for all* i*, and, for some* $\{s_i \in \mathbb{C}\langle x\rangle\}_i$*, it holds that*

$$s_i(x)g_i(x) + g_i(x)s_i(x)^* + \nabla_x g_i(q(x)) \in M(\mathbf{g}) + I(\mathbf{h}), \quad i = 1, \ldots, m, \tag{4.4}$$

then the trajectory $X(t)$ *also satisfies:*

$$g_i(X(t)) \geq 0, \quad i = 1, \ldots, m, \tag{4.5}$$

for $t \in [0, \epsilon]$.

Proof. Consider the system of ordinary differential equations (ODEs)

$$\begin{aligned} \frac{dX(t)}{dt} &= q(X), \\ X(0) &= X^\star. \end{aligned} \tag{4.6}$$

Since $X_1^\star, \ldots, X_n^\star$ are bounded, we can apply the Cauchy-Kovalevskaya theorem (in Banach spaces, see, e.g., [47] or [48][3]) and conclude that there exists a ball in the complex plane of radius $\epsilon > 0$ and with center at 0 where the solution of this differential equation is analytic. In particular, for $t \in [-\epsilon, \epsilon]$, $X(t)$ exists and, from the equation above, it satisfies the boundary conditions (4.3b) and (4.3c).

Being polynomials of a tuple of analytic operators, $\{h_j(X(t))\}_j$ are also analytic in the region $\{t \in \mathbb{C} : |t| \leq \epsilon\}$. Therefore, for $t \in [-\epsilon, \epsilon]$,

$$h_j(X(t)) = \sum_{k=0}^{\infty} \frac{t^k}{k!} \frac{d^k h_j(X(\tau))}{d\tau^k}\Big|_{\tau=0}. \tag{4.7}$$

Due to relation (4.1), we have that

$$\begin{aligned} \frac{dh_j(X(t))}{dt} &= \nabla_x h_j \left(\frac{dX(t)}{dt} \right) = \nabla_x h_j(q(X(t))) \\ &= \sum_{j',l} r^+_{jj'l}(X(t))\, h_{j'}(X(t))\, r^-_{jj'l}(X(t)). \end{aligned} \tag{4.8}$$

By induction, it therefore holds that

$$\frac{d^k h_j(X(t))}{dt^k} = \sum_{j',l} p^+_{jj'kl}(X(t))\, h_{j'}(X(t))\, p^-_{jj'kl}(X(t)), \tag{4.9}$$

for some polynomials $\{p^\pm_{jj'kl} : j, j', k, l\} \subset \mathbb{C}\langle x \rangle$. Since $h_{j'}(X(0)) = 0$ for all j', we have that

$$\frac{d^k h_j(X(\tau))}{d\tau^k}\Big|_{\tau=0} = 0, \quad \forall k. \tag{4.10}$$

Consequently, by Eq. (4.7), $h_j(X(t)) = 0$, for $t \in [-\epsilon, \epsilon]$.

Let us now assume that Eq. (4.5) holds. Define the function $G_i(t) := \omega_i(t) g_i(X(t)) \omega_i(t)^*$, where $\omega_i(t)$ is the solution of the differential equation:

$$\frac{d\omega_i(t)}{dt} = \omega_i(t) s_i(X(t)), \quad \omega_i(0) = \mathbb{I}, \tag{4.11}$$

which, by the Cauchy-Kovalevskaya theorem, is analytic for short times t. For short times, $\omega(t)$ is also close to the identity, and thus invertible. Let $\epsilon' > 0$ be such that $\epsilon \geq \epsilon'$ and $\{\omega_i(t)\}_i$ are analytic and invertible for $t \in [0, \epsilon']$.

Being the product of three analytic functions, the function $G_i(t)$ is also analytic in $t \in [0, \epsilon']$. Taking differentials, we find

$$\begin{aligned} \frac{dG_i(t)}{dt} &= \omega_i(t)(s_i(X(t))g_i(X(t)) + g_i(X(t))s_i^*(X(t)) + \nabla_x g_i(q(X(t)))\omega_i(t)^* \\ &\overset{(4.4)}{=} \omega_i(t) \left(\sum_l s^*_{il}(X(t)) s_{il}(X(t)) + \sum_{i',l} s^*_{ii'l}(X(t)) g_{i'}(X(t)) s_{ii'l}(X(t)) \right. \end{aligned}$$

[3] The basic idea of the proof of the Cauchy-Kovalevskaya theorem is fairly simple. Differentiate the ODE to compute a formal power series for $X(t)$ about 0. The method of majorants was developed by Cauchy and Kovalevskaya to prove that this power series has a positive radius of convergence and thus determines an analytic function that solves the ODE.

$$+\sum_{j,l} s^{+}_{ijl}(X(t))h_j(X(t))s^{-}_{ijl}(X(t))\Bigg)\,\omega_i(t)^*$$
$$\overset{(4.3a)}{=} \omega_i(t)\left(\sum_{l} s^{*}_{il}(X(t))s_{il}(X(t)) + \sum_{i',l} s^{*}_{ii'l}(X(t))\omega_{i'}(t)^{-1}G_{i'}(t)(\omega_{i'}(t)^{-1})^* s_{ii'l}(X(t))\right)\omega_i(t)^*. \tag{4.12}$$

Since all the operators are bounded, this equation might be solved through any standard numerical method, e.g.: Euler's explicit method. Take $\Delta > 0$ and consider the following time discretization:

$$\begin{aligned} G_i(k+1;\Delta) &= G_i(k;\Delta) + \Delta\omega_i(\Delta k)\left(\sum_{l} s^{*}_{il}(X(\Delta k))s_{il}(X(\Delta k))\right)\omega_i(\Delta k)^* \\ &+ \Delta\omega_i(\Delta k)\left(\sum_{i'l} s^{*}_{ii'l}(X(\Delta k))\omega_{i'}(\Delta k)^{-1}G_{i'}(k;\Delta))(\omega_{i'}(\Delta k)^{-1})^* s_{ii'l}(X(\Delta k))\right)\omega_i(\Delta k)^*. \end{aligned} \tag{4.13}$$

Clearly, starting from positive semidefinite operators

$$G_i(0;\Delta) = g_i(X^\star) \geq 0, \quad i = 1,\ldots,m, \tag{4.14}$$

it holds that $G_i(k;\Delta) \geq 0$, for all k.

Suppose that the Euler explicit method converged: namely, for any $t \in [0,\epsilon']$, $\lim_{k\to\infty} G_i(k;\frac{t}{k}) = G_i(t)$, for all i. Then, the observation above would imply that $\omega_i(t)g_i(X(t))\omega_i(t)^* \geq 0$, for $t \in [0,\epsilon']$ and for all i. Since $\omega_i(t)$ is invertible, this would imply that $g_i(X(t)) \geq 0$ and thus Eq. (4.5) would hold.

To finish the proof, it thus suffices to show that the Euler method converges. This is not difficult: we just need to follow the proof for ordinary differential equations. First, for fixed Δ, define $t_k := k\Delta$ and the error

$$\delta_k := \max_i \|G_i(t_k) - G_i(k;\Delta)\|. \tag{4.15}$$

Also, rewrite the right-hand side of Eq. (4.12) as

$$\frac{dG(t)}{dt} = \rho(t, G(t)). \tag{4.16}$$

Note that $\rho(t,G(t))$ is affine linear in $G(t)$. This implies that it is Lipschitz, i.e., for all $t \in [0,\epsilon']$, it holds that $\|\rho(t,G) - \rho(t,G')\| \leq K\|G - G'\|$, for some $K \in \mathbb{R}^+$. Eq. (4.16) implies that

$$G(t) = \int_0^t d\tau\rho(\tau, G(\tau)).^{4} \tag{4.17}$$

Integrating by parts, we have that

$$G(t_{j+1}) = G(t_j) + \rho(t_j, G(t_j))\Delta + \int_{t_j}^{t_{j+1}} d\tau(t_{j+1}-\tau)\frac{d}{d\tau}\rho(\tau, G(\tau)). \tag{4.18}$$

Subtracting on both sides by $G(j+1;\Delta) = G(j;\Delta) + \Delta\rho(t_j, G(j;\Delta))$, we arrive at

$$\begin{aligned} \delta_{j+1} &\leq (1+K\Delta)\delta_j + \left\|\int_{t_j}^{t_{j+1}} d\tau(t_{j+1}-\tau)\frac{d}{d\tau}\rho(\tau,G(\tau))\right\| \\ &\leq (1+K\Delta)\delta_j + \left(\int_{t_j}^{t_{j+1}}(t_{j+1}-\tau)^2 d\tau\right)^{1/2}\left(\int_{t_j}^{t_{j+1}} d\tau\left\|\frac{d}{d\tau}\rho(\tau,G(\tau))\right\|^2\right)^{1/2} \\ &\leq (1+K\Delta)\delta_j + \gamma\Delta^{3/2}, \end{aligned} \tag{4.19}$$

[4] Here and below we only use the Riemann integral in Banach spaces, where the basic properties are essentially the same as those in Euclidean spaces. For details and a deeper study in integration in Banach spaces we refer to one the standard works, e.g., [49–51]

with

$$\gamma = \frac{1}{\sqrt{3}} \left(\int_0^{\epsilon'} d\tau \left\| \frac{d}{d\tau} \rho(\tau, G(\tau)) \right\|^2 \right)^{1/2}. \tag{4.20}$$

It can be verified, by induction, that

$$\delta_j \leq e^{(\gamma+K)t_j} \Delta^{1/2}. \tag{4.21}$$

Indeed, $\delta_0 = 0 \leq \Delta^{1/2}$. Now, suppose that the equation (4.21) above holds for j. Then, by Eq. (4.19), we have that

$$\begin{aligned} \delta_{j+1} &\leq (1 + K\Delta) e^{(\gamma+K)t_j} \Delta^{1/2} + \gamma \Delta^{3/2} \\ &\leq (1 + K\Delta) e^{(\gamma+K)t_j} \Delta^{1/2} + e^{(\gamma+K)t_j} \gamma \Delta^{3/2} \\ &= (1 + (K+\gamma)\Delta) e^{(\gamma+K)t_j} \Delta^{1/2} \\ &\leq e^{(K+\gamma)\Delta} e^{(\gamma+K)t_j} \Delta^{1/2} \\ &= e^{(\gamma+K)t_{j+1}} \Delta^{1/2}. \end{aligned} \tag{4.22}$$

For any $j \in \{0, \ldots, \left\lfloor \frac{\epsilon'}{\Delta} \right\rfloor\}$, it thus follows that $\delta_j \leq e^{(\gamma+K)\epsilon'} \Delta^{1/2}$, which tends to zero in the limit $\Delta \to 0$. Thus, the Euler explicit method converges. □

Proof of Theorem 4.1. It is enough to prove the statement for $\epsilon_i = \epsilon > 0$, for all i. Let $\sigma^\star$ be a bounded minimizer of Problem (2.13). For fixed $k \in \mathbb{N}$, consider the semidefinite program

$$\begin{aligned} \mathbb{P} := \min_{\epsilon, \mu, \lambda} \ & \epsilon \\ \text{s.t. } & \epsilon \geq 0, \\ & \mu_i^k(pp^*) + \epsilon \|p\|_2^2 \geq 0, \quad \forall p \in \mathbb{C}\langle x \rangle,\ \deg(p) \leq k, && \text{(4.23a)} \\ & \mu_i^k(pg_l p^*) + \epsilon \|p\|_2^2 \geq 0, \quad \forall p \in \mathbb{C}\langle x \rangle,\ \deg(p) \leq k - \left\lceil \frac{\deg(g_l)}{2} \right\rceil,\ i = 1, \ldots, m,\ l = 1, \ldots, m, && \text{(4.23b)} \\ & \mu_i^k(I(\mathbf{h})_{2k}) = 0, i = 1, \ldots, m, && \text{(4.23c)} \\ & \mu_i^k(sg_i) = \mu_i^k(g_i s) = 0, \quad \forall s \in \mathbb{C}\langle x \rangle,\ \deg(s) \leq 2k - \deg(g_i) \quad i = 1, \ldots, m, && \text{(4.23d)} \\ & \lambda_j^k(I(\mathbf{h})_{2k}) = 0, j = 1, \ldots, m', && \text{(4.23e)} \\ & \sigma^\star \left(\nabla_x f(p(x)) \right) - \sum_i \mu_i^k \left(\nabla_x g_i(p(x)) \right) - \sum_j \lambda_j^k \left(\nabla_x h_j(p(x)) \right) = 0, \quad \forall p \in \mathbb{C}\langle x \rangle^n, && \text{(4.23f)} \\ & \qquad \deg\left(\nabla_x f \right), \deg\left(\nabla_x g_i \right), \deg\left(\nabla_x h_j \right) \leq 2k - \deg(p), && \text{(4.23g)} \end{aligned}$$

where the minimization takes place over Hermitian functionals $\{\mu_i^k\}_i, \{\lambda_j^k\}_j$. We next prove that this problem has feasible points.

First, let us focus on Eqs. (4.23c)–(4.23f). Let $\mathcal{M}$ be the set of polynomials of the form $r + r^*$ and $i(r - r^*)$, where $r \in \langle x \rangle$. Note that all the elements of $\mathcal{M}$ are Hermitian. Moreover, $\mathcal{M}_k$, the set of elements of $\mathcal{M}$ with degree smaller than or equal to k, satisfies $\text{span}(\mathcal{M}_k) = \mathbb{C}\langle x \rangle_k$. With this notation, satisfying Eqs. (4.23c)–(4.23f) with Hermitian functionals can be seen equivalent to the feasibility problem:

$$\begin{aligned} & \exists \{\mu_i^k(p) : p \in \mathcal{M}_{2k}\}_i, \{\lambda_j^k(p) : p \in \mathcal{M}_{2k}\}_j \subset \mathbb{R}, \\ \text{s.t. } & \mu_i^k \left(r^+ h_j r^- + (r^-)^* h_j (r^+)^* \right) = 0,\ \forall r^+, r^- \in \langle x \rangle,\ \deg(r^-) + \deg(r^+) \leq \deg(h_j),\ \forall i, j \\ & \mu_l^k \left(i(r^+ h_j r^- - (r^-)^* h_j (r^+)^*) \right) = 0,\ \forall r^+, r^- \in \langle x \rangle,\ \deg(r^-) + \deg(r^+) \leq \deg(h_j),\ \forall j, l, \end{aligned}$$

$$
\begin{aligned}
&\mu_l^k(sg_l + g_l s^*) = 0, \quad \forall s \in \langle x \rangle,\, \deg(s) \le 2k - \deg(g_l),\, \forall l,\\
&\mu_l^k(i(sg_l - g_l s^*)) = 0, \quad \forall s \in \langle x \rangle,\, \deg(s) \le 2k - \deg(g_l),\, \forall l,\\
&\lambda_l^k\left(r^+ h_j r^- + (r^-)^* h_j (r^+)^*\right) = 0,\ \forall r^+, r^- \in \langle x \rangle,\, \deg(r^-) + \deg(r^+) \le \deg(h_j),\, \forall j, l,\\
&\lambda_l^k\left(i(r^+ h_j r^- - (r^-)^* h_j (r^+)^*)\right) = 0,\ \forall r^+, r^- \in \langle x \rangle,\, \deg(r^-) + \deg(r^+) \le \deg(h_l),\, \forall j, l,\\
&\sum_i \mu_i^k\left(\nabla_x g_i(r)\right) + \sum_j \lambda_j^k\left(\nabla_x h_j(r)\right) = \sigma^\star\left(\nabla_x f(r)\right), \quad \forall r \in \mathcal{M}^n,\\
&\qquad \deg\left(\nabla_x f\right), \deg\left(\nabla_x g_i\right), \deg\left(\nabla_x h_j\right) \le 2k - \deg(r).
\end{aligned} \tag{4.24}
$$

The left-hand side of each constraint can be expressed as a real linear combination of the real variables defined in the first line. The feasibility problem (4.24) is therefore a real system of linear equations. Note that the first conditions are homogeneous (namely, satisfied by taking all variables equal to zero); the only non-homogeneous conditions can be found in the last equality in the line next to the last of Eq. (4.24).

An arbitrary real linear combination of the equalities of the last line results in an expression of the form

$$
\sum_i \mu_i^k\left(\nabla_x g_i(p)\right) + \sum_j \lambda_j^k\left(\nabla_x h_j(p)\right) = \sigma^\star\left(\nabla_x f(p)\right), \tag{4.25}
$$

where the vector of polynomials $p \in \mathbb{C}\langle x \rangle^n$ has Hermitian entries. By basic linear algebra, the system is not solvable iff, for some $p \in \mathbb{C}\langle x \rangle^n$, the right-hand side of the equation above is non-zero, whereas the left-hand side can be expressed a real linear combination of the homogeneous constraints. Since the homogeneous constraints do not relate the different variable types (namely, those that go with each functional), this would imply that the polynomial evaluated by λ_j^k belongs to $I(\mathbf{h})_{2k}$, and the polynomial evaluated by μ_i^k is of the form $s_i g_i + g_i s_i^* + r_i$, for some polynomials $s_i \in \mathbb{C}\langle x \rangle$, $r_i \in I(\mathbf{h})_{2k}$.

We have just proven that there exist Hermitian λ, μ satisfying Eqs. (4.23c)–(4.23f) iff, for all Hermitian $p \in \mathbb{C}\langle x \rangle^n$ and (not necessarily Hermitian) $\{s_i\}_i \subset \mathbb{C}\langle x \rangle$, the conditions

$$
\begin{aligned}
&\deg\left(\nabla_x f\right), \deg\left(\nabla_x g_i\right), \deg\left(\nabla_x h_j\right) \le 2k - \deg(p),\\
&\deg(s_i) \le 2k - \deg(g_i), i = 1, \ldots, m,
\end{aligned} \tag{4.26}
$$

and

$$
\begin{aligned}
&\nabla_x g_i(p) - s_i g_i - g_i s_i^* \in I(\mathbf{h})_{2k}, i = 1, \ldots, m,\\
&\nabla_x h_j(p) \in I(\mathbf{h})_{2k}, j = 1, \ldots, m',
\end{aligned} \tag{4.27}
$$

imply that

$$
\sigma^\star(\nabla_x f(p)) = 0. \tag{4.28}
$$

Let $(\mathcal{H}^\star, X^\star, \psi^\star)$ be the result of applying the GNS construction on $\sigma^\star$, and suppose that $p \in \mathbb{C}\langle x \rangle^n$ satisfies Eqs. (4.26) and (4.27). Then, $X^\star, q := \pm p$ satisfy the conditions of Lemma 4.3, and thus there exist two feasible operator trajectories $\{X^\pm(t) : t \in [0, \delta]\} \subset B(\mathcal{H}^\star)$ such that

$$
X^\pm(0) = X^\star,\ \left.\frac{dX^\pm}{dt}\right|_{t=0} = \pm p(X^\star). \tag{4.29}
$$

Hence,

$$
\pm\sigma^\star(\nabla_x f(p)) = \sigma^\star(\nabla_x f(\pm p)) = \psi^\star\left(\nabla_x f(\pm p(X^\star))\Big|_{X=X^\star}\right) = \left.\frac{d\psi^\star\left(f(X^\pm(t))\right)}{dt}\right|_{t=0} \ge 0, \tag{4.30}
$$

where the last inequality follows from the fact that $X = X^\star$ minimizes the expression $\psi^\star(f(X))$ over tuples $X \in B(\mathcal{H}^\star)^n$ of feasible operators. Thus, Eq. (4.28) holds whenever p, s satisfy Eqs. (4.26), (4.27).

It follows that there exist μ^k, λ^k satisfying conditions (4.23c)–(4.23f). For each i, the corresponding moment and localizing matrices of μ_i^k might not be positive semidefinite, but, one can find $\epsilon > 0$ large enough such that conditions (4.23a), (4.23b) hold. We therefore conclude that the SDP (4.23) is feasible.

The dual of (4.23) is the following SDP:

$$\begin{aligned}
\mathbb{P}^* := \max_{q,s,Z} \; & -\sigma^\star(\nabla_x f(q)) \\
\text{s.t. } & \nabla_x g_i(q) + s_i g_i + g_i s_i^* - \sum_{a,b} (Z_i)_{ab} o_a g_i o_b^* - \sum_{l,a,b} (Z_{i,l})_{ab} o_a^l g_l (o_b^l)^* \in I(\mathbf{h})_{2k}, \\
& \nabla_x h_j(q) \in I(\mathbf{h})_{2k}, \quad j = 1, \ldots, m', \\
& Z_i \geq 0, \; Z_{i,l} \geq 0, \quad i, k = 1, \ldots, m, \\
& \sum_{i,l} \operatorname{tr}(Z_{i,l}) \leq 1.
\end{aligned} \tag{4.31}$$

where $\{o_k\}_k$, $\{o_k^l\}_k$ are, respectively, monomial bases for polynomials of degree k and $k - \left\lceil \frac{\deg(g_l)}{2} \right\rceil$, and Z_i, $Z_{i,l}$ are Hermitian matrices.

That is, in Problem (4.31) one needs to maximize $-\sigma^\star(\nabla_x f(q))$ over the polynomials $q, \{s_i\}_i$ such that $\nabla_x h_j(q) = 0$ for all j, and the polynomials $\nabla_x g_i(q) + s_i g_i + g_i s_i^*$, $i = 1, \ldots, m$ are sums of weighted squares satisfying certain normalization constraints.

We claim that the solution of (4.31) is zero. This value can be achieved, e.g., by taking $q = Z_i = Z_{i,l} = s_i = 0$ for all i, l. To see that zero is the optimal value, consider any feasible point q, s of (4.31). By definition, $q, X^\star$ satisfy the conditions of Lemma 4.3. Thus, there exists $\delta > 0$ and a trajectory of feasible operators $\{X(t) : t \in [0, \delta]\}$ such that

$$X(0) = X^\star, \; \frac{dX}{dt}\Big|_{t=0} = q(X^\star). \tag{4.32}$$

Hence,

$$-\sigma^\star(\nabla_x f(q)) = -\frac{d\psi^\star(f(X(t)))}{dt}\Big|_{t=0} \leq 0. \tag{4.33}$$

The solution of Problem (4.31) is therefore zero.

Now, the SDP (4.23) is bounded from below by 0 and it admits strictly feasible points (by taking ϵ large enough). By Slater's criterion, the problem thus satisfies strong duality, and so the solutions of (4.23), (4.31) coincide. This implies that one can find feasible points of Problem (4.23) for any $\epsilon > 0$. Hence, conditions (3.30) hold for arbitrary $\{\epsilon_i\}_i \subset \mathbb{R}^+$. □

Remark 4.4. Note that our derivation of Theorem 4.1 does not make full use of the fact that $(\mathcal{H}^\star, X^\star, \psi^\star)$ is an optimal solution of Problem 2.5. It is enough to assume that, for a fixed state $\psi^\star$, the operators $X^\star$ minimize the objective function $\psi^\star(f(X))$ (under the corresponding operator constraints). With this in mind, it is easy to re-derive, via Theorem 4.1, the first-order optimality conditions for synchronous non-local games presented in [1, Prop. 7.1], see Appendix B.

4.1.1. Setting ϵ_i to zero

In Remark 4.2, the presence of $\epsilon_i > 0$ in Theorem 4.1 was attributed to the impossibility of bounding the norm of the positive semidefinite multipliers $\{\mu_i\}_i$. If this intuition were accurate, then one would expect that any algebraic bound on the norm of μ_i would allow one to set $\epsilon_i = 0$ in Eq. (3.30). This is exactly what the next theorem shows.

Theorem 4.5. *Consider an NPO problem (2.13) that satisfies the Archimedean condition. For some set of indices $I \subset \{1,\dots,m\}$, let there exist $r \in \mathbb{R}^+$ and a tuple of Hermitian polynomials $q \in \mathbb{C}\langle x\rangle^n$ such that*

$$\begin{aligned}
&\nabla_x g_i(q) + s_i g_i + g_i s_i^* - r \in M(\mathbf{g}) + I(\mathbf{h}), \quad \forall i \in I,\\
&\nabla_x g_i(q) + s_i g_i + g_i s_i^* \in M(\mathbf{g}) + I(\mathbf{h}), \quad \forall i \notin I,\\
&\nabla_x h_j(q) \in I(\mathbf{h}), \quad j = 1,\dots,m'.
\end{aligned} \tag{4.34}$$

for some polynomials $\{s_i : i = 1,\dots,m\}$. Then, there exist Hermitian linear functionals $\{\mu_i : \mathbb{C}\langle x\rangle \to \mathbb{C}\}_{i\in I}$, with $\mu_i(M(\mathbf{g}) + I(\mathbf{h})) \geq 0$ for all $i \in I$, such that, for any $k \in \mathbb{N}$ and $\{\epsilon_i : i \notin I\} \subset \mathbb{R}^+$, there exist Hermitian functionals $\{\mu_i^k : \mathbb{C}\langle x\rangle_{2k} \to \mathbb{C}\}_{i\notin I}$, $\{\lambda_j^k : \mathbb{C}\langle x\rangle_{2k} \to \mathbb{C}\}_j$ satisfying

$$\begin{aligned}
&\mu_i^k(pp^*) + \epsilon_i \|p\|_2^2 \geq 0, \quad \forall p \in \mathbb{C}\langle x\rangle,\ \deg(p) \leq k,\ \forall i \notin I\\
&\mu_i^k(pg_l p^*) + \epsilon_i \|p\|_2^2 \geq 0, \quad \forall p \in \mathbb{C}\langle x\rangle,\ \deg(p) \leq k - \left\lceil \frac{\deg(g_l)}{2} \right\rceil,\ l = 1,\dots,m,\ \forall i \notin I\\
&\mu_i^k(I(\mathbf{h})_{2k}) = 0, \quad \forall i \notin I,\\
&\mu_i^k(sg_i) = \mu_i^k(g_i s) = 0, \quad \forall s \in \mathbb{C}\langle x\rangle,\ \deg(s) \leq 2k - \deg(g_i),\ \forall i \notin I,\\
&\lambda_j^k(I(\mathbf{h})_{2k}) = 0, \quad j = 1,\dots,m',\\
&\sum_{i\notin I} \mu_i^k\left(\nabla_x g_i(p(x))\right) + \sum_j \lambda_j^k\left(\nabla_x h_j(p(x))\right) = \sigma^\star\left(\nabla_x f(p(x))\right) - \sum_{i\in I} \mu_i\left(\nabla_x g_i(p(x))\right), \quad \forall p \in \mathbb{C}\langle x\rangle^n,\\
&\qquad \deg\left(\nabla_x f\right), \deg\left(\nabla_x g_i\right), \deg\left(\nabla_x h_j\right) \leq 2k - \deg(p).
\end{aligned} \tag{4.35}$$

In particular, the essential ncKKT conditions (3.30) hold with $\epsilon_i = 0$ for all $i \in I$.

Remark 4.6. Note that, if weak ncKKT holds, then one can use Eq. (4.34) to upper bound the norm of the Lagrange multipliers $\{\mu_i : i \in I\}$. Indeed, take $p = q$ in Eq. (3.27). By Eq. (4.34) it then holds that

$$\sigma^\star(\nabla_x f(q)) = \sum_i \mu_i(\nabla_x g_i(q)) = \sum_i \mu_i(\nabla_x g_i(q) + s_i g_i + g_i s_i^*) \geq r \sum_i \mu_i(1). \tag{4.36}$$

That is, for $i \in I$, $\mu_i(1)$, the norm of μ_i, is bounded above by $\frac{\sigma^\star(\nabla_x f(q))}{r}$.

To prove Theorem 4.5, we will need the following simple lemma.

Lemma 4.7. *Consider the system of linear equations in the variables $y \in \mathbb{R}^s$ given by*

$$A \cdot y = b, \tag{4.37}$$

where A is an $r \times s$ real matrix and $b \in \mathbb{R}^r$. Then, there exists a constant $K \in \mathbb{R}^+$ such that, for any $b' \in \mathrm{im}(A) \subset \mathbb{R}^r$ and any solution y of (4.37), there exists a solution y' of the system

$$A \cdot y' = b', \tag{4.38}$$

with $\|y - y'\|_2 \leq K\|b - b'\|_2$.

Proof. If system (4.37) is solvable, then any solution y thereof can be expressed as

$$y = A^+ \cdot b + \bar{y}, \tag{4.39}$$

where $\bar{y} \in \ker(A)$ and A^+ denotes the Moore-Penrose inverse of A. Given y, b', the expression $y' = \bar{y} + A^+ \cdot b'$ is thus a solution of Eq. (4.38), as long as Eq. (4.38) is solvable. Moreover,

$$\|y - y'\|_2 \leq \|A^+\|\|b - b'\|_2. \tag{4.40}$$

Defining $K := \|A^+\|$, we arrive at the statement of the lemma. □

Proof of Theorem 4.5. By Theorem 4.1, for any $k \in \mathbb{N}$ and any $\{\epsilon_i\}_i \subset \mathbb{R}^+$, there exist functionals $\{\mu_i^k\}_i$, $\{\lambda_j^k\}_j$ satisfying Eqs. (3.30). Without loss of generality, let us assume in the following that $\epsilon_i \leq 1$ for all i.

We next prove that the existence of a polynomial q satisfying conditions (4.34) implies that we can choose the functionals $\{\mu_i^k : i \in I\}$ such that the terms $\{\mu_i^k(p) : p \in \mathcal{M}_{2k}\}$ are bounded. Remember, from the proof of Theorem 4.1, that $\mathcal{M}_j$ is the (finite) set of polynomials of the form $s + s^*$ or $i(s - s^*)$, with $s \in \langle x \rangle_j$.

First, by virtue of the first three lines of Eq. (3.30), we have that, for any polynomial $p \in M(\mathbf{g})_{2k} + I(\mathbf{h})_{2k}$, with SOS decomposition given by Eq. (2.11) and

$$2\deg(p_l),\ 2\deg(p_{il}) + \deg(g_i),\ \deg(p_{jl}^+) + \deg(p_{jl}^+) + \deg(h_j) \leq 2k, \tag{4.41}$$

it holds that

$$\mu_i^k(p) + \epsilon_i \left(\sum_j \|p_j\|_2^2 + \sum_l \|p_{il}\|_2^2 \right) \geq 0. \tag{4.42}$$

Take $p = q$ (the polynomial tuple in Eq. (4.34)). Provided that k is large enough so that the SOS decompositions in Eq. (4.34) can be evaluated by $\{\mu_i^k\}_i$, this implies that there exist constants $\{K_i\}_i \subset \mathbb{R}^+$, independent of $\{\epsilon_i\}_i$, such that

$$\sigma^\star(\nabla_x f(q)) + \sum_i \epsilon_i K_i = \sum_i \mu_i^k(\nabla_x g_i(q)) + \sum_i \epsilon_i K_i \geq r \sum_{i \in I} \mu_i^k(1). \tag{4.43}$$

Since $\sigma^\star(\nabla_x f(q))$ is bounded (by the Archimedean condition) and $\mu_i^k(1) + 1 \geq \mu_i^k(1) + \epsilon_i \geq 0$, it follows that the values $\{|\mu_i^k(1)| : i \in I\}$ are upper bounded by a constant $C \in \mathbb{R}^+$.

Also by the Archimedean condition, for any Hermitian polynomial p, there exists $K_p \in \mathbb{R}^+$ such that $K_p \pm p$ is an SOS. Eq. (4.42) then implies that, for any $p \in \mathbb{C}\langle x \rangle$, there exists a constant C_p, independent of $\{\epsilon_i\}_i$, such that, for k large enough,

$$K_p \mu_i^k(1) \pm \mu_i^k(p) + \epsilon_i C_p \geq 0. \tag{4.44}$$

If k is also large enough to accommodate the SOS decompositions in Eq. (4.34), we conclude by Eq. (4.43) that $\{\mu_i^k(p) : i \in I\}$ are also bounded.

Thus, for given $k' \in \mathbb{N}$, we can find $k \geq k'$ and $C \in \mathbb{R}^+$ such that any feasible set of Lagrange multipliers μ^k, λ^k satisfies

$$|\mu_i^k(p)| \leq C, \quad \forall i \in I, \tag{4.45}$$

for all polynomials $p \in \mathcal{M}_{2k'}$. Obviously, the restrictions $\lambda^{k'}, \mu^{k'}$ of said Lagrange multipliers λ^k, μ^k to the smaller domain $\mathbb{C}\langle x \rangle_{2k'}$ also satisfy conditions (3.30), with the replacement $k \to k'$. Moreover, by the argument above, for all $p \in \mathbb{C}\langle x \rangle_{2k'}$, $i \in I$, $\mu_i^{k'}(p)$ is bounded.

We have just proven that, for any $k \in \mathbb{N}$, one can choose $\{\mu_i^k : i \in I\}$ bounded, independently of the value of $\{\epsilon_i\}_i$. Now, by Theorem 4.1, for each $k \in \mathbb{N}$ there exists feasible $\mu^k, \lambda^k : \mathbb{C}\langle x \rangle_{2k} \to \mathbb{C}$ satisfying Eq. (3.30) for $\epsilon_i = \frac{1}{k}$, with $\{\mu_i^k : i \in I\}$ bounded. Consider now the sequence of functionals $(\{\mu_i^k : \mathbb{C}\langle x \rangle \to \mathbb{C}\}_{i \in I})_k$ (we extend μ_i^k from $\mathbb{C}\langle x \rangle_{2k}$ to $\mathbb{C}\langle x \rangle$ by mapping all monomials of degree $> k$ to zero). Now, let $i \in I$. Since, for all monomials $m \in \mathbb{C}\langle x \rangle$, there exists L_m such that $|\mu_i^k(m)| \leq L_m$ for all k, we can construct a linear, invertible transformation $\mathbb{L}$ such that $|\mathbb{L} \circ \mu_i^k(m)| \leq 1$ for all monomials $m \in \mathbb{C}\langle x \rangle$. By the Banach-Alaoglu theorem [52, Theorem IV.21], the sequence $(\{\tilde{\mu}_i^k\}_{i \in I})_k$, with $\tilde{\mu}_i := \mathbb{L} \circ \mu_i^k$, thus admits a converging subsequence indexed by $(k_\alpha)_\alpha$, call $\{\tilde{\mu}_i : \mathbb{C}\langle x \rangle \to \mathbb{C}\}_{i \in I}$ its limit. Finally, we define $\mu_i := \mathbb{L}^{-1} \circ \tilde{\mu}_i$. By construction, we have that, for all $k \in \mathbb{N}$ and $i \in I$,

$$\lim_{\alpha \to \infty} \mu_i^{k_\alpha}\Big|_{\mathbb{C}\langle x \rangle_{2k}} = \mu_i\Big|_{\mathbb{C}\langle x \rangle_{2k}}. \tag{4.46}$$

This implies that, for $i \in I$ and any $\delta > 0$,

$$M^k(\mu_i) + \delta\mathbb{I} \geq 0,\ M_l^k(\mu_i) + \delta\mathbb{I} \geq 0. \tag{4.47}$$

It follows that $M^k(\mu_i),\ M_l^k(\mu_i) \geq 0$, for all k, and so $\{\mu_i\}_{i\in I}$ are positive linear functionals compatible with the problem constraints.

Now, fix $k \in \mathbb{N}$. For any $\alpha \in \mathbb{N}$ such that $k \leq k_\alpha$, there exist $\{\mu_i^{k,\alpha}\}_{i\notin I}$, $\{\lambda_j^{k,\alpha}\}_j$ satisfying the first five lines of Eqs. (4.35) and such that

$$\sum_{i\notin I} \mu_i^{k,\alpha}(\nabla_x g_i(p)) + \sum_j \lambda_j^{k,\alpha}(\nabla_x h_j(p)) = \sigma^\star(\nabla_x f(p)) - \sum_{i\in I} \mu_i^{k_\alpha}(\nabla_x g_i(p)), \tag{4.48}$$

for all $p \in \mathbb{C}\langle x\rangle$ with appropriately constrained degree. Indeed, $\{\mu_i^{k,\alpha}\}_{i\notin I}$, $\{\lambda_j^{k,\alpha}\}_j$ can be chosen to be the restrictions to $\mathbb{C}\langle x\rangle_{2k}$ of $\{\mu_i^{k_\alpha}\}_{i\notin I}$, $\{\lambda_j^{k_\alpha}\}_j$. The variables $\{\mu_i^{k,\alpha} : i \notin I\}, \{\lambda_j^{k,\alpha}\}_j$ in the equation above, identified by their evaluations on a finite set of polynomials, can thus be seen as a solution of the real system of linear equations

$$\begin{aligned}
&\mu_l^k(s^+h_j s^- + (s^-)^* h_j (s^+)^*) = \mu_l^k(i(s^+h_j s^- - (s^-)^* h_j (s^+)^*)) = 0,\\
&\qquad \forall s^+, s^- \in \langle x\rangle,\ \deg(s^+) + \deg(s^-) \leq 2k - \deg(h_j),\ \forall j, \forall l \notin I\\
&\mu_l^k(sg_l + g_l s^*) = \mu_l^k(i(sg_l - g_l s^*)) = 0, \quad \forall s \in \langle x\rangle,\ \deg(s) \leq 2k - \deg(g_l),\ l \notin I,\\
&\lambda_l^k(s^+h_j s^- + (s^-)^* h_j (s^+)^*) = \lambda_l^k(i(s^+h_j s^- - (s^-)^* h_j (s^+)^*)) = 0,\\
&\qquad \forall s^+, s^- \in \mathcal{M},\ \deg(s^+) + \deg(s^-) \leq 2k - \deg(h_j),\ \forall j, l,\\
&\sum_{i\notin I} \mu_i^k(\nabla_x g_i(p)) + \sum_j \lambda_j^k(\nabla_x h_j(p)) = c^\alpha(p),\\
&\qquad \forall p \in \mathcal{M}^n, \deg\left(\nabla_x f\right), \deg\left(\nabla_x g_i\right), \deg\left(\nabla_x h_j\right) \leq 2k - \deg(p),
\end{aligned} \tag{4.49}$$

The "constant vector" $c^\alpha(p)$ is given by

$$c^\alpha(p) := \sigma^\star(\nabla_x f(p)) - \sum_{i\in I} \mu_i^{k_\alpha}(\nabla_x g_i(p)).$$

Since the system (4.49) is solvable for all α, then it is also solvable in the limit $\alpha \to \infty$ (because the image of the corresponding matrix of coefficients, being finite-dimensional, is a closed subspace). By Lemma 4.7, there exists $K \in \mathbb{R}^+$ such that, for any α, there exists a solution $\{\mu_i^k : i \notin I\}, \{\lambda_j^k\}_j$ of the system

$$\mu_l^k(s^+h_j s^- + (s^-)^* h_j (s^+)^*) = \mu_l^k(i(s^+h_j s^- - (s^-)^* h_j (s^+)^*)) = 0,$$
$$\forall s^+, s^- \in \langle x\rangle,\ \deg(s^+) + \deg(s^-) \leq 2k - \deg(h_j),\ \forall j, \forall l \notin I \tag{4.50}$$
$$\mu_l^k(sg_l + g_l s^*) = \mu_l^k(i(sg_l - g_l s^*)) = 0, \quad \forall s \in \langle x\rangle,\ \deg(s) \leq 2k - \deg(g_l),\ l \notin I, \tag{4.51}$$
$$\lambda_l^k(s^+h_j s^- + (s^-)^* h_j (s^+)^*) = \lambda_l^k(i(s^+h_j s^- - (s^-)^* h_j (s^+)^*)) = 0,$$
$$\forall s^+, s^- \in \mathcal{M},\ \deg(s^+) + \deg(s^-) \leq 2k - \deg(h_j), \forall j, l,$$
$$\sum_{i\notin I} \mu_i^k(\nabla_x g_i(p)) + \sum_j \lambda_j^k(\nabla_x h_j(p)) = \lim_{\alpha\to\infty} c^\alpha(p) = \sigma^\star(\nabla_x f(p)) - \sum_{i\in I} \mu_i(\nabla_x g_i(p)),$$
$$\forall p \in \mathcal{M}^n, \deg\left(\nabla_x f\right), \deg\left(\nabla_x g_i\right), \deg\left(\nabla_x h_j\right) \leq 2k - \deg(p), \tag{4.52}$$

with

$$\sum_{p\in\mathcal{M}_{2k}} \left(\sum_{i\notin I} (\mu_i^k(p) - \mu_i^{k,\alpha}(p))^2 + \sum_j (\lambda_j^k(p) - \lambda_j^{k,\alpha}(p))^2 \right) \leq K^2 \|c - c^\alpha\|_2^2. \tag{4.53}$$

Thus, there exists a sequence of real positive numbers $(\delta_\alpha)_\alpha$, with $\lim_{\alpha\to\infty}\delta_\alpha = 0$, such that

$$\|M^k(\mu_i^k) - M^k(\mu_i^{k,\alpha})\|, \|M_l^k(\mu_i^k) - M_l^k(\mu_i^{k,\alpha})\| \leq \delta_\alpha, \forall i \notin I. \tag{4.54}$$

Since, by construction,

$$M^k(\mu_i^{k,\alpha}) + \frac{1}{k_\alpha}\mathbb{I} \geq 0, M_l^k(\mu_i^{k,\alpha}) + \frac{1}{k_\alpha}\mathbb{I} \geq 0, \tag{4.55}$$

we have that

$$M^k(\mu_i^k) + \left(\frac{1}{k_\alpha} + \delta_\alpha\right)\mathbb{I} \geq 0, M_l^k(\mu_i^k) + \left(\frac{1}{k_\alpha} + \delta_\alpha\right)\mathbb{I} \geq 0, \forall i \notin I. \tag{4.56}$$

The quantity $\frac{1}{k_\alpha} + \delta_\alpha$ tends to zero as α tends to infinity; it follows that, for any $\epsilon \in \mathbb{R}^+$ and for all $k \in \mathbb{N}$, there exist functionals $\{\mu_i^k : i \notin I\}, \{\lambda_j^k\}_j$ such that Eq. (4.35) is satisfied for $\epsilon_i = \epsilon$, for all $i \notin I$, with $\{\mu_i\}_{i\in I}$ being positive linear functionals compatible with the problem constraints. This implies that Eq. (4.35) is satisfied for any $\{\epsilon_i\}_{i\notin I} \subset \mathbb{R}^+$, as long as $\epsilon_i \geq \epsilon$ for all $i \notin I$. Since $\epsilon \in \mathbb{R}^+$ is arbitrary, the theorem has been proven. □

4.2. Normed KKT conditions

In this section, we will investigate tractable conditions under which the normed ncKKT conditions from Definition 3.15 hold. To find them, we will adapt known sufficient criteria for the classical case.

In Section 4.2.2, we present an algebraic, non-commutative version of the Mangasarian–Fromovitz Constraint Qualification (MFCQ), see Eqs. (2.3), (2.4) and prove that it suffices to guarantee that the corresponding NPO problem satisfies the normed ncKKT conditions. Before we proceed, though, we need to formulate an algebraic notion of gradient linear independence for sets of non-commutative polynomial equality constraints. This is the subject of the next section.

4.2.1. Linearly independent gradients

Given a number of non-commutative polynomial equality constraints $\{h_j(x) = 0\}$, we need to find a meaning for the expression "their gradients are linearly independent". Our starting point is the classical meaning of the term.

Let $\{x_i\}_i$ be commuting variables. Then the gradient vectors $\{\partial_x h_j\}_j$ are independent iff there exist vectors $v_1(x), \ldots, v_{m'}(x)$ such that

$$\langle \partial_x h_j | v_k(x)\rangle = \delta_{j,k}, \quad j,k = 1,\ldots,m'. \tag{4.57}$$

Now, define the matrices

$$\begin{aligned} \hat{P}_j(x) &:= |v_j(x)\rangle\langle\partial_x h_j|, \quad j = 1,\ldots,m', \\ \hat{P}_0(x) &:= \mathbb{I} - \sum_{k=1}^{m'} \hat{P}_k(x). \end{aligned} \tag{4.58}$$

It is easy to see that, for all $z \in \mathbb{R}^n$, these matrices satisfy

$$\begin{aligned} &\sum_k \hat{P}_k(x)|z\rangle = |z\rangle, \\ &\langle \partial_x h_j | \hat{P}_k(x)|z\rangle = 0, \quad \forall j \neq k, \\ &\langle \partial_x h_j | z\rangle = 0 \;\rightarrow\; \hat{P}_j(x)|z\rangle = 0. \end{aligned} \tag{4.59}$$

Note that, in general classical polynomial optimization problems, the matrices $\{\hat{P}_j(x)\}_j$ are rational functions of x. In the following lines, though, we will assume that they are actually polynomials. This algebraic constraint will restrict our definition of gradient linear independence more than usual, but otherwise will allow us to generalize this notion to non-commutative systems.

In classical systems, variables form a vector $x = (x_1, \ldots, x_n)$ of scalars, and the gradient $\partial_x h$ of a function h is also an n-dimensional vector of scalars. To find out how h will change if we move the variables in some direction z, we compute the scalar product $\partial_x h \cdot z$, thus obtaining a scalar.

In non-commutative systems, variables form a vector $x = (x_1, \ldots, x_n)$ of non-commuting objects, and the gradient $\nabla_x h(\bullet)$ of a polynomial $h(x)$ can be regarded as a linear map from n-tuples of polynomials $p = (p_1, \ldots, p_n) \in \mathbb{C}\langle x\rangle^n$ to a single polynomial $\nabla_x h(p(x))$. A non-commutative analog of relations (4.59) would thus demand the existence of $m'+1$ n-tuples of polynomials $P_0(x,z), \ldots, P_{m'}(x,z)$ in the variables $x = (x_1, \ldots, x_n)$, $z = (z_1, \ldots, z_n)$, linear on z. Each such n-tuple of polynomials $P_k(x,z)$ would play the role that the vector $\hat{P}_k(x)|z\rangle$ played in relations (4.59). Correspondingly, the tuples $P_0, \ldots, P_{m'}$ should satisfy the following conditions:

$$\sum_k P_k(X,Z) = Z, \tag{4.60a}$$

$$\nabla_x h_j(P_k(X,Z)) = 0, \quad \forall j \neq k, \tag{4.60b}$$

$$\nabla_x h_j(Z) = 0 \;\rightarrow\; P_j(X,Z) = 0. \tag{4.60c}$$

We will regard the algebraic version of constraints (4.60a)–(4.60c) as the non-commutative generalization of gradient linear independence.

Definition 4.8. *We say that a set of equality constraints $\{h_j(x) = 0 : j = 1, \ldots, m'\}$ has linearly independent gradients if there exist n-tuples of symmetric polynomials in $2n$ variables $P_0(x,z), P_1(x,z), \ldots, P_{m'}(x,z)$, linear in the z variables, such that*

$$\sum_{j=0}^{m'} P_j(x,p) - p \in I(\mathbf{h})^n, \quad \forall p \in \mathbb{C}\langle x\rangle^n, \tag{4.61a}$$

$$\nabla_x h_j(P_k(x,p)) \in I(\mathbf{h}), \quad \forall k \neq j,\ p \in \mathbb{C}\langle x\rangle^n, \tag{4.61b}$$

$$\exists \beta^+, \beta^- : (P_j(x,p))_k - \sum_l \left(\beta^+_{jkl}(x)\nabla_x h_j(p)(\beta^-_{jkl}(x))^* + \beta^-_{jkl}(x)\nabla_x h_j(p)(\beta^+_{jkl}(x))^*\right) \in I(\mathbf{h}),$$

$$j = 1, \ldots, m', \quad k = 1, \ldots, n, \quad p \in \mathbb{C}\langle x\rangle^n. \tag{4.61c}$$

Here $(P)_k$ denotes the k^{th} component of the n-tuple P.

Remark 4.9. One can replace Eq. (4.61c) by the weaker[5], problem-dependent constraints

$\exists \beta^+, \beta^-, \gamma^+, \gamma^-, s,$ such that

$$\nabla_x g_i\,(P_j(x,p)) - \sum_l \left(\beta^+_{ijl}(x)\nabla_x h_j(p)(\beta^-_{ijl}(x))^* + \beta^-_{ijl}(x)\nabla_x h_j(p)(\beta^+_{ijl}(x))^*\right) + s_i(p)g_i + g_i s_i(p) \in I(\mathbf{h}),$$

$$\nabla_x f\,(P_j(x,p)) - \sum_l \left(\gamma^+_{jl}(x)\nabla_x h_j(p)(\gamma^-_{jl}(x))^* + \gamma^-_{jl}(x)\nabla_x h_j(p)(\gamma^+_{jl}(x))^*\right) \in I(\mathbf{h}),$$

$$j = 1, \ldots, m', \quad k = 1, \ldots, n, \quad p \in \mathbb{C}\langle x\rangle^n. \tag{4.62}$$

Indeed, the proofs of Lemmas 4.13, 4.18 and Theorem 4.17 below follow through with such a modified definition.

[5] That eq. (4.61c) implies eq. (4.62) can be seen by substituting $P_j(x,p)$ by minus the second term of eq. (4.61c) in $\nabla_x g_i(P_j(x,p)), \nabla_x f(P_j(x,p))$. Due to the linearity of the gradient, the result will be an expression like minus the second terms of both lines of eq. (4.62), plus some polynomial in $I(\mathbf{h})$.

Unless the quotient space $\mathbb{C}\langle x\rangle/I(\mathbf{h})$ is finite-dimensional, verifying that conditions (4.61) hold for all $p \in \mathbb{C}\langle x\rangle^n$ is a challenging endeavor. The next proposition provides a practical, sufficient condition to guarantee gradient linear independence.

Proposition 4.10. *Let $\{h_j\}_j$ be a set of Hermitian polynomials, and let $\{r_k\}_k \subset \mathbb{C}\langle x\rangle$ be such that $[r_k, x_i] \in I(\mathbf{h})$ for all i, k. Suppose that there exist n-tuples of symmetric polynomials in $2n$ variables $P_0(x,z), P_1(x,z), \ldots, P_{m'}(x,z)$, linear in the z variables, such that*

$$\sum_{j=0}^{m'} P_j(x,z) - z \in (\mathcal{J}^Z)^n, \tag{4.63a}$$

$$\nabla_x h_j(P_k(x,z)) \in \mathcal{J}^Z, \quad \forall k \neq j, \tag{4.63b}$$

$$\exists \beta^+, \beta^- : (P_j(x,z))_k - \sum_l \beta^+_{jkl}(x)\nabla_x h_j(z)(\beta^-_{jkl}(x))^* - \beta^-_{jkl}(x)\nabla_x h_j(z)(\beta^+_{jkl}(x))^* \in \mathcal{J}^Z,$$
$$j = 1, \ldots, m', \quad k = 1, \ldots, n, \tag{4.63c}$$

where $\mathcal{J}^Z$ is the set of polynomials $q(x,z)$, linear in $z = (z_1, \ldots, z_n)$, of the form

$$q(x,z) = \sum_{j,l} p^+_{jl}(x,z) h_j(x) p^-_{jl}(x,z) + \sum_{k,l,i} q^+_{ikl}(x)[r_k(x), z_i] q^-_{ikl}(x). \tag{4.64}$$

Then, the set of equality constraints $\{h_j(x) = 0\}_j$ has linearly independent gradients.

The proof is obvious.

Example 4.11. Let $\{h_j(x) = 0\}_{j=1}^{m'}$ be of the form

$$h_j(x) = \sum_k \nu_{jk} x_k + b_j, \tag{4.65}$$

with $\{\nu_{jk}\}_{j,k} \cup \{b_j\} \subset \mathbb{R}$. If the matrix ν has linearly independent rows, then the gradients of $\{h_j\}_j$ are linearly independent. Indeed, let ν have rank m'. Then, there exist vectors $\{v^j\}_j \subset \mathbb{R}^n$ such that $\sum_k v^j_k \cdot \nu_{lk} = \delta_{jl}$. Define thus

$$\begin{aligned} (P_j(x,z))_k &:= v^j_k \sum_l \nu_{jl} z_l, \quad j = 1, \ldots, m', \\ P_0(x,z) &:= z - \sum_j P_j(x,z). \end{aligned} \tag{4.66}$$

The newly defined $\{P_j\}_j$ satisfy Eqs. (4.63). That is: for linear constraints, the commutative and non-commutative definitions of gradient linear independence coincide.

Example 4.12. Consider the equality constraints $\{h_j(x) = 0\}_{j=1}^n$ for

$$h_j(x) := x_j^2 - 1, \quad j = 1, \ldots, n. \tag{4.67}$$

This system also satisfies gradient linear independence. Indeed, define

$$\begin{aligned} (P_j(x,z))_k &:= \delta_{jk} \frac{1}{2}(z_j + x_j z_j x_j), \\ (P_0(x,z))_k &:= \frac{1}{2}(z_k - x_k z_k x_k). \end{aligned} \tag{4.68}$$

Then it can be easily verified that the n-tuples of polynomials $P_0, \ldots, P_n$ satisfy conditions (4.63).

The linear independence constraints are very restrictive: if a relaxed form of the weak ncKKT conditions holds, then so does normed ncKKT. This is proved in the following lemma.

Lemma 4.13. *Let the constraints $\{g_i(x) \geq 0\}_i \cup \{h_j(x) = 0\}_j$ be Archimedean, and let there exist bounded positive functionals σ, $\{\mu_i\}_i$ satisfying $\sigma(M(\mathbf{g}) + I(\mathbf{h})), \mu_i(M(\mathbf{g}) + I(\mathbf{h})) \geq 0$, for all i, with $\sigma(1)$, $\mu_i(1) \leq K$, for some $K \in \mathbb{R}^+$, $K > 1$ and such that, for all $k \in \mathbb{N}$, the system of linear equations*

$$\begin{aligned} &\lambda_j^k(I(\mathbf{h})_{2k}) = 0, \quad j = 1, \ldots, m', \\ &\sum_j \lambda_j^k(\nabla_x h_j(p(x))) = \sigma(\nabla_x f(p(x))) - \sum_i \mu_i(\nabla_x g_i(p(x)), \quad \forall p \in \mathbb{C}\langle x\rangle^n, \\ &\qquad \deg(t) \leq 2k + 1 - \deg(f) + 1,\ 2k + 1 - \deg(g_i), \quad i = 1, \ldots, m, \end{aligned} \tag{4.69}$$

has a solution λ^k. Furthermore, let the gradients of the constraints $\{h_j = 0\}_j$ be linearly independent. Then, the normed ncKKT conditions hold, with $\|\lambda^{\pm}\|_{\mathrm{SOS}} \leq O(K)$.

Proof. By Eq. (4.61b), if we set $p = P_0(x, p')$ in Eq. (4.69), we arrive at

$$\sigma(\nabla_x f(P_0(x, p'))) - \sum_i \mu_i(\nabla_x g_i(P_0(x, p'))) = 0, \tag{4.70}$$

provided that p' has sufficiently small degree (so that all relevant intermediate polynomials have degree equal to or below $2k$). Taking the limit $k \to \infty$, it follows that the relation above holds for all $p' \in \mathbb{C}\langle x\rangle$.

Now, consider the linear functional $\lambda_j : \mathbb{C}\langle x\rangle \to \mathbb{C}$, defined by

$$\begin{aligned} \lambda_j(p) := \sigma &\left(\nabla_x f \left(\left(\sum_l \beta_{jkl}^+ p(\beta_{jkl}^-)^* + \beta_{jkl}^- p(\beta_{jkl}^+)^* \right)_k \right) \right) \\ &- \sum_i \mu_i \left(\nabla_x g_i \left(\left(\sum_l \beta_{jkl}^+ p(\beta_{jkl}^-)^* + \beta_{jkl}^- p(\beta_{jkl}^+)^* \right)_k \right) \right), \end{aligned} \tag{4.71}$$

where the polynomials $\{\beta_{jkl}^{\pm}\}_{jkl}$ are the ones appearing in Eq. (4.61c).

Note that, for $r, t \in \mathbb{C}\langle x\rangle$ and any positive functional ω, the functional

$$\lambda(p) := \omega(rpt^* + tpr^*) \tag{4.72}$$

equals the difference between two positive functionals, namely,

$$\lambda(p) = \omega^+(p) - \omega^-(p), \tag{4.73}$$

for

$$\omega^{\pm}(p) := \frac{1}{2}\omega((r \pm t)p(r \pm t)^*). \tag{4.74}$$

Now, we expand the arguments of the functionals on the right-hand side of (4.71), as polynomials of the form $\sum_i s_i p t_i^* + t_i p s_i^*$. Since σ, $\{\mu_i\}$ are bounded positive functionals, we find that the right-hand side of Eq. (4.71) can be decomposed as a finite sum of differences of positive functionals. Thus,

$$\lambda_j = \lambda_j^+ - \lambda_j^-, \tag{4.75}$$

where the positive functionals $\lambda_j^{\pm}$ inherit from σ, $\{\mu_i\}_i$ the property of being non-negative on $M(\mathbf{g}) + I(\mathbf{h})$. The values $\lambda_j^{\pm}(1)$ correspond to expressions of the form

$$\lambda_j^{\pm}(1) = \sigma(s^{\pm}) + \sum_i \mu_i(s_i^{\pm}), \tag{4.76}$$

where $s^{\pm}, s_i^{\pm}$ are SOS. By the Archimedean condition, there exists $\nu \in \mathbb{R}^+$ such that the polynomials $\{\nu - s^{\pm}\} \cup \{\nu - s_i^{\pm}\}_i$ are SOS. Together with the constraints $\sigma(1), \mu_i(1) \leq K$, it follows that $\lambda^{\pm}(1) \leq K(m+1)\nu$.

It remains to be seen that the newly defined λ's satisfy condition (3.26c). Taking $p = P_0(x, p')$ in (3.26c) and invoking Eq. (4.61b) and Eq. (4.70) we have that Eq. (3.26c) is satisfied as long as p is of the form $p = P_0(x, p')$.

Next, set $p = P_j(x, p')$ in Eq. (3.26c), for some $j \in \{1, \ldots, m'\}$, $p' \in \mathbb{C}\langle x \rangle^n$. By Eq. (4.61b), all terms of the form $\lambda_k(\nabla_x h_k(P_j(x, p')))$ with $k \neq j$, drop. Furthermore, by Eqs. (4.61a), (4.61b), we have that

$$\lambda_j\left(\nabla_x h_j(p')\right) = \sum_k \lambda_j\left(\nabla_x h_j(P_k(x, p'))\right) = \lambda_j\left(\nabla_x h_j(P_j(x, p'))\right). \tag{4.77}$$

It follows that

$$\begin{aligned}
\lambda_j(\nabla_x h_j(P_j(x, p'))) &= \lambda_j(\nabla_x h_j(p')) \\
&= \sigma\left(\nabla_x f\left(\sum_l \beta_{jkl}^+ \nabla_x h_j(p') \beta_{jkl}^-\right)\right) - \sum_i \mu_i\left(\nabla_x g_i\left(\sum_l \beta_{jkl}^+ \nabla_x h_j(p') \beta_{jkl}^-\right)\right) \\
&= \sigma(\nabla_x f(P_j(x, p'))) - \sum_i \mu_i(\nabla_x g_i(P_j(x, p'))),
\end{aligned} \tag{4.78}$$

where the last equality is a consequence of Eq. (4.61c). Hence, Eq. (3.26c) holds for $p = P_j(x, p')$, for $j = 0, \ldots, m'$.

By Eq. (4.61a), any tuple of polynomials p' can be expressed as a sum of terms of the form $\{P_k(x, p')\}_k$ (modulo elements of $I(\mathbf{h})$), and so Eq. (3.26c) holds in general. □

4.2.2. Non-commutative Mangasarian–Fromovitz Constraint Qualification

We are ready to present a non-commutative, algebraic version of the Mangasarian–Fromovitz constraint qualification.

Definition 4.14 (ncMFCQ)**.** *Consider an NPO Problem* (2.13)*. We say that the problem satisfies* non-commutative Mangasarian–Fromovitz constraint qualification (ncMFCQ) *if, on one hand, the equality constraints have linearly independent gradients and, on the other hand, for* $i = 1, \ldots, m$*, there exist* $r \in \mathbb{R}^+$ *and a tuple of Hermitian polynomials* $q \in \mathbb{C}\langle x \rangle^n$ *such that*

$$\nabla_x g_i(q) + s_i g_i + g_i s_i^* - r \in M(\mathbf{g}) + I(\mathbf{h}), \quad i = 1, \ldots, m, \tag{4.79a}$$

$$\nabla_x h_j(q) \in I(\mathbf{h}), \quad j = 1, \ldots, m', \tag{4.79b}$$

for some polynomials $\{s_i : i = 1, \ldots, m\} \subset \mathbb{C}\langle x \rangle$*.*

Remark 4.15. Note the similarities between ncMFCQ and its classical analog MFCQ (Eqs. (2.3), (2.4)). For starters, both qualifications demand linear independence of the gradients of the equality constraints. In addition, Eq. (4.79b) is clearly the algebraic analog of the second line of Eq. (2.4). It remains to justify Eq. (4.79a). As we argue in Appendix A, the active constraints associated to the inequality $g_i(x) \geq 0$ are $\{\omega(g_i(X)) \geq 0 : \omega \geq 0, \omega(g_i(X^\star)) = 0\}$. Now, let ω be a state satisfying $\omega(g_i(X^\star)) = 0$ and let Eq. (4.79a) hold. Then, we have that

$$\frac{d}{dt}\omega(g_i(X^\star + tq(X^\star)))\Big|_{t=0} = \omega(\nabla_x g_i(q(X^\star))) \geq \omega(-s_i g_i - g_i s_i^* + r) = r > 0. \tag{4.80}$$

In other words, small variations of the optimal operators $X^\star$ in the $q(X^\star)$ direction are positive. Condition (4.79a) is therefore the algebraic, non-commutative analog of the first line in (2.4). Notice that, contrarily to the literature on classical polynomial optimization, we demand the conditions to hold for all feasible points, not just the optimal ones.

Example 4.16. Let $x = (y_1, \ldots, y_c, z_1, \ldots, z_d)$, and let $\{h_j(y) = 0\}_j$ be a set of equality constraints with linearly independent gradients. Let the remaining constraints be inequalities of the form:

$$g_i(x) = 1 + \tilde{g}_i(x) \geq 0, i = 1, \ldots, m, \tag{4.81}$$

where $\tilde{g}_i(x)$ is homogeneous in $z = (z_1, \ldots, z_d)$ with degree $r_i > 0$.

The constraints $\{h_j(y) = 0\}_j \cup \{g_i(x) \geq 0\}_i$ satisfy ncMFCQ. Indeed, choose $(q(x))_k = 0$, for $k = 1, \ldots, c$, and $(q(x))_{k+c} = -z_k$, for $k = 1, \ldots, d$. Then, on one hand, $\nabla_x h_j(q) = 0$, for all j, i.e., Eq. (4.79b) holds. On the other hand, $\nabla_x g_i(x)(q) = -r_i \tilde{g}_i(x)$, and so

$$r_i g_i(x) + \nabla_x g_i(x)(q) - r_i = 0. \tag{4.82}$$

That is, Eq. (4.79a) holds for $r := \min_i r_i$. Since, by assumption, the equality constraints have linearly independent gradients, the three conditions defining ncMFCQ are satisfied.

The next theorem is the non-commutative generalization of a celebrated result in classical optimization, which states the validity of the KKT conditions under Mangasarian–Fromovitz constraint qualification [20].

Theorem 4.17. *Consider an NPO Problem (2.13) that satisfies both the Archimedean condition and ncMFCQ. Then, the problem satisfies the normed ncKKT conditions.*

Proof. Eq. (4.79a) and (4.79b) represent the conditions of Theorem 4.5 for $I = \{1, \ldots, m\}$. Thus, there exist positive linear functionals $\{\mu_i\}_{i=1}^m$, compatible with the constraints of the problem $\{g_i(x) \geq 0\}_i \cup \{h_j(x) = 0\}_j$, such that, for all k, the system of linear equations

$$\begin{aligned} &\lambda_j^k(s^+ h_k s^-) = 0, \quad j, k = 1, \ldots, m', \ \deg(s^+ h_k s^-) \leq 2k, \\ &\sum_j \lambda_j^k(\nabla_x h_j(p(x))) = \sigma^\star(\nabla_x f(p(x))) - \sum_i \mu_i(\nabla_x g_i(p(x)), \quad \forall p \in \mathbb{C}\langle x\rangle^n, \end{aligned} \tag{4.83}$$

where the degree of p is appropriately bounded, has a solution λ^k. The conditions of Lemma 4.13 are, therefore, met. Hence, the problem satisfies the normed ncKKT conditions. □

Like its classical counterpart, ncMFCQ allows one to bound the norm of the Lagrange multipliers $\{\mu_i\}_i$, $\{\lambda_j\}_j$. This implies that each of the SDP relaxations of the normed KKT conditions corresponds to a bounded optimization problem.

Lemma 4.18. *Assume that Problem (2.13) is Archimedean and satisfies ncMFCQ (and thus normed ncKKT). Then we can, without loss of generality, bound the state multipliers $\{\mu_i\}_i$ and $\{\lambda_j^\pm\}_j$ in Eq. (3.26). That is, we can find $K \in \mathbb{R}^+$ such that*

$$\mu_i(1), \lambda_j^\pm(1) \leq K, \quad \forall i, j. \tag{4.84}$$

Proof. By Remark 4.6, we have that $\sum_i \mu_i(1) \leq \frac{1}{r}\sigma^\star(\nabla_x f(q))$. In turn, provided that the original NPO problem is Archimedean, there exists $\eta \in \mathbb{R}^+$ such that

$$\eta - \nabla_x f(q) \tag{4.85}$$

admits an SOS decomposition. Hence, for each i, $\mu_i(1) \leq \frac{\eta}{r}$.

The set of linear functionals $\{\lambda_j\}_j$, restricted to the set of polynomials of degree $2k$ or smaller, is a solution of Eq. (4.69). The conditions of Lemma 4.13 are met, and thus there exist bounded linear functionals $\{\tilde{\lambda}_j^\pm : \mathbb{C}\langle x\rangle \to \mathbb{C}\}_j$, with norms bounded by $O(\frac{\eta}{r})$, compatible with the problem constraints and satisfying Eq. (3.26c). □

4.3. Strong KKT conditions

It is observed in practice that many natural NPO problems admit an exact SOS resolution, see, e.g., [12]. Namely, for some $k \in \mathbb{N}$, the k^{th} level of the hierarchy of SDP relaxations (2.17) has a (feasible) maximizer achieving the exact solution of the problem. The next theorem shows that, in any such scenario, strong ncKKT holds.

Theorem 4.19. *Consider the NPO Problem (2.5), and let $f - p^\star \in M(\mathbf{g}) + I(\mathbf{h})$. Then, Problem (2.5) satisfies strong ncKKT.*

Proof. Let $(\mathcal{H}^\star, X^\star, \psi^\star)$ be a minimizer of Problem (2.5). For simplicity, in the following we use its abstract, functional form $\sigma^\star : \mathbb{C}\langle x\rangle \to \mathbb{C}$, namely:

$$\sigma^\star(p) = \psi^\star(p(X^\star)). \tag{4.86}$$

By the hypotheses of the theorem, there exist polynomials $s_l, s_{il}, s_{jl}^+, s_{jl}^-$ such that

$$f - p^\star = \sum_l s_l s_l^* + \sum_{i,l} s_{il} g_i s_{il}^* + \sum_{j,l} s_{jl}^+ h_j (s_{jl}^-)^* + s_{jl}^- h_j (s_{jl}^+)^*. \tag{4.87}$$

In addition,

$$\sigma^\star(f - p^\star) = 0. \tag{4.88}$$

It follows that

$$\begin{aligned} \sigma^\star(s_l s_l^*) &= 0, \quad \forall l, \\ \sigma^\star(s_{il} g_i s_{il}^*) &= 0, \quad \forall i, l. \end{aligned} \tag{4.89}$$

These relations, in turn, imply that, for any $q \in \mathbb{C}\langle x\rangle$,

$$\sigma^\star(s_l q) = \sigma^\star(q s_l^*) = 0, \quad \forall l, \tag{4.90a}$$

$$\sigma^\star(s_{il} g_i q) = \sigma^\star(q g_i s_{il}^*) = 0, \quad \forall i, l. \tag{4.90b}$$

Indeed, the first relation follows from the Cauchy-Schwarz inequality or the positive semidefiniteness of the 2×2 matrix

$$\begin{pmatrix} \sigma^\star(qq^*) & \sigma^\star(qs_l^*) \\ \sigma^\star(s_l q^*) & \sigma^\star(s_l s_l^*) \end{pmatrix}. \tag{4.91}$$

The second one, from the positive semidefiniteness of

$$\begin{pmatrix} \sigma^\star(qg_i q^*) & \sigma^\star(qg_i s_{il}^*) \\ \sigma^\star(s_{il} g_i q^*) & \sigma^\star(s_{ij} g_i s_{ij}^*) \end{pmatrix}. \tag{4.92}$$

Now, for $\delta \in \mathbb{R}$ and an arbitrary vector of Hermitian polynomials $p = (p_i)_{i=1}^n$, let us define a new state through the relation $\sigma^\delta(a) := \sigma^\star(\pi^\delta(a))$, where $\pi^\delta : \mathbb{C}\langle x\rangle \to \mathbb{C}\langle x\rangle$ is the homomorphism given by $\pi^\delta(x_i) = x_i + \delta \cdot p_i(x)$. This linear functional σ^δ is indeed a state, since $\sigma^\delta(qq^*) \geq 0$ for all $q \in \mathbb{C}\langle x\rangle$. However, it does not necessarily satisfy feasibility conditions of the form $\sigma^\delta(qg_i q^*) \geq 0$, $\sigma^\delta(s^+ h_j s^-) = 0$.

We apply the state σ^δ on both sides of Eq. (4.87). Taking into account Eqs. (4.88), (4.90), and the chain rule of differentiation, the result is

$$\delta\sigma^\star(\nabla_x f(p(x))) + O(\delta^2) = \delta \sum_i \mu_i(\nabla_x(g_i(p(x))) + \delta \sum_j \lambda_j(\nabla_x(h_j(p(x))) + O(\delta^2), \tag{4.93}$$

where μ_i denotes the positive linear functional given by

$$\mu_i(q) := \sigma^\star(\sum_l s_{il} q s_{il}^*), \tag{4.94}$$

and λ_j is the linear functional

$$\lambda_j(q) := \sigma^\star(\sum_l s_{jl}^+ q (s_{jl}^-)^* + \sum_l s_{jl}^- q (s_{jl}^+)^*). \tag{4.95}$$

Note that λ_j can be expressed as the difference of two positive functionals $\lambda_j^\pm$, namely:

$$\lambda_j^\pm(q) := \frac{1}{2}\sum_l \sigma^\star((s_{jl}^+ \pm s_{jl}^-) q (s_{jl}^+ \pm s_{jl}^-)^*). \tag{4.96}$$

Collecting the terms in Eq. (4.93) that depend linearly on δ, we have that

$$\sigma^\star(\nabla_x f(p(x))) = \sum_i \mu_i(\nabla_x(g_i(p(x))) + \sum_j \lambda_j(\nabla_x(h_j(p(x))). \tag{4.97}$$

This is condition (3.25b).

Finally, the positive linear functionals $\{\mu_i\}_i$ satisfy complementary slackness (3.25a), for

$$\mu_i(g_i) = \sum_l \sigma^\star(s_{il} g_i s_{il}^*) = 0, \quad \forall i, \tag{4.98}$$

by Eq. (4.89).

It only rests to show that $\{\mu_i\}_i$ and $\{\lambda_j^\pm\}_j$ can be expressed as

$$\mu_i(p) = \tilde{\mu}_i(p(X^\star)),\ \lambda_j^\pm(p) = \tilde{\lambda}_j^\pm(p(X^\star)), \tag{4.99}$$

for some functionals $\{\tilde{\mu}_i : \mathcal{A}(X^\star) \to \mathbb{C}\}_i$, $\{\tilde{\lambda}_j^\pm : \mathcal{A}(X^\star) \to \mathbb{C}\}_j$. This last bit follows from Eq. (4.86) and the fact that both sets of functionals are defined in terms of $\sigma^\star$. □

The reader might question the practical use of Theorem 4.19. How can one know in advance that a given NPO problem will admit an exact SOS resolution? To answer this question, we need to examine the relation between positive, non-negative and SOS polynomials.

Given a set of constraints $\{g_i(x) \geq 0\}_i \cup \{h_j(x) = 0\}_j$, a Hermitian polynomial r is said to be positive (non-negative) if $r(X) > 0$ ($r(X) \geq 0$), for all tuples of operators X satisfying the problem constraints. Think of the polynomial $f - p^\star$, where $p^\star$ is the solution of (2.5). This polynomial is non-negative, but not positive.

As we pointed out in Section 2.2, if r is SOS, then it is also non-negative. The converse statement, however, does not hold: for some Archimedean constraints, there exist non-negative polynomials r that are not SOS. It is thus not a surprise that some instances of Problem (2.13) do not admit an SOS resolution.

However, some sets of non-commuting constraints $\{g_i(x) \geq 0\}_i \cup \{h_j(x) = 0\}_j$ have the property that any non-negative Hermitian polynomial is SOS. Such sets of constraints are said to generate an *Archimedean closed* quadratic module (or set of SOS polynomials) [53]. In the following, we provide two families of constraints that are known to generate Archimedean closed quadratic modules.

4.3.1. Equality constraints with a faithful finite-dimensional $*$-representation

Proposition 4.20. *Consider a set of equality constraints $\{h_j(x) = 0\}_j$ such that, for some finite-dimensional Hilbert space $\mathcal{H}^\star$, there exist a $*$-homomorphism $\pi : \mathbb{C}\langle x\rangle \to B(\mathcal{H}^\star)$, with $\ker(\pi) = I(\mathbf{h})$. Then, the quadratic module generated by $\{h_j(x) = 0\}_j$ is Archimedean closed.*

Proof. Define $X_k^\star := \pi(x_k)$, denote by $C^*(X^\star)$ the unital C^*-algebra generated by $X_1^\star, \ldots, X_n^\star$ and let $p \in \mathbb{C}\langle x\rangle$ be an arbitrary non-negative polynomial. Since $\pi(p) = p(X^\star) \in \mathcal{A}(X^\star)$ is a non-negative operator, then one can define its square root $p(X^\star)^{1/2}$. Due to the finite dimensionality of $\mathcal{A}(X^\star)$, there exists a polynomial $s \in \mathbb{C}\langle x\rangle$ such that $p(X^\star)^{1/2} = s(X^\star)$. We thus have that

$$p(X^\star) - s(X^\star)s(X^\star)^* = 0. \tag{4.100}$$

Hence, $\pi(p - ss^*) = 0$, and so $p - ss^* \in I(\mathbf{h})$. We have just shown that p is SOS. □

Remark 4.21. A set of constraints satisfying the conditions of Proposition 4.20 is the Pauli algebra (6.1), (6.2) used to model many-body quantum systems, see Section 6.1.

Corollary 4.22. *Let $\{h_j(x) = 0\}_j$ be a set of constraints satisfying the conditions of Proposition 4.20. Then, any problem of the form (2.13) with just equality constraints $\{h_j(x) = 0\}_j$ satisfies strong ncKKT.*

4.3.2. Convexity

As it turns out, any set of convex inequality constraints defines an Archimedean closed quadratic module. To make this statement precise, though, we need to recall the definition of convexity for non-commutative polynomials, over an algebra or in general.

Definition 4.23. *A Hermitian non-commutative polynomial p is convex in the C^*-algebra $\mathcal{A}$ if, for any two n-tuples of Hermitian operators $Y_1, Y_2 \in \mathcal{A}^n$, it holds that*

$$p\left(\delta Y_1 + (1-\delta)Y_2\right) \leq \delta p(Y_1) + (1-\delta)p(Y_2), \tag{4.101}$$

for all $\delta \in \mathbb{R}$, $0 \leq \delta \leq 1$. If p is convex for all C^*-algebras, then we call it matrix convex *[54].*

As shown in [54], $p \in \mathbb{C}\langle x\rangle$ is matrix convex iff its *Hessian* is *matrix positive*, i.e., if the polynomial

$$\frac{d^2 p(x+th)}{dt^2}\Big|_{t=0} \tag{4.102}$$

is a sum of squares of the Hermitian variables $x_1, \ldots, x_n, h_1, \ldots, h_n$. In [54] it is also proven that matrix convex polynomials have degree at most two.

Now we are ready to state a sufficient criterion for Archimedean closure.

Theorem 4.24. *Let the Archimedean Problem (2.13) be such that*

(a) *$\{g_i\}_i$ are matrix concave;*

(b) *the equality constraints $\{h_j(x) = 0\}_j$ are affine linear, i.e., of the form*

$$h_j(x) = \sum_k \nu_{jk} x_k - b_j, \tag{4.103}$$

and linearly independent.

In addition, let there exist $r \in \mathbb{R}^+$ *and an* n*-tuple* $q \in \mathbb{C}\langle x\rangle^n$ *of Hermitian polynomials such that*

$$g_i(q(x)) - r \tag{4.104}$$

is SOS, for $i = 1, \ldots, m$*, and*

$$h_j(q(x)) \in I(\mathbf{h}), \quad j = 1, \ldots, m'. \tag{4.105}$$

Then, the constraints $\{g_i(x) \geq 0\}_i \cup \{h_j(x) = 0\}_j$ *generate the Archimedean closed quadratic module* $M(\mathbf{g}) + I(\mathbf{h})$*. Thus, by Theorem 4.19, Problem (2.13) satisfies the strong ncKKT conditions.*

In the proof of the theorem we shall make use of Schur complements [55]:

Lemma 4.25. *The block matrix*

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{12}^* & I \end{bmatrix}$$

is positive (semi)definite iff $a_{11} - a_{12}a_{12}^*$ *is positive (semi)definite.*

Proof. This is a special case of [55, Theorem 1.12]. Alternately, observe that

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{12}^* & I \end{bmatrix} = \begin{bmatrix} I & a_{12} \\ 0 & I \end{bmatrix} \begin{bmatrix} a_{11} - a_{12}a_{12}^* & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} I & a_{12} \\ 0 & I \end{bmatrix}^*. \qquad \square$$

Proof of Theorem 4.24. Let $(\mathcal{H}^\star, \sigma^\star, X^\star)$ be a solution of Problem (2.5). By (4.104) and (4.105), the choice $\hat{X} = q(X^\star)$ satisfies the constraints

$$g_i(\hat{X}) > 0 \quad \text{for all } i, \qquad h_j(\hat{X}) = 0 \quad \text{for all } j. \tag{4.106}$$

We solve the system of linear equations $\{h_j = 0\}_j$. Without loss of generality, we express the last $n - s$ variables in terms of $x_1, \ldots, x_s$. By back substitution, our Problem (2.5) is thus equivalent to one with $m' = 0$, i.e., one without equality constraints. More precisely, if a polynomial p has a SOS decomposition (without h_j's) after this back substitution, then the original form of p has a SOS decomposition as in (2.11). It thus suffices to consider an Archimedean Problem (2.5) with $m' = 0$.

By the Helton-McCullough structure theorem for concave non-commutative polynomials [54, Corollary 7.1], each g_i is of the form

$$g_i(x) = c_i + \Lambda_{i0}(x) - \sum_{j=1}^{N} \Lambda_{ij}(x)^* \Lambda_{ij}(x) \tag{4.107}$$

for some $c_i \in \mathbb{R}$ and homogeneous linear $\Lambda_{ij}(x)$. (By Carathéodory's theorem on convex hulls in finite dimensions [56, Theorem I.2.3], the length N of the sum of squares in (4.107) can be chosen independently of i.) Such a g_i is the Schur complement of a linear pencil, namely,

$$G_i(x) = \begin{bmatrix} c_i + \Lambda_{i0}(x) & \Lambda_{i1}(x)^* & \cdots & \Lambda_{iN}(x)^* \\ \Lambda_{i1}(x) & 1 & & \\ \vdots & & \ddots & \\ \Lambda_{iN}(x) & & & 1 \end{bmatrix} \tag{4.108}$$

Form the large block diagonal pencil $L(x) := G_1(x) \oplus \cdots \oplus G_m(x)$ of size $m(N+1) \times m(N+1)$.

From $g_i(\hat{X}) > 0$ for all i, we deduce from Lemma 4.25 that $L(\hat{X}) > 0$. For a unit vector $\psi \in \mathcal{H}^\star$, consider $\hat{x} := \psi^* \hat{X} \psi \in \mathbb{R}^n$. Since L is linear and ψ is a unit vector, $L(\psi^* \hat{X} \psi) = (I_{m(N+1)} \otimes \psi)^* L(\hat{X})(I_{m(N+1)} \otimes \psi)$. Letting $0 \neq \eta \in \mathbb{C}^{m(N+1)}$, we have

$$\begin{aligned} \langle L(\hat{x})\eta|\eta\rangle &= \langle L(\psi^* \hat{X} \psi)\eta|\eta\rangle = \langle (I_{m(N+1)} \otimes \psi)^* L(\hat{X})(I_{m(N+1)} \otimes \psi)\eta|\eta\rangle \\ &= \langle L(\hat{X})(I_{m(N+1)} \otimes \psi)\eta|(I_{m(N+1)} \otimes \psi)\eta\rangle = \langle L(\hat{X})(\eta \otimes \psi)|(\eta \otimes \psi)\rangle > 0. \end{aligned} \tag{4.109}$$

Now again by Lemma 4.25, $g_i(\hat{x}) > 0$ for all i. By translating x by $-\hat{x}$, we may assume w.l.o.g. that $g_i(0) > 0$. By rescaling we further reduce to $g_i(0) = 1$, i.e., all g_i are monic. We are now in a position to apply the convex Positivstellensatz [57, Theorem 1.2] to deduce that $f - p^\star$ has a SOS decomposition. Thus, by Theorem 4.19, Problem (2.13) admits the strong ncKKT conditions. □

Corollary 4.26. *Let the Problem (2.13) be such that*

(a) $\{g_i\}_i$ *are matrix concave;*

(b) *the equality constraints* $\{h_j(x) = 0\}_j$ *are affine linear, and linearly independent.*

If there exists a feasible point X for Problem (2.13) such that $g_i(X) > 0$ for all i, then Problem (2.13) satisfies the strong ncKKT conditions.

Proof. Immediate from the proof of Theorem 4.24. □

Remark 4.27. The Convex Positivstellensatz of [57] has minimal degree. Enforcing the strong ncKKT conditions on an NPO problem with matrix convex constraints is therefore superfluous: the first level of relaxation (2.17) with degree high enough to encode the problem will already achieve convergence.

5. PARTIAL OPERATOR OPTIMALITY CONDITIONS

As we will see in Section 6.2, the computation of the maximum quantum value of a linear Bell functional only seems to satisfy essential ncKKT. In this particular problem, and some interesting variants thereof [43, 58, 59], the non-commuting variables $x = (x_1, \ldots, x_n)$ can be partitioned as $x = (y, z)$, and the only constraints relating the parts y and z are commutation relations of the form:

$$[z_k, y_l] = 0, \quad \forall k, l. \tag{5.1}$$

As it turns out, if the remaining constraints on z are convex and $f(y, z)$ is convex on z, then a partial form of strong ncKKT holds for the variables z. Similarly, if the remaining constraints on z satisfy ncMFCQ, then the variables z will satisfy a form of normed ncKKT. Both scenarios thus allow us to derive stronger SDP constraints than the mere essential ncKKT would permit, further boosting the speed of convergence of the vanilla SDP hierarchy [39].

In this section, we define these partial forms of operator optimality and find sufficient criteria to ensure that they hold. One of these criteria will be used in Section 6.2 to derive new SDP constraints for maximizations over Bell functionals.

5.1. Partial normed ncKKT conditions

Theorem 5.1. *Consider an Archimedean NPO (2.13) with variables $x = (y, z)$, with $z = (z_1, \ldots, z_q)$ such that*

(a) *The only constraints involving both types of variables y and z are the following:*

$$[y_r, z_s] = 0, \quad \forall r, s. \tag{5.2}$$

(b) *The remaining constraints involving variables z are*

$$\begin{aligned} &\hat{g}_i(z) \geq 0, \quad i = 1, \ldots, m_Z, \\ &\hat{h}_j(z) = 0, \quad j = 1, \ldots, m'_Z, \end{aligned} \tag{5.3}$$

and satisfy ncMFCQ.

Then, for any bounded solution $\sigma^\star$ of Problem (2.13), there exist Hermitian linear functionals $\{\mu_i : \mathbb{C}\langle z\rangle \to \mathbb{C}\}_i$, $\{\lambda_i^\pm : \mathbb{C}\langle z\rangle \to \mathbb{C}\}_j$, non-negative on $M(\{\hat{g}_i(z)\}_i) + I(\{\hat{h}_j(z)\}_j)$, such that

$$\mu_i(\hat{g}_i) = 0, \quad i = 1, \ldots, m_Z, \tag{5.4a}$$

$$\sigma^\star\left(\nabla_z f(p)\right) - \sum_i \mu_i\left(\nabla_z \hat{g}_i(p)\right) - \sum_j \lambda_j\left(\nabla_z \hat{h}_j(p)\right) = 0, \quad \forall p \in \mathbb{C}\langle z\rangle^q, \tag{5.4b}$$

with $\lambda_j = \lambda_j^+ - \lambda_j^-$, for $j = 1, \ldots, m_Z$.

Proof. Let $\sigma^\star$ be a solution of Problem (2.13). For fixed $k \in \mathbb{N}$, consider the following SDP:

$$\begin{aligned} &\min_{\epsilon,\mu,\lambda} \epsilon \\ &\text{s.t. } \epsilon \geq 0, \\ &\quad \mu_i^k(pp^*) + \epsilon\|p\|_2^2 \geq 0, \quad \forall p \in \mathbb{C}\langle z\rangle,\ \deg(p) \leq k,\ i = 1, ., , , m_Z, \\ &\quad \mu_i^k(p\hat{g}_l p^*) + \epsilon\|p\|_2^2 \geq 0, \quad \forall p \in \mathbb{C}\langle z\rangle,\ \deg(p) \leq k - \left\lceil \frac{\deg(\hat{g}_l)}{2} \right\rceil,\ i, l = 1, \ldots, m_Z, \\ &\quad \mu_i^k(s^+\hat{h}_j s^-) = 0, \quad \forall s^+, s^- \in \mathbb{C}\langle z\rangle,\ \deg(s^+) + \deg(s^-) \leq 2k - \deg(\hat{h}_j),\ j = 1, \ldots, m'_Z, \\ &\quad \mu_i^k(s\hat{g}_i) = \mu_i^k(\hat{g}_i s) = 0, \quad \forall s \in \mathbb{C}\langle z\rangle,\ \deg(s) \leq 2k - \deg(g_i), i = 1, \ldots, m, \\ &\quad \lambda_j^k(s^+\hat{h}_j s^-) = 0, \quad \forall s^+, s^- \in \mathbb{C}\langle z\rangle,\ \deg(s^+) + \deg(s^-) \leq 2k - \deg(h_j),\ j = 1, \ldots, m'_Z, \\ &\quad \sigma^\star\left(f'(p)\right) - \sum_i \mu_i^k\left(\nabla_z \hat{g}_i(p)\right) - \sum_j \lambda_j^k\left(\nabla_z h_j(p)\right) = 0, \quad \forall p \in \mathbb{C}\langle z\rangle^q, \\ &\qquad \deg_z\left(f'(p)\right),\ \deg\left(\nabla_z \hat{g}_i\right),\ \deg\left(\nabla_z \hat{h}_j\right) \leq 2k - \deg(p), \end{aligned} \tag{5.5}$$

where

$$f'(p) = \lim_{\delta \to 0} \frac{f(Y^\star, Z^\star + \delta p(Z^\star)) - f(Y^\star, Z^\star)}{\delta}, \tag{5.6}$$

and the expression $\deg_z(\bullet)$ denotes the degree of $\bullet$ with respect to the variables z.

Invoking Lemma 4.3 as in the proof of Theorem 4.1, we find that this SDP problem has a solution for every $\epsilon > 0$. Next, using relations (4.79a), (4.79b), like in the proof of Theorem 4.5, it is shown that Problem (5.5) is also feasible for $\epsilon = 0$. Moreover, we can replace the variables μ_i^k by bounded positive functionals $\{\mu_i\}_i$, compatible with the constraints $\{\hat{g}_i(z) \geq 0\}_i \cup \{\hat{h}_j(z) = 0\}_j$. Finally, we invoke the linear independence of the gradients of $\{\hat{h}_j\}_j$ as in Lemma 4.13 to prove that $\{\lambda_j^k\}_j$ can be replaced by linear functionals. □

5.2. Partial strong ncKKT conditions

Theorem 5.2. *Consider an NPO Problem (2.13), not necessarily Archimedean, with variables $x = (y, z)$, with $z = (z_1, \dots, z_q)$ such that*

(a) *The only constraints involving both types of variables y and z are the following:*

$$[y_r, z_s] = 0, \quad \forall r, s. \tag{5.7}$$

(b) *The remaining constraints involving variables of type z are*

$$\begin{aligned} &\hat{g}_i(z) \geq 0, \quad i = 1, \dots, m_Z, \\ &\hat{h}_j(z) = 0, \quad j = 1, \dots, m'_Z, \end{aligned} \tag{5.8}$$

where $\{\hat{g}_i\}_i$ are matrix concave non-commutative polynomials and $\{\hat{h}_j\}_j$ are affine linear polynomials with linearly independent gradients.

(c) *There exist $r \in \mathbb{R}^+$ and polynomials $Q = (Q_1, \dots, Q_q)$ such that*

$$\hat{g}_i(Q(z)) - r \tag{5.9}$$

is a weighted sum of squares, for $i = 1, \dots, m_Z$, and

$$\hat{h}_j(Q(z)) = \sum_{l,j'} s_{jj'l}(z) \hat{h}_{j'}(z) s'_{jj'l}(z), \tag{5.10}$$

for some polynomials $s_{jj'l}, s'_{jj'l}$, for $j = 1, \dots, m'_Z$.

(d) *Let $\{\tilde{g}_i(y) \geq 0\}_i \cup \{\tilde{h}_j(y) = 0\}_j$ be the remaining constraints on y. Then $f(y, z)$ satisfies*

$$\begin{aligned} \frac{d^2 f(y, z + th)}{dt^2}\Big|_{t=0} &= \sum_{l,k} \zeta^+_{lk}(y, z, h)[y_k, z_l]\zeta^-_{lk}(y, z, h) + \eta^+_{lk}(y, z, h)[y_k, h_l]\eta^-_{lk}(y, z, h) \\ &+ \sum_l \theta_l(z, h, y)\theta^*_l(z, h, y) + \sum_i \theta_{il}(z, h, y)\tilde{g}_i(y)\theta^*_{il}(z, h, y) + \sum_j \iota^+_{jl}(z, h, y)\tilde{h}_j(y)\iota^-_{jl}(z, h, y), \end{aligned} \tag{5.11}$$

for some polynomials $\zeta^\pm, \eta, \theta, \iota^\pm$.

Let $(\mathcal{H}^\star, X^\star, \sigma^\star)$ be any bounded solution of Problem (2.5), with $X^\star = (Y^\star, Z^\star)$, and denote by $C^*(Z^\star)$ the C^*-algebra generated by $Z^\star_1, \dots, Z^\star_q$. Then, there exist positive linear functionals $\{\mu_i : C^*(Z^\star) \to \mathbb{C}\}_i$ and bounded Hermitian linear functionals $\{\lambda_i : C^*(Z^\star) \to \mathbb{C}\}_j$ such that

$$\mu_i(\hat{g}_i(Z^\star)) = 0, \quad i = 1, \dots, m_Z, \tag{5.12a}$$

$$\sigma^\star\left(\nabla_z f(p)\right) - \sum_i \mu_i\left(\nabla_z \hat{g}_i(p)\right) - \sum_j \lambda_j\left(\nabla_z \hat{h}_j(p)\right) = 0, \quad \forall p \in C^*(Z^\star)^q. \tag{5.12b}$$

The following lemma is the key to arrive at this result.

Lemma 5.3. *Let $\mathcal{A}$ be a C^*-algebra, call $\mathcal{A}_h$ its set of Hermitian elements. Let $Z = (Z_1, \dots, Z_q)$ be a set of Hermitian operator variables and let $\tilde{f} : \mathcal{A}^q \to \mathbb{C}$ be a Hermitian convex function[6].*

[6] Namely, $f(\delta Z^1 + (1-\delta)Z^2) \leq \delta f(Z^1) + (1-\delta) f(Z^2)$, for all $\delta \in \mathbb{R}$, $0 \leq \delta \leq 1$, $Z^1, Z^2 \in \mathcal{A}^q_h$.

Given some concave *non-commutative polynomials* $\{\hat{g}_i\}$ *and linearly independent affine linear polynomials* $\{\hat{h}_i\}$*, consider the optimization problem*

$$\begin{aligned} \min_{Z\in\mathcal{A}^q}\ & \hat{f}(Z) \\ s.t.\ & \hat{g}_i(Z)\geq 0, \quad i=1,\dots,m_Z, \\ & \hat{h}_j(Z)=0, \quad j=1,\dots,m'_Z. \end{aligned} \tag{5.13}$$

Suppose that an optimal solution exists, call it $Z^\star$*.*

Further assume that Problem (5.13) *admits a strictly feasible point, i.e., there exists a feasible tuple* $\hat{Z}\in\mathcal{A}^q$ *such that*

$$\begin{aligned} &\hat{g}_i(\hat{Z})>0, \quad i=1,\dots,m_Z, \\ &\hat{h}_j(\hat{Z})=0, \quad j=1,\dots,m'_Z. \end{aligned} \tag{5.14}$$

Then, there exist positive linear functionals $\mu_i:\mathcal{A}\to\mathbb{C}$*,* $i=1,\dots,m_Z$*, and bounded Hermitian linear functionals* $\lambda_j:\mathcal{A}\to\mathbb{C}$*,* $j=1,\dots,m'$ *satisfying*

$$\mu_i(\hat{g}_i(Z^\star))=0, \quad i=1,\dots,m_Z, \tag{5.15}$$

such that $Z^\star$ *is a solution of the unconstrained optimization problem*

$$\min_{Z\in\mathcal{A}_h^q}\mathcal{L}(Z;\mu,\lambda), \tag{5.16}$$

with

$$\mathcal{L}(Z;\mu,\lambda):=\hat{f}(Z)-\sum_i\mu_i(\hat{g}_i(Z))-\sum_j\lambda_j(\hat{h}_j(Z)). \tag{5.17}$$

Proof. It suffices to follow the classical proof of the Slater criterion for strong duality (cf. [17, §4.2]). Given $Z^\star$, we define the sets:

$$\begin{aligned} A:=&\{(r,S,T): r\in\mathbb{R},\ S\in\mathcal{A}_h^{m_Z},\ T\in\mathcal{A}_h^{m'_Z}, \\ &\qquad \exists Z_1,\dots,Z_n\in\mathcal{A}_h,\hat{f}(Z)\leq r,\ -\hat{g}_i(Z)\leq S_i,\ \hat{h}_j(Z)=T_j,\ \forall i,j\}, \\ B:=&\{(\nu,0,0):\nu<\hat{f}(Z^\star)\}. \end{aligned} \tag{5.18}$$

Clearly, $A\cap B=\emptyset$. Also, both sets are convex. Since they live in a real normed space (namely, $\mathbb{R}\times\mathcal{A}^{m_Z+m'_Z}$) and B is open, the Hahn-Banach separation theorem [52, Theorem V.4(a)] implies that there exists a separating linear functional (ϕ,μ,λ) and $\alpha\in\mathbb{R}$ such that

$$\phi r+\sum_i\mu_i(S_i)+\sum_j\lambda_j(T_j)\geq\alpha, \quad \forall(r,S,T)\in A, \tag{5.19a}$$

$$\phi\nu\leq\alpha, \quad \forall(\nu,0,0)\in B. \tag{5.19b}$$

Note that, from the definition of A, for any $y\in\mathcal{A}$, $(r,S,T)\in A$ implies that $(r,S',T)\in A$, with $S'_i=S_i+yy^*$, and $S'_j=S_j$, for $j\neq i$. Now, suppose that there exists y such that $\mu_i(yy^*)<0$. Then, we could make the left-hand side of Eq. (5.19a) arbitrarily small, just by replacing S_i with S_i+uyy^*, with $u\in\mathbb{R}^+$ sufficiently large. It follows that, for all i, $\mu_i(yy^*)\geq 0$, i.e., $\{\mu_i\}_i$ are positive linear functionals of $\mathcal{A}$.

Notice as well that we can choose ν to be arbitrarily small in Eq. (5.19b). It follows that $\phi\geq 0$. We next prove that $\phi>0$.

From Eqs. (5.19a), (5.19b) we have that

$$\phi\hat{f}(Z)-\sum_i\mu_i(\hat{g}_i(Z))-\sum_j\lambda_j(\hat{h}_i(Z))\geq\alpha\geq\phi\hat{f}(Z^\star), \quad \forall Z\in\mathcal{A}^{\overline{\pi}}. \tag{5.20}$$

Now, take $Z = \hat{Z}$. We have that

$$\phi(\hat{f}(\hat{Z}) - \hat{f}(Z^\star)) \geq \sum_i \mu_i(\hat{g}_i(\hat{Z})). \tag{5.21}$$

Now, suppose that $\phi = 0$. Then, the equation above implies that $\mu_i(\hat{g}_i(\hat{Z})) = 0$, for all i. Since $\hat{g}_i(\hat{Z}) > 0$, it follows that $\mu_i = 0$ for all i. Hence we deduce that $\phi = 0$ implies $\mu_i = 0$ for all i. Therefore, Eq. (5.19a) implies that

$$\sum_j \lambda_j(\hat{h}_j(Z)) \geq \alpha, \quad \forall Z. \tag{5.22}$$

This can only be true if the left-hand side does not depend on Z at all. Now, let $\hat{h}_j(Z) := \sum_k \beta_{jk} Z_k - b_j$. Non-dependence on Z_k implies that the functional $\sum_j \beta_{jk}\lambda_j$ vanishes, for all k. Now, take any $W \in \mathcal{A}$ such that there exists l with $\lambda_l(W) \neq 0$. Then, we have that $\sum_j \beta_{jk}\lambda_j(W) = 0$ for all k, and thus the rows of the matrix β are not linearly independent. It follows that $\lambda_j = 0$ for all j. However, that would imply that the separating linear functional (ϕ, μ, λ) is zero, which contradicts the Hahn-Banach theorem.

From all the above it follows that $\phi > 0$. Dividing Eq. (5.20) by ϕ, we have that

$$\hat{f}(Z) - \sum_i \tilde{\mu}_i(\hat{g}_i(Z)) - \sum_j \tilde{\lambda}_j(\hat{h}_j(Z)) \geq \hat{f}(Z^\star), \quad \forall Z, \tag{5.23}$$

where $\tilde{\mu}_i := \frac{1}{\phi}\mu_i$ are positive linear functionals and $\tilde{\lambda}_j := \frac{1}{\phi}\lambda_j$ are linear functionals.

Finally, take $Z = Z^\star$ in Eq. (5.23). We arrive at:

$$f(Z^\star) - \sum_i \tilde{\mu}_i(\hat{g}_i(Z^\star)) \geq \hat{f}(Z^\star). \tag{5.24}$$

This can only be true if the second term of the left-hand-side of the equation above vanishes, i.e., if $Z^\star$ satisfies the complementary slackness condition (5.15). In that case,

$$\mathcal{L}(Z^\star, \tilde{\mu}, \tilde{\lambda}) = \hat{f}(Z^\star), \tag{5.25}$$

and so, by the above equation and (5.23), $Z^\star$ is a global solution of the unconstrained problem (5.16). □

Proof of Theorem 5.2. Let $(\mathcal{H}^\star, \sigma^\star, X^\star)$ be a solution of Problem (2.5), with $X^\star = (Y^\star, Z^\star)$. Call $\mathcal{A}$ the algebra generated by $Z^\star$. Since the only relations connecting y with z are the commutation relations (5.7), it follows that the solution $p^\star$ of Problem (2.13) satisfies

$$\begin{aligned} p^\star = \min_{Z \in \mathcal{A}^q} \ & \hat{f}(Z) \\ \text{s.t. } & \hat{g}_i(Z) \geq 0, \quad i = 1, \ldots, m_Z, \\ & \hat{h}_j(Z) = 0, \quad j = 1, \ldots, m'_Z, \end{aligned} \tag{5.26}$$

with the function $\hat{f} : \mathcal{A} \to \mathbb{C}$ defined as:

$$\hat{f}(Z) = \sigma^\star(f(Y^\star, Z)). \tag{5.27}$$

Moreover, one of the minimizers of (5.26) is $Z = Z^\star$. This function is convex by virtue of Eq. (5.11), which ensures that its Hessian is non-negative [54].

In addition, by Eqs. (5.9), (5.10), we know that the choice $\hat{Z} = Q(Z^\star)$ satisfies the constraints $\hat{g}_i(\hat{Z}) > 0$ for all i, $\hat{h}_j(\hat{Z}) = 0$ for all j. We can thus invoke Lemma 5.3 and conclude that $Z^\star$ is the solution of the unconstrained Problem (5.16), for some positive linear functionals $\{\mu_i\}_i$ and

bounded Hermitian linear functionals $\{\lambda_j\}_j$. Next, for any q-tuple of symmetric polynomials p on z, consider the following trajectory in $\mathcal{A}^q$

$$Z(t) := Z^\star + tp(Z^\star). \tag{5.28}$$

Since $Z^\star$ is a minimizer of Problem (5.16), it follows that

$$\begin{aligned} 0 &= \frac{d\mathcal{L}(Z(t);\mu,\lambda)}{dt}\Big|_{t=0} \\ &= \sigma^\star\left(\nabla_z f(Y^\star, z)(p(Z^\star))\right) - \sum_i \mu_i^k \left(\nabla_z \hat{g}_i(p)\Big|_{Z=Z^\star}\right) - \sum_j \lambda_j \left(\nabla_z \hat{h}_j(p)\Big|_{Z=Z^\star}\right). \end{aligned} \tag{5.29}$$

Since this relation is valid for arbitrary $p \in \mathbb{C}\langle z\rangle^q$, we arrive at the statement of the theorem. □

6. APPLICATIONS

6.1. Many-body quantum systems

A consequence of the state optimality condition (3.2) is that the computation of the properties of condensed matter systems at zero temperature admits an NPO formulation. Consider, for instance, an n-qubit quantum system. Each such qubit or subsystem j has an associated set of operators $\sigma_x^j, \sigma_y^j, \sigma_z^j$, which form a *Pauli algebra*:

$$\begin{aligned} &(\sigma_x^j)^2 = (\sigma_y^j)^2 = (\sigma_z^j)^2 = 1, \\ &\sigma_x^j\sigma_y^j - i\sigma_z^j = \sigma_y^j\sigma_z^j - i\sigma_x^j = \sigma_z^j\sigma_x^j - i\sigma_y^j = 0. \end{aligned} \tag{6.1}$$

In a sense, these operators represent everything we can measure in any such subsystem. Being independent systems, the operators of different subsystems commute:

$$[\sigma_a^j, \sigma_b^k] = 0, \quad a, b \in \{x, y, z\},\ j \neq k. \tag{6.2}$$

The set of constraints (6.1), (6.2) admits (up to unitary equivalence) a unique irreducible operator representation $\pi : \mathbb{C}\langle x\rangle \to B(\mathbb{C}^2)^{\otimes n}$, with

$$\begin{aligned} \pi(\sigma_x^j) &:= \mathbb{I}_2^{\otimes j-1} \otimes \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \otimes \mathbb{I}_2^{\otimes n-j}, \\ \pi(\sigma_y^j) &:= \mathbb{I}_2^{\otimes j-1} \otimes \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \otimes \mathbb{I}_2^{\otimes n-j}, \\ \pi(\sigma_z^j) &:= \mathbb{I}_2^{\otimes j-1} \otimes \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \otimes \mathbb{I}_2^{\otimes n-j}. \end{aligned} \tag{6.3}$$

It is easy to see that π satisfies the conditions of Proposition 4.20.

The n qubits jointly interact through a 2-local Hamiltonian. This is an operator of the form

$$H(\sigma) = \sum_{j>k}^{n} P_{jk}(\sigma^j, \sigma^k), \tag{6.4}$$

where $\sigma^j := (\sigma_x^j, \sigma_y^j, \sigma_z^j)$ and P_{jk} is a polynomial of degree 2. At zero temperature, the system is described by one of the eigenvectors of $\pi(H)$ with minimum eigenvalue. Any such eigenvector is called a *ground state*.

For large n, computing $E_0(H)$ is Quantum-Merlin-Arthur-hard (**QMA**-hard) [60]. Quantum chemists [4–6] (and, more recently, condensed matter physicists [10, 61–63]) use NPO to lower bound $E_0(H)$. In essence, they relax the problem

$$\begin{aligned} E_0(H) = &\min \rho(H) \\ &\text{s.t. } (\sigma_x^j)^2 = (\sigma_y^j)^2 = (\sigma_z^j)^2 = 1, \quad j = 1, \dots, n, \\ &\quad \sigma_x^j\sigma_y^j - i\sigma_z^j = \sigma_y^j\sigma_z^j - i\sigma_x^j = \sigma_z^j\sigma_x^j - i\sigma_y^j = 0, \quad j = 1, \dots, n, \\ &\quad [\sigma_a^j, \sigma_b^k] = 0, \quad a, b \in \{x, y, z\},\ j \neq k \end{aligned} \tag{6.5}$$

through hierarchies of SDPs.

Knowing the ground state energy of a condensed matter system is very useful: if positive, it signals that the system is unstable; if negative, its absolute value corresponds to the minimum energy required to disintegrate it.

However, both physicists and chemists are also interested in estimating other properties of the set of ground states. Take, for instance, the magnetization of the sample. Basic quantum mechanics teaches us that the magnetization M of a condensed matter system at zero temperature lies in $[M^-, M^+]$, with

$$M^{\pm} := \mp \min\{\mp\langle\psi|\pi(\sum_j \sigma_z^j)|\psi\rangle : \langle\psi| \bullet |\psi\rangle \in \mathrm{Gr}(\pi(H))\}. \tag{6.6}$$

For instance, Wang *et al.* [63] study a relaxation of this problem. First, using variational methods, they derive an upper bound E_0^+ on $E_0(H)$. Next, they relax the NPO problem:

$$\begin{aligned} \bar{M}^{\pm} = &\mp \min\{\mp\rho(\sum_j \sigma_z^j)\} \\ &\text{s.t. } (\sigma_x^j)^2 = (\sigma_y^j)^2 = (\sigma_z^j)^2 = 1, \quad j = 1, \dots, n, \\ &\quad \sigma_x^j\sigma_y^j - i\sigma_z^j = \sigma_y^j\sigma_z^j - i\sigma_x^j = \sigma_z^j\sigma_x^j - i\sigma_y^j = 0, \quad j = 1, \dots, n, \\ &\quad [\sigma_a^j, \sigma_b^k] = 0, \quad a, b \in \{x, y, z\},\ j \neq k, \\ &\quad \rho(H) \leq E_0^+. \end{aligned} \tag{6.7}$$

Any SDP relaxation of the problem above of order k will produce two quantities $\bar{M}_k^{\pm}$, with the property that $M \in [\bar{M}_k^-, \bar{M}_k^+]$.

However, the method proposed by Wang *et al.* [63] is only feasible when good variational methods for the considered Hamiltonian are available. Indeed, given a loose upper bound E_0^+ on $E_0(H)$, one should not expect great results. Correspondingly, the numerical results of [63] are remarkable for 1D quantum systems. Those have Hamiltonians of the form $H = \sum_j P_{j,j+1}(\sigma^j, \sigma^{j+1})$, and one can obtain good approximations to their ground state energies via tensor network state methods [64–66]. The results of [63] are not that good for 2D systems, namely, qubit systems with a Hamiltonian of the form (6.12) below. For such systems, current variational tools are very imprecise [67].

The state optimality condition (3.2) allows us to formulate Problem (6.6) as the following NPO:

$$\begin{aligned} m^{\pm} = &\mp \min \rho(\mp \sum_j \sigma_z^j) \\ &\text{s.t. } (\sigma_x^j)^2 = (\sigma_y^j)^2 = (\sigma_z^j)^2 = 1, \quad j = 1, \dots, n, \\ &\quad \sigma_x^j\sigma_y^j - i\sigma_z^j = \sigma_y^j\sigma_z^j - i\sigma_x^j = \sigma_z^j\sigma_x^j - i\sigma_y^j = 0, \quad j = 1, \dots, n, \\ &\quad [\sigma_a^j, \sigma_b^k] = 0, \quad a, b = x, y, z; \quad j \neq k, \\ &\quad \rho \in \mathrm{Gr}(H). \end{aligned} \tag{6.8}$$

In turn, the last constraint can be modeled by enforcing the relations:

$$\sigma([H,p]) = 0 \tag{6.9a}$$

$$\sigma\left(p^*Hp - \frac{1}{2}\{H,p^*p\}\right) \geq 0. \tag{6.9b}$$

Note that the quotient of $\mathbb{C}\langle\{\sigma_a^i : a = 1,2,3, i = 1,\ldots,n\}\rangle$ with the Pauli constraints (6.1), (6.2) has dimension 4^n. Hence, by Remark 3.10 and Theorem 3.9, the resulting SDP hierarchy converges to the solution of Problem (6.8).

The advantage of the formulation (6.8) with respect to (6.7) is that it does not require any upper bound on $E_0(H)$. Problem (6.8) is thus appropriate to tackle 2D and 3D systems, and even spin glasses [68].

We illustrate our technique by bounding the ground state energy and magnetization of a translation-invariant Heisenberg model in 1D and 2D with periodic boundary conditions. All calculations were done using the toolkit for non-commutative polynomial optimization Moment [69], the modeller YALMIP [70], and the solver MOSEK [71]. The code used to produce these results is available at `https://github.com/ajpgarner/heisenberg_energy`, together with code doing the direct diagonalization of the sparse Hamiltonian used to compute the exact values we compare to.

For simplicity, the only symmetry of the problem we exploited was translation invariance. A full use of the symmetries of problem, as done in Ref. [63], leads to dramatic improvements in performance.

In the 1D case, the Hamiltonian reads

$$H = \frac{1}{4}\sum_{i=0}^{n-1}\sum_{a\in\{x,y,z\}} \sigma_a^i\sigma_a^{i\oplus 1}, \tag{6.10}$$

where addition $\oplus$ is modulo n.

Results for the energy are shown in Table 6.1. Here, the lower bounds are much tighter than the upper bounds. This is because when calculating the lower bound the state optimality condition (3.2) is only a tightening of the SDP: without it, the SDP hierarchy (2.15) would converge anyway to the ground state energy. In calculating the upper bound, however, the state optimality condition (3.2) is doing all the work, as without it the SDP would converge to the system's maximum energy.

n	Lower bound	Exact value	Upper bound
6	−0.4671	−0.4671	−0.4671
7	−0.4079	−0.4079	−0.4079
8	−0.4564	−0.4564	−0.4564
9	−0.4251	−0.4219	−0.4189
10	−0.4515	−0.4515	−0.4306
11	−0.4460	−0.4290	−0.4020
12	−0.4492	−0.4489	−0.3886
13	−0.4475	−0.4330	−0.3987
14	−0.4518	−0.4474	−0.3013
15	−0.4506	−0.4356	−0.3001
16	−0.4509	−0.4464	−0.3013
17	−0.4501	−0.4373	−0.3004

Table 6.1. Ground state energy per site of 1D Heisenberg model. Up to $n = 10$ we use all nearest-neighbor monomials of degree up to 4, for $n = 11$ until $n = 13$ degree up to 3, and for higher n degree up to 2.

The magnetization is given by

$$M = \sum_{i=0}^{n-1} \sigma_z^i. \tag{6.11}$$

Note that H has the symmetry $\sigma_x^{\otimes n} H \sigma_x^{\otimes n} = H$, whereas the magnetization obeys $\sigma_x^{\otimes n} M \sigma_x^{\otimes n} = -M$. This implies that if the magnetization of the ground state $|g\rangle$ is m, then $\sigma_x^{\otimes n}|g\rangle$ will also be a ground state with magnetization $-m$. If these states are equal (up to a global phase), this implies that $m = 0$. Otherwise the ground state is degenerate and both alternatives show up. We have found numerically that for even n the magnetization per site (i.e., the magnetization divided by the number n of qubits) is always zero, and for odd n it is $\pm 1/n$.

Since the SDP respects the same symmetries as the original problem, if it gives $-m$ as a lower bound to the magnetization, it will give m as an upper bound. Therefore we have reported the numerical results only for the lower bound of the magnetization, together with the lowest exact value. Results are shown in Table 6.2.

n	Lower bound	Lowest exact value
6	0	0
7	−0.1469	−0.1429
8	0	0
9	−0.1118	−0.1111
10	−0.0315	0
11	−0.1379	−0.0909
12	−0.1422	0
13	−0.1378	−0.0769
14	−0.1780	0
15	−0.1742	−0.0667
16	−0.1715	0
17	−0.1693	−0.0588

Table 6.2. Magnetization per site of 1D Heisenberg model. Up to $n = 10$ we use all nearest-neighbor monomials of degree up to 4, for $n = 11$ until $n = 13$ degree up to 3, and for higher n degree up to 2.

In the 2D case the Hamiltonian reads

$$H = \frac{1}{4} \sum_{i,j=0}^{L-1} \sum_{a \in \{x,y,z\}} \sigma_a^{i,j} (\sigma_a^{i \oplus 1, j} + \sigma_a^{i, j \oplus 1}), \tag{6.12}$$

where $\sigma_a^{i,j}$ denotes the Pauli matrix a at the (i, j) site of the square lattice. Results for the energy are shown in Table 6.3, and for the magnetization in Table 6.4.

n	Lower bound	Exact value	Upper bound
3^2	−0.4637	−0.4410	−0.3647
4^2	−0.7077	−0.7018	−0.2383
5^2	−0.6732		−0.2388
6^2	−0.7086		−0.2389

Table 6.3. Ground state energy per site of 2D Heisenberg model. For $L = 3$ we used all nearest-neighbor monomials of degree up to 3. For $L = 4$ we used degree up to 3 for the moment matrix and up to 2 for the state optimality condition (3.2). For higher L we used degree up to 2.

n	Lower bound	Lower exact value
3^2	-0.2097	-0.1111
4^2	-0.2106	0
5^2	-0.3124	
6^2	-0.3161	

Table 6.4. Magnetization per site of 2D Heisenberg model. For $L = 3$ we used all nearest-neighbor monomials of degree up to 3. For $L = 4$ we used degree up to 3 for the moment matrix and up to 2 for the state optimality condition (3.2). For higher L we used degree up to 2.

Problem (6.8) can also be adapted to deal with the thermodynamic limit, $n = \infty$. In that case, we demand the Hamiltonian to have a special symmetry called translation invariance. For one-dimensional materials, H would be of the form:

$$H = \sum_{j=-\infty}^{\infty} P(\sigma^j, \sigma^{j+1}). \tag{6.13}$$

The reader could be worried by the fact that there are infinitely many operator variables. However, we can take the state ρ to be translation-invariant, i.e., invariant under the $*$-isomorphisms

$$\pi_R(\sigma_a^j) = \sigma_a^{j+1}, \quad \pi_L(\sigma_a^j) = \sigma_a^{j-1}. \tag{6.14}$$

In that case, one can relax the problem of minimizing the energy-per-site $e_0(H) := \min_\rho \rho(P(\sigma^1, \sigma^2))$ to

$$\begin{aligned} e_0^n := & \min \rho(P(\sigma^1, \sigma^2)) \\ \text{s.t. } & (\sigma_x^j)^2 = (\sigma_y^j)^2 = (\sigma_z^j)^2 = 1, \quad j = 1, \dots, n, \\ & \sigma_x^j \sigma_y^j - i\sigma_z^j = \sigma_y^j \sigma_z^j - i\sigma_x^j = \sigma_z^j \sigma_x^j - i\sigma_y^j = 0, \quad j = 1, \dots, n, \\ & [\sigma_a^j, \sigma_b^k] = 0, \quad a, b \in \{x, y, z\},\ j \neq k \\ & \rho(p) = \rho(\pi_L(p)) = \rho(\pi_R(p)), \quad \text{for } \ p, \pi_L(p), \pi_R(p) \in \mathbb{C}\langle \sigma^1, \dots, \sigma^n \rangle. \end{aligned} \tag{6.15}$$

It can be proven that $\lim_{n\to\infty} e_0^n$ coincides with the energy-per-site in the thermodynamic limit. The *bootstrap technique* adds to this NPO the first optimality condition (6.9a) [61, 72]. Note that, if $p \in \mathbb{C}\langle \sigma^2, \dots, \sigma^{n-1} \rangle$ has degree k, then the expression in (6.9a) has degree $k+1$ and only involves the variables $\sigma^1, \dots, \sigma^n$.

The bootstrap technique thus allows computing lower bounds on $e_0(H)$. It cannot be used, however, to bound other properties of the ground states of H.

Things change dramatically when we add the optimality condition (6.9b), for it also allows us to bound whatever local property of the system, such as the magnetization. For $1D$ Hamiltonians (6.13), if p has degree k and depends on the variables $\sigma^2, \dots, \sigma^{n-1}$, the polynomial in Eq. (6.9b) will be of degree $2k + 1$ and only depend on $\sigma^1, \dots, \sigma^n$. Thus, even though we are working in the thermodynamic limit, the state optimality condition can be evaluated. This is, in fact, the case for any translation-invariant scenario in arbitrarily many spatial dimensions.

For any local property o, the corresponding SDP hierarchies will converge to the exact interval of allowed values for o (at zero temperature).

In order to illustrate our technique we bounded the ground state properties of the 1D Heisenberg model, with Hamiltonian given by

$$H = \frac{1}{4} \sum_{i=-\infty}^{\infty} \sum_{a \in \{x,y,z\}} \sigma_a^i \sigma_a^{i+1}. \tag{6.16}$$

Results for the energy are shown in Table 6.5.

n	Lower bound	Upper bound
6	-0.4671	-0.3751
7	-0.4564	-0.3930
8	-0.4564	-0.4004
9	-0.4520	-0.4045
10	-0.4516	-0.4069
11	-0.4500	-0.4084
12	-0.4490	-0.4097

Table 6.5. Ground state energy per site of 1D Heisenberg model in the thermodynamic limit. For comparison the exact value is $1/4 - \log(2) \approx -0.4431$. We used all nearest-neighbor monomials of degree up to 4.

6.2. The curious case of quantum Bell inequalities

Consider a quantum bipartite Bell experiment [32, 33], where two separate parties conduct measurements on an entangled quantum state. The first party, Alice, conducts measurement $x \in \{1, ..., n\}$ and obtains outcome $a \in \{1, ..., d\}$. The second party, Bob, respectively calls $y \in \{1, ..., n\}, b \in \{1, ..., d\}$ his measurement setting and outcome. This configuration defines an *nndd Bell scenario*[7]. If Alice and Bob conduct many experiments, then they can estimate the probabilities $P = (P(a, b|x, y) : x, y = 1, \ldots, n;\ a, b = 1, \ldots, d)$. Given a linear functional C on P (also called a Bell functional), we wish to determine the minimum value of

$$C(P) := \sum_{a,b,x,y} C(a, b, x, y) P(a, b|x, y) \tag{6.17}$$

compatible with quantum mechanics. This leads us to formulate the following NPO:

$$\begin{aligned} c^\star = \min \sigma & \left(\frac{1}{2} \sum_{a,b,x,y} C(a, b, x, y) \{E_{a|x}, F_{b|y}\} \right) \\ \text{s.t. } & E_{a|x} \geq 0, \quad \forall a, x, \\ & \sum_a E_{a|x} - \mathbb{I} = 0, \quad \forall x, \\ & F_{b|y} \geq 0, \quad \forall b, y, \\ & \sum_b F_{b|y} - \mathbb{I} = 0, \quad \forall y, \\ & [E_{a|x}, F_{b|y}] = 0, \quad \forall a, b, x, y. \end{aligned} \tag{6.18}$$

As we can appreciate, taking the partition $X = (E, F)$, the NPO satisfies the conditions of Theorem 5.2, with $Q^A_{a|x}(E) = Q^B_{b|y}(F) = \frac{1}{d}$ for all a, b, x, y.

It so happens that the solution $(E^\star, F^\star)$ of Problem (6.18) can be chosen such that the non-commuting variables are all projectors [73]. That is,

$$\begin{aligned} (E^\star_{a|x})^2 &= E^\star_{a|x}, \quad \forall a, x, \\ (F^\star_{b|y})^2 &= F^\star_{b|y}, \quad \forall b, y. \end{aligned} \tag{6.19}$$

The standard SDP hierarchy of relaxations for problem (6.18) under restrictions (6.19) is usually dubbed the Navascués-Pironio-Acín (NPA) hierarchy [2, 3]. Since we will next define extra

[7] In general bipartite Bell scenarios, the two parties can have different numbers of measurements and different numbers of possible outcomes. We will not consider such scenarios in this article.

constraints, based on optimality conditions, to boost its speed of convergence, we will in the following refer to NPA without optimality conditions as "vanilla NPA".

Next, we apply Theorem 5.2 independently to Alice's algebra $\mathcal{A}$ (generated by the projectors $E^\star$'s) and to Bob's algebra $\mathcal{B}$ (generated by the projectors $F^\star$'s). The partial strong ncKKT conditions imply that we can, not only demand the state σ to be compatible with relations (6.19), but also the Lagrangian multipliers of Alice's $\mu^A_{a|x}, \lambda^A_x$ and Bob's $\mu^B_{b|y}, \lambda^B_y$ to be respectively compatible with the constraints $\{E^2_{a|x} - E_{a|x} = 0\}_{x,a} \cup \{1 - \sum_a E_{a|x} = 0\}_x$ and $\{F^2_{b|y} - F_{b|y} = 0\}_{y,b} \cup \{1 - \sum_b F_{b|y} = 0\}_y$.

Calling $\mathcal{P}_E$ ($\mathcal{P}_F$) the set of polynomials on the E's ($F's$), the operator optimality relations for the E's read:

$$\mu^A_{a|x}(ss^*) \geq 0, \quad \forall a, x, \ \forall s \in \mathcal{P}_E, \tag{6.20a}$$

$$\mu^A_{a|x}\Big(s\big((E_{a'|x'})^2 - E_{a'|x'}\big)s'\Big) = 0, \quad \forall a, a', x, x', \ \forall s, s' \in \mathcal{P}_E \tag{6.20b}$$

$$\mu^A_{a|x}\Big(s\Big(\sum_{a'} E_{a'|x'} - 1\Big)s'\Big) = 0, \quad \forall a, x, x', \ \forall s, s' \in \mathcal{P}_E, \tag{6.20c}$$

$$\lambda^A_{x}\Big(s\big((E_{a|x'})^2 - E_{a|x'}\big)s'\Big) = 0, \quad \forall a, x, x', \ \forall s, s' \in \mathcal{P}_E \tag{6.20d}$$

$$\lambda^A_{x}\Big(s\Big(\sum_{a} E_{a|x} - 1\Big)s'\Big) = 0, \quad \forall x, \ \forall s, s' \in \mathcal{P}_E, \tag{6.20e}$$

$$\mu^A_{a|x}(E_{a|x}) = 0, \quad \forall a, x, \tag{6.20f}$$

$$\sigma\left(\frac{1}{2}\sum_{b,y} C(a,b,x,y)\{p, F_{b|y}\}\right) = \mu^A_{a|x}(p) + \lambda^A_x(p), \quad \forall a, x, \ \forall p \in \mathcal{P}_E. \tag{6.20g}$$

The operator optimality relations for the F's are the same, under the replacements $E \to F$, $a \to b$, $x \to y$, $\mathcal{P}_E \to \mathcal{P}_F$, $A \to B$. The reader can find the full optimization problem, including the state optimality conditions in Appendix C.

6.2.1. Only two outcomes

When a, b can only take two values, it is customary to rewrite Problem (6.18) in terms of "dichotomic operators" $A_x := E_{1|x} - E_{2|x}, B_y := F_{1|y} - F_{2|y}$. The problem to solve is thus

$$\begin{aligned} &\min \sigma(H) \\ &\text{s.t.} \ \frac{1 - A_x}{2} \geq 0, \ \frac{1 + A_x}{2} \geq 0, \quad \forall x, \\ &\quad \frac{1 - B_y}{2} \geq 0, \ \frac{1 + B_y}{2} \geq 0, \quad \forall y, \\ &\quad [A_x, B_y] = 0, \quad \forall x, y \end{aligned} \tag{6.21}$$

where H is the Bell polynomial

$$\frac{1}{2}\sum_{x,y} c_{xy}\{A_x, B_y\} + \sum_x d_x A_x + \sum_y e_y B_y. \tag{6.22}$$

To simplify notation, we define the polynomials

$$\mathcal{F}_x := \sum_y \frac{1}{2} c_{xy} B_y + \frac{1}{2} d_x \mathbb{I}, \tag{6.23a}$$

$$\mathcal{G}_y := \sum_x \frac{1}{2} c_{xy} A_x + \frac{1}{2} e_y \mathbb{I}, \tag{6.23b}$$

which allow us to express H as

$$H = \sum_x \{\mathcal{F}_x, A_x\} + \sum_y e_y B_y = \sum_y \{\mathcal{G}_y, B_y\} + \sum_x d_x A_x. \tag{6.24}$$

As before, it can be shown that the minimizers $(A^\star, B^\star)$ can be chosen such that

$$(A_x^\star)^2 = (B_y^\star)^2 = 1. \tag{6.25}$$

Thus, once more we can apply Theorem 5.2 to the algebra $\mathcal{A}$ generated by $A_1^\star, \ldots, A_n^\star$ and conclude that one can add new state multipliers μ_x^+, μ_x^- to the problem, with the properties:

$$\mu_x^\pm(s(A_{x'}^2 - 1)s') = 0, \quad \forall x, x', \ \forall s, s' \in \mathcal{A}, \tag{6.26a}$$

$$\mu_x^\pm\left(\frac{1 \pm A_x}{2}\right) = 0, \quad \forall x, \tag{6.26b}$$

$$\sigma\left(\{p, \mathcal{F}_x\}\right) = \mu_x^+(p) - \mu_x^-(p), \quad \forall x, \ \forall p \in \mathcal{A}. \tag{6.26c}$$

If one does not wish to introduce new variables $\mu_x^\pm$ into the NPO problem, it is easy to get a relaxation of the conditions above that only involves evaluations with the already existing variable σ.

Let $E_x^\pm := \frac{1 \pm A_x}{2}$. From Eq. (6.26b) and the positivity of $\mu_x^\pm$, an analogous argument to the one used to derive Eqs. (4.90) shows that

$$\mu_x^\pm\left(E_x^\pm p\right) = \mu_x^\pm\left(p E_x^\pm\right) = 0 \quad \forall p \in \mathcal{A}. \tag{6.27}$$

Taking $p = \{A_x, q\}$ we find that

$$\mu_x^\pm\left(\{A_x, q\}\right) = \mp\mu_x^\pm\left(q + A_x q A_x\right) = 0. \tag{6.28}$$

Thus, if we set $p = -\{A_x, q\}$ in Eq. (6.26c), we arrive that

$$-\sigma\left(\{\{A_x, q\}, \mathcal{F}_x\}\right) = \mu^+(q + A_x q A_x) + \mu^-(q + A_x q A_x). \tag{6.29}$$

In particular, taking $q = ss^*$, we have that

$$-\sigma\left(\{\{A_x, ss^*\}, \mathcal{F}_x\}\right) = \mu^+(ss^* + A_x ss^* A_x) + \mu^-(ss^* + A_x ss^* A_x) \geq 0, \quad \forall s \in \mathcal{P}_E. \tag{6.30}$$

Setting $p = [A_x, q]$ in Eq. (6.26c) and using Eq. (6.27), we obtain another useful constraint:

$$\sigma\left(\{[A_x, q], \mathcal{F}_x\}\right) = 0, \quad \forall q \in \mathcal{A}. \tag{6.31}$$

Constraints (6.30), (6.31) are, respectively, extra positivity and linear conditions that one can apply to the already existing variables of the 'quantum NPO' (6.18).

6.2.2. Numerical implementation

In order to implement numerically the constraints (6.30) and (6.31), together with the analogous constraints for Bob and the state optimality conditions (3.2), we express them in terms of a basis of monomials. Let $\{m_i^A\}_i$ and $\{m_i^B\}_i$ be a basis of monomials belonging to Alice's and Bob's algebra of operators, and $\{m_i\}_i$ a basis for the entire algebra. Then the equality constraints become

$$\sigma\left(\mathcal{F}_x[A_x, m_i^A]\right) = 0, \tag{6.32a}$$

$$\sigma\left(\mathcal{G}_y[B_y, m_i^B]\right) = 0, \tag{6.32b}$$

$$\sigma\left([H, m_i]\right) = 0, \tag{6.32c}$$

where we are using the fact that $\mathcal{F}_x$ commutes with every element of Alice's algebra, and the analogous condition for Bob. The positivity conditions (6.30) are equivalent to the positive semidefiniteness of the matrices $\{\alpha^x\}_x, \{\beta^y\}_y, \gamma$, with elements given by

$$\alpha^x_{ij} := -\sigma\left(\mathcal{F}_x\{A_x, m_i^{A^*} m_j^A\}\right), \tag{6.33a}$$

$$\beta^y_{ij} := -\sigma\left(\mathcal{G}_y\{B_y, m_i^{B^*} m_j^B\}\right), \tag{6.33b}$$

$$\gamma_{ij} := \sigma\left(m_i^* H m_j - \frac{1}{2}\{H, m_i^* m_j\}\right). \tag{6.33c}$$

Note that, when dealing with Bell inequalities, it is more usual to formulate the problem as a maximization instead of a minimization [2]. One can adapt the KKT constraints for maximization by simply flipping the sign of the positivity conditions (6.33).

Most commercial SDP solvers are based on interior-point methods, as they guarantee fast and accurate results. In order for interior point algorithms to work reliably, it is however vital to ensure that the problem we are solving is strictly feasible, that is, that there exists a point that satisfies all the equality constraints and has strictly positive eigenvalues in the positive semidefiniteness constraints [74]. Although the vanilla NPA hierarchy is always strictly feasible [75], this is in general not true when additional constraints are enforced [76]. This is in fact the case here, as the matrix γ will necessarily have linearly dependent columns, and therefore some of its eigenvalues will be zero. To see that, we use Eq. (6.32c) to rewrite Eq. (6.33c) as

$$\gamma_{ij} = \sigma(m_i^*[H, m_j]). \tag{6.34}$$

This implies that a sufficient condition for some columns $\gamma_{_j}$ to be linearly dependent is that the corresponding operators $[H, m_j]$ are linearly dependent. This is always the case if $m_j = \mathbb{I}$, as $[H, \mathbb{I}] = 0$, or if $\{m_j\}_j$ is a set of monomials that can express H itself, as $[H, H] = 0$. Additional linear dependencies can show up for specific choices for H. In the Bell inequalities we consider in this section, the only additional dependencies that appeared were in the case of the tilted CHSH (6.35), for which $[H, \{A_0, A_1\}] = [H, \{B_0, B_1\}] = 0$. We removed these dependencies simply by removing enough monomials from the set used to define γ. We verified numerically that after doing that, the problem was always strictly feasible.

We illustrate the technique by minimizing Bell functionals in the 2222, 3322, and 4422 Bell scenarios. All calculations were done using the toolkit for non-commutative polynomial optimization Moment [69], the modeller YALMIP [70], and the arbitrary-precision solver SDPA-GMP [77]. We are particularly interested in checking whether we have achieved convergence at some level. To do so, we verify that the moment matrix is flat [15, 78–81] or, in physicists' slang, that it has a *rank loop* [3]. We remark under the corresponding table whether it holds.

We start with a tilted version of the CHSH inequality [82], with an additional $\tau(A_0 + B_0)$ term [83, 84]. In full correlation notation the coefficients table is given by

$$\left(\begin{array}{c|cc} 0 & \tau & 0 \\ \hline \tau & 1 & 1 \\ 0 & 1 & -1 \end{array}\right). \tag{6.35}$$

This table, which represents a Hermitian polynomial on the operators $1, \{A_x\}_x, \{B_y\}_y$ has to be understood as follows: the rows are labeled by the operators $O = (1, A_1, A_2)$; the columns, by the operators $O' = (1, B_1, B_2)$. The O_j, O'_k-th entry of the table corresponds to the coefficient multiplying $\frac{1}{2}\{O_j, O'_k\}$.

For two different values of τ, we upper bound the maximum average of this polynomial, see Tables 6.6 and 6.7. As τ tends to 1 the level at which the NPA hierarchy converges exactly seems to get ever higher.

level	NPA	NPA+KKT
2	3.9003 2967	3.9003 1859
3	3.9001 6474	3.9001 6389
4	3.9001 6389	

Table 6.6. Results for the tilted CHSH inequality with $\tau = 0.95$. For comparison, the analytical value is 3.9001 6389 9372 [85]. With the KKT constraints we get a rank loop at level 3, and without at level 4.

level	NPA	NPA+KKT
2	3.9800 1157	3.9800 1078
3	3.9800 0416	3.9800 0280
4	3.9800 0217	3.9800 0132
5	3.9800 0156	
6	3.9800 0132	

Table 6.7. Results for the tilted CHSH inequality with $\tau = 0.99$. For comparison, the analytical value is 3.9800 0132 8893 [85]. With the KKT constraints we find a rank loop at level 4; without, at level 7.

Our next example is the well-studied I3322 inequality [86–88]. In full correlation notation the coefficients table is given by

$$\frac{1}{4}\left(\begin{array}{c|cc} 0 & -1 & -1 & 0 \\ \hline -1 & -1 & -1 & -1 \\ -1 & -1 & -1 & 1 \\ 0 & -1 & 1 & 0 \end{array}\right). \tag{6.36}$$

The results are shown in Table 6.8. Its maximal violation is conjectured to occur only for an infinite-dimensional system [87], and a rank loop implies the existence of a finite-dimensional system achieving the maximum. Therefore we expected to find no rank loop here, as was indeed the case. For increased efficiency the calculations here were done without the positivity conditions (6.33), as they did not seem to improve the results.

level	NPA	NPA+KKT
2	1.2509 3972	1.2509 3965
3	1.2508 7556	1.2508 7554
4	1.2508 7540	1.2508 7538
5	1.2508 7538	

Table 6.8. Results for the I3322 inequality. For comparison, the best known lower bound is 1.2508 7538 4513. No rank loop was found.

Our final example is the I^{20}_{4422} inequality [89], that had a gap between the best known lower bound and the best known upper bound (see Table IV of Ref. [90]). In full correlation notation the coefficients table is given by

$$\frac{1}{4}\left(\begin{array}{c|cccc} -12 & -1 & -1 & -2 & 4 \\ \hline -1 & -1 & 1 & 1 & 2 \\ -1 & 1 & -1 & 1 & 2 \\ -2 & 1 & 1 & -2 & 2 \\ 4 & 2 & 2 & 2 & -2 \end{array}\right). \tag{6.37}$$

The results are shown in Table 6.9. For increased efficiency the calculations here were done without the positivity conditions (6.33), as they did not seem to improve the results.

level	NPA	NPA + KKT
2	0.5070 6081	0.5020 4577
3	0.4677 5783	0.4676 7939
4	0.4676 7939	

Table 6.9. Results for the inequality I^{20}_{4422}. For comparison, the best known lower bound is 0.4676 7939. With the KKT constraints we get a rank loop at level 3, and without at level 4.

We have also experimented with the weak ncKKT conditions from Definition 3.16, even though they are not proven to hold for this problem. In all cases we got the same numerical answer as with the partial KKT conditions, except in the tilted CHSH case (6.35), where we got numerical problems.

7. CONCLUSION

In this work, we have generalized the KKT optimality conditions to non-commutative polynomial optimization problems (NPO). These enforce new equality and positive semidefinite conditions on the already existing hierarchies of SDPs used in NPO.

The state optimality conditions (Eq. (3.2)) and essential ncKKT (Eq. (3.30)), a loose version of the operator optimality conditions, hold for all problems. In contrast, normed and strong ncKKT conditions need to be justified through some constraint qualification. The existence of an SOS certificate to solve the NPO problem is enough to guarantee that the strong ncKKT conditions hold. This property is difficult to verify for most NPO problems *a priori*. However, we found that it is satisfied by all Archimedean NPO problems with strictly feasible convex constraints or a faithful finite-dimensional representation.

In addition, we generalized a known 'classical' qualification constraint: Mangasarian–Fromovitz Constraint Qualification (MFCQ), which legitimates the use of normed ncKKT conditions in NPO. We also presented very mild conditions that guarantee that at least some relaxed form of either the normed or the strong ncKKT conditions holds. Those conditions are satisfied in the NPO formulation of quantum nonlocality, and thus have immediate practical applications.

We tested the effectiveness of the non-commutative KKT conditions by upper bounding the maximal violation of bipartite Bell inequalities in quantum systems. We found that the partial ncKKT conditions do improve the speed of convergence of the SDP hierarchy, sometimes achieving convergence at a finite level. This hints that the collapse of Lasserre's hierarchy of SDP relaxations [25] under the KKT constraints, proven in [27], might extend to the non-commutative case.

Similarly, we applied the state optimality conditions to bound the local properties of ground states of many-body quantum systems. Prior to our work, there was no mathematical tool capable of delivering rigorous bounds that did not rely on variational methods, see [63]. It is intriguing whether the state optimality conditions can be integrated within renormalization flow techniques, like those proposed in [62]. That would allow one to skip several levels of the SDP hierarchy through a careful (Hamiltonian-dependent) trimming of irrelevant degrees of freedom, thus delivering much tighter bounds on key physical properties.

ACKNOWLEDGMENTS

This project was funded within the QuantERA II Programme that has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement No 101017733, and from the Austrian Science

Fund (FWF), project I-6004. M.A and A.G. acknowledge funding from the FWF stand-alone project P 35509-N. M.A. also acknowledges support by the European Union–Next Generation UE/MICIU/Plan de Recuperación, Transformación y Resiliencia/Junta de Castilla y León, by the Department of Education of the Junta de Castilla y León and FEDER Funds (Reference: CLU-2023-1-05), and by the Spanish Agencia Estatal de Investigación, Grants No. RYC2023-044074-I and PID2024-161725OA-I00. I.K.'s work was performed within the project COMPUTE, funded within the QuantERA II Programme that has received funding from the EU's H2020 research and innovation programme under the GA No 101017733. I.K. was also supported by the Slovenian Research Agency program P1-0222 and grants J1-50002, N1-0217, J1-60011, J1-50001, J1-3004 and J1-60025. Partially supported by the Fondation de l'Ecole polytechnique as part of the Gaspard Monge Visiting Professor Program. I.K. thanks Ecole Polytechnique and Inria Paris Saclay for hospitality during the preparation of this manuscript. T.V. acknowledges the support of the European Union (QuantERA eDICT, CHIST-ERA MoDIC), the National Research, Development and Innovation Office NKFIH (Grants No. 2019-2.1.7-ERA-NET-2020-00003, No. 2023-1.2.1-ERA_NET-2023-00009 and No. K145927) and the 'Frontline' Research Excellence Program of the NKFIH (No. KKP133827).

[1] J. W. Helton, H. Mousavi, S. S. Nezhadi, V. I. Paulsen, and T. B. Russell, Synchronous values of games, Annales Henri Poincaré **25**, 4357–4397 (2024).
[2] M. Navascués, S. Pironio, and A. Acín, Bounding the set of quantum correlations, Physical Review Letters **98**, 010401 (2007), arXiv:quant-ph/0607119.
[3] M. Navascués, S. Pironio, and A. Acín, A convergent hierarchy of semidefinite programs characterizing the set of quantum correlations, New Journal of Physics **10**, 073013 (2008), arXiv:0803.4290 [quant-ph].
[4] M. Nakata, H. Nakatsuji, M. Ehara, M. Fukuda, K. Nakata, and K. Fujisawa, Variational calculations of fermion second-order reduced density matrices by semidefinite programming algorithm, The Journal of Chemical Physics **114**, 8282 (2001).
[5] D. A. Mazziotti, Realization of quantum chemistry without wave functions through first-order semidefinite programming, Physical Review Letters **93**, 213001 (2004).
[6] D. A. Mazziotti, Quantum many-body theory from a solution of the n-representability problem, Physical Review Letters **130**, 153001 (2023), arXiv:2304.08570 [quant-ph].
[7] T. Barthel and R. Hübener, Solving condensed-matter ground-state problems by semidefinite relaxations, Physical Review Letters **108**, 200404 (2012), arXiv:1106.4966 [cond-mat.str-el].
[8] T. Baumgratz and M. B. Plenio, Lower bounds for ground states of condensed matter systems, New Journal of Physics **14**, 023027 (2012), arXiv:1106.5275 [quant-ph].
[9] A. Haim, R. Kueng, and G. Refael, Variational-correlations approach to quantum many-body problems (2020), arXiv:2001.06510 [cond-mat.str-el].
[10] B. Requena, G. Muñoz-Gil, M. Lewenstein, V. Dunjko, and J. Tura, Certificates of quantum many-body properties assisted by machine learning, Physical Review Research **5**, 013097 (2023), arXiv:2103.03830 [quant-ph].
[11] S. Lawrence, Semidefinite programs at finite fermion density, Physical Review D **107**, 094511 (2023), arXiv:2211.08874 [hep-lat].
[12] S. Pironio, M. Navascués, and A. Acín, Convergent relaxations of polynomial optimization problems with noncommuting variables, SIAM Journal on Optimization **20**, 2157 (2010), arXiv:0903.4368 [math.OC].
[13] J. W. Helton and S. McCullough, A Positivstellensatz for non-commutative polynomials, Transactions of the American Mathematical Society **356**, 3721 (2004).
[14] A. C. Doherty, Y.-C. Liang, B. Toner, and S. Wehner, The quantum moment problem and bounds on entangled multi-prover games, in *2008 23rd Annual IEEE Conference on Computational Complexity* (IEEE, 2008) pp. 199–210, arXiv:0803.4373 [quant-ph].
[15] S. Burgdorf, I. Klep, and J. Povh, *Optimization of polynomials in non-commuting variables*, Vol. 2 (Springer, 2016).
[16] L. Vandenberghe and S. Boyd, Semidefinite programming, SIAM Review **38**, 49 (1996).

[17] Y. Nesterov and A. Nemirovskii, *Interior-point polynomial algorithms in convex programming*, SIAM Stud. Appl. Math., Vol. 13 (Philadelphia, PA: SIAM, Society for Industrial and Applied Mathematics, 1994).

[18] W. Karush, Minima of functions of several variables with inequalities as side constraints, M. Sc. Dissertation. Dept. of Mathematics, Univ. of Chicago (1939).

[19] H. W. Kuhn and A. W. Tucker, Nonlinear programming, in *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 2 (University of California Press, 1951) pp. 481–493.

[20] J. Nocedal and S. Wright, *Numerical Optimization* (Springer New York, 2006).

[21] H. Fawzi, O. Fawzi, and S. O. Scalet, Certified algorithms for equilibrium states of local quantum hamiltonians, Nature Communications **15**, 10.1038/s41467-024-51592-3 (2024).

[22] M. Laurent, Sums of squares, moment matrices and optimization over polynomials, in *Emerging Applications of Algebraic Geometry*, edited by M. Putinar and S. Sullivant (Springer New York, New York, NY, 2009) pp. 157–270.

[23] J. B. Lasserre, *An introduction to polynomial and semi-algebraic optimization*, Camb. Texts Appl. Math. (Cambridge: Cambridge University Press, 2015).

[24] J. Nie, *Moment and polynomial optimization*, MOS-SIAM Ser. Optim., Vol. 31 (Philadelphia, PA: Society for Industrial and Applied Mathematics (SIAM), 2023).

[25] J. B. Lasserre, Global optimization with polynomials and the problem of moments, SIAM Journal on Optimization **11**, 796 (2001).

[26] P. Parrilo, Semidefinite programming relaxations for semialgebraic problems, Mathematical Programming, Series B **96** (2003).

[27] J. Nie, An exact Jacobian SDP relaxation for polynomial optimization, Mathematical Programming **137**, 225 (2013), arXiv:1006.2418 [math.OC].

[28] A. W. Harrow, A. Natarajan, and X. Wu, An improved semidefinite programming hierarchy for testing entanglement, Communications in Mathematical Physics **352**, 881 (2017), arXiv:1506.08834 [quant-ph].

[29] M. Navascués, S. Pironio, and A. Acín, SDP relaxations for non-commutative polynomial optimization, in *Handbook on Semidefinite, Conic and Polynomial Optimization*, edited by M. F. Anjos and J. B. Lasserre (Springer US, New York, NY, 2012) pp. 601–634.

[30] M. Takesaki, *Theory of operator algebras I.*, 2nd ed., Encycl. Math. Sci., Vol. 124 (Berlin: Springer, 2002).

[31] M. Putinar, Positive polynomials on compact semi-algebraic sets, Indiana University Mathematics Journal **42**, 969 (1993).

[32] J. S. Bell, On the Einstein Podolsky Rosen paradox, Physics Physique Fizika **1**, 195–200 (1964).

[33] B. S. Tsirel'son, Quantum analogues of the Bell inequalities. the case of two spatially separated domains, Journal of Soviet Mathematics **36**, 557 (1987).

[34] N. Brunner, D. Cavalcanti, S. Pironio, V. Scarani, and S. Wehner, Bell nonlocality, Reviews of Modern Physics **86**, 419 (2014), arXiv:1303.2849 [quant-ph].

[35] P. W. Anderson, *Basic Notions of Condensed Matter Physics*, edited by P. W. Anderson (CRC Press, 2018).

[36] I. Vidav, On some regular rings, Acad. Serbe Sci., Publ. Inst. Math. **13**, 73 (1959).

[37] J. Cimprič, A representation theorem for Archimedean quadratic modules on *-rings., Can. Math. Bull. **52**, 39 (2009).

[38] I. Klep and M. Schweighofer, A nichtnegativstellensatz for polynomials in noncommuting variables, Israel Journal of Mathematics **161**, 17 (2007).

[39] S. Pironio, A. Acín, S. Massar, A. B. de la Giroday, D. N. Matsukevich, P. Maunz, S. Olmschenk, D. Hayes, L. Luo, T. A. Manning, and C. Monroe, Random numbers certified by Bell's theorem, Nature **464**, 1021 (2010).

[40] I. Gelfand and M. Neumark, On the imbedding of normed rings into the ring of operators in Hilbert space, Rec. Math. [Mat. Sbornik] N.S **12**, 197 (1943).

[41] I. E. Segal, Irreducible representations of operator algebras, Bulletin of the American Mathematical Society **53**, 73–88 (1947).

[42] M. F. Anjos and J. B. Lasserre, *Handbook on semidefinite, conic and polynomial optimization*, Vol. 166 (Springer Science & Business Media, 2011).

[43] I. Klep, V. Magron, J. Volčič, and J. Wang, State polynomials: positivity, optimization and nonlinear Bell inequalities, Math. Program. **207**, 645 (2024).

[44] D. S. Kaliuzhnyi-Verbovetskyi and V. Vinnikov, *Foundations of free noncommutative function the-*

ory, Math. Surv. Monogr., Vol. 199 (Providence, RI: American Mathematical Society (AMS), 2014).
[45] J. W. Helton, I. Klep, and S. McCullough, Free convex algebraic geometry, in *Semidefinite optimization and convex algebraic geometry* (Philadelphia, PA: Society for Industrial and Applied Mathematics (SIAM), 2013) pp. 341–405.
[46] T. S. Motzkin, The arithmetic-geometric inequality, in *Inequalities. Proc. of Sympos. at Wright-Patterson AFB*, edited by O. Shisha (1965) pp. 205–224.
[47] P. C. Rosenbloom, The majorant method, Proc. Sympos. Pure Math. 4, 51-72 (1961).
[48] M. Shinbrot and R. R. Welland, The Cauchy-Kowalewskaya theorem, J. Math. Anal. Appl. **55**, 757 (1976).
[49] J. Dieudonné, *Foundations of modern analysis. Enlarged and corrected printing* (New York-London: Academic Press. xv, 387 p., 1969).
[50] H. Cartan, *Differential calculus* (Hermann, Paris; Houghton Mifflin Co., Boston, MA, 1971) p. 160, exercises by C. Buttin, F. Rideau and J. L. Verley, Translated from the French.
[51] A. Ambrosetti and G. Prodi, *A primer of nonlinear analysis*, Camb. Stud. Adv. Math., Vol. 34 (Cambridge: Cambridge University Press, 1995).
[52] M. Reed and B. Simon, *Methods of modern mathematical physics. I: Functional analysis.* (1980).
[53] N. Ozawa, About the Connes embedding conjecture, Japanese Journal of Mathematics **8**, 147–183 (2013), arXiv:1212.1700 [math.OA].
[54] J. W. Helton and S. McCullough, Convex noncommutative polynomials have degree two or less, SIAM Journal on Matrix Analysis and Applications **25**, 1124 (2004).
[55] F. Zhang, *The Schur Complement and Its Applications*, Numerical Methods and Algorithms (Springer US, 2006).
[56] A. Barvinok, *A Course in Convexity*, Graduate studies in mathematics (American Mathematical Society, 2002).
[57] J. W. Helton, I. Klep, and S. McCullough, The convex Positivstellensatz in a free algebra, Advances in Mathematics **231**, 516–534 (2012), arXiv:1102.4859 [math.RA].
[58] L. T. Ligthart and D. Gross, The inflation hierarchy and the polarization hierarchy are complete for the quantum bilocal scenario, Journal of Mathematical Physics **64**, 10.1063/5.0143792 (2023).
[59] E. Wolfe, A. Pozas-Kerstjens, M. Grinberg, D. Rosset, A. Acín, and M. Navascués, Quantum inflation: A general approach to quantum causal compatibility, Physical Review X **11**, 10.1103/physrevx.11.021043 (2021).
[60] Y.-K. Liu, M. Christandl, and F. Verstraete, Quantum computational complexity of the N-representability problem: QMA complete, Physical Review Letters **98**, 110503 (2007), arXiv:quant-ph/0609125.
[61] X. Han, Quantum many-body bootstrap (2020), arXiv:2006.06002 [cond-mat.str-el].
[62] I. Kull, N. Schuch, B. Dive, and M. Navascués, Lower bounds on ground-state energies of local hamiltonians through the renormalization group, Physical Review X **14**, 021008 (2024).
[63] J. Wang, J. Surace, I. Frérot, B. Legat, M.-O. Renou, V. Magron, and A. Acín, Certifying ground-state properties of many-body systems, Physical Review X **14**, 031006 (2024).
[64] J. I. Cirac, D. Perez-Garcia, N. Schuch, and F. Verstraete, Matrix product states and projected entangled pair states: Concepts, symmetries, theorems, Reviews of Modern Physics **93**, 045003 (2021), arXiv:2011.12127 [quant-ph].
[65] Z. Landau, U. Vazirani, and T. Vidick, A polynomial time algorithm for the ground state of one-dimensional gapped local Hamiltonians, Nature Physics **11**, 566 (2015), arXiv:1307.5143 [quant-ph].
[66] F. Verstraete and J. I. Cirac, Matrix product states represent ground states faithfully, Physical Review B **73**, 094423 (2006), arXiv:cond-mat/0505140.
[67] N. Schuch, M. M. Wolf, F. Verstraete, and J. I. Cirac, Computational complexity of projected entangled pair states, Physical Review Letters **98**, 140506 (2007), arXiv:quant-ph/0611050.
[68] M. Mézard, G. Parisi, and M. A. Virasoro, *Spin glass theory and beyond: An Introduction to the Replica Method and Its Applications*, Vol. 9 (World Scientific Publishing Company, 1987).
[69] A. J. Garner and M. Araújo, Introducing moment: A toolkit for semi-definite programming with moment matrices (2024), arXiv:2406.15559 [quant-ph].
[70] J. Löfberg, YALMIP: a toolbox for modeling and optimization in MATLAB, in *Proceedings of the CACSD Conference* (Taipei, Taiwan, 2004) pp. 284–289, `https://yalmip.github.io/`.
[71] MOSEK ApS, *The MOSEK Optimization Suite* (2024), `https://docs.mosek.com/latest/intro/index.html`.
[72] X. Han, S. A. Hartnoll, J. Kruthoff, *et al.*, Bootstrapping matrix quantum mechanics, Physical Review Letters **125**, 041601 (2020).

[73] M. Junge, M. Navascues, C. Palazuelos, D. Perez-Garcia, V. B. Scholz, and R. F. Werner, Connes' embedding problem and Tsirelson's problem, Journal of Mathematical Physics **52**, 012102 (2011), arXiv:1008.1142 [math-ph].

[74] D. Drusvyatskiy and H. Wolkowicz, The many faces of degeneracy in conic optimization, Foundations and Trends in Optimization **3**, 77 (2017), arXiv:1706.03705 [math.OC].

[75] A. Tavakoli, A. Pozas-Kerstjens, P. Brown, and M. Araújo, Semidefinite programming relaxations for quantum correlations, Reviews of Modern Physics **96**, 02551 (2024), arXiv:2307.02551 [quant-ph].

[76] M. Araújo, Comment on 'Geometry of the quantum set on no-signaling faces', Physical Review A **107**, 036201 (2023), arXiv:2302.03529 [quant-ph].

[77] M. Nakata, A numerical evaluation of highly accurate multiple-precision arithmetic version of semidefinite programming solver: SDPA-GMP, -QD and -DD., in *2010 IEEE International Symposium on Computer-Aided Control System Design* (2010) pp. 29–34, `https://github.com/nakatamaho/sdpa-gmp`, `https://github.com/nakatamaho/sdpa-qd`, `https://github.com/nakatamaho/sdpa-dd`.

[78] D. Henrion and J.-B. Lassere, Detecting global optimality and extracting solutions in GloptiPoly, in *Positive polynomials in control.* (Berlin: Springer, 2005) pp. 293–310.

[79] J. Nie, Certifying convergence of Lasserre's hierarchy via flat truncation, Mathematical Programming **142**, 485 (2013), arXiv:1106.2384 [math.OC].

[80] M. Laurent, Optimization over polynomials: selected topics, in *Proceedings of the International Congress of Mathematicians (ICM 2014), Seoul, Korea, August 13–21, 2014. Vol. IV: Invited lectures* (Seoul: KM Kyung Moon Sa, 2014) pp. 843–869.

[81] V. Magron and J. Wang, *Sparse polynomial optimization. Theory and practice*, Ser. Optim. Appl., Vol. 5 (Singapore: World Scientific, 2023) arXiv:2208.11158 [math.OC].

[82] J. F. Clauser, M. A. Horne, A. Shimony, and R. A. Holt, Proposed experiment to test local hidden-variable theories, Physical Review Letters **23**, 880 (1969).

[83] P. H. Eberhard, Background level and counter efficiencies required for a loophole-free Einstein-Podolsky-Rosen experiment, Physical Review A **47**, R747 (1993).

[84] Y.-C. Liang, T. Vértesi, and N. Brunner, Semi-device-independent bounds on entanglement, Physical Review A **83**, 022108 (2011), arXiv:1012.1513 [quant-ph].

[85] N. Gigena, E. Panwar, G. Scala, M. Araújo, M. Farkas, and A. Chaturvedi, Self-testing tilted strategies for maximal loophole-free nonlocality, npj Quantum Information **11**, 82 (2025), arXiv:2405.08743 [quant-ph].

[86] D. Collins and N. Gisin, A relevant two qubit Bell inequality inequivalent to the CHSH inequality, Journal of Physics A: Mathematical and General **37**, 1775 (2004), arXiv:quant-ph/0306129.

[87] K. F. Pál and T. Vértesi, Maximal violation of a bipartite three-setting, two-outcome Bell inequality using infinite-dimensional quantum systems, Physical Review A **82**, 022116 (2010), arXiv:1006.3032 [quant-ph].

[88] D. Rosset, SymDPoly: symmetry-adapted moment relaxations for noncommutative polynomial optimization (2018), arXiv:1808.09598.

[89] N. Brunner and N. Gisin, Partial list of bipartite Bell inequalities with four binary settings, Physics Letters A **372**, 3162 (2008), arXiv:0711.3362 [quant-ph].

[90] K. F. Pál and T. Vértesi, Quantum bounds on Bell inequalities, Physical Review A **79**, 022120 (2009), arXiv:0810.1615 [quant-ph].

[91] I. Klep, V. Magron, and J. Volčič, Optimization over trace polynomials, Annales Henri Poincaré **23**, 67–100 (2021).

Appendix A: Heuristic derivation of the first-order optimality conditions

The KKT conditions can be non-rigorously derived by considering infinitesimal variations over $x^\star$ of the Lagrangian functional

$$\mathcal{L}_c(x,\mu,\lambda) = f(x) - \sum_{i\in\mathbb{A}(x^\star)} \mu_i g_i(x) - \sum_j \lambda_j h_j(x), \tag{A.1}$$

with $\{\mu_i\}_i \subset \mathbb{R}^+$, $\{\lambda_j\}_j \subset \mathbb{R}$.

In this work, we seek the non-commutative analogs of the first-order optimality conditions (2.2). To find them, we start by writing a Lagrangian for Problem (2.5).

Let $(\mathcal{H}^\star, X^\star, \psi^\star)$ be any bounded solution of Problem (2.5). For technical convenience, rather than considering general variations of the operators $X_1^\star, \ldots, X_n^\star$ within $B(\mathcal{H}^\star)$, we will demand those to be within $C^*(X^\star)$, the unital C^*-algebra generated by the operators $X^\star := (X_1^\star, \ldots, X_n^\star)$. Consequently, from now on we regard the state $\psi^\star$ as a linear, positive, normalized functional on $C^*(X^\star)$.

Note that any operator inequality constraint $g(\bar{X}) \geq 0$ can be interpreted as an infinity of inequality constraints of the form $\omega(g(\bar{X})) \geq 0$ for all states $\omega : C^*(X^\star) \to \mathbb{C}$. The set of active constraints at $X^\star$ thus corresponds to $\{\omega(g(\bar{X})) \geq 0 : \omega \geq 0,\ \omega(g(X^\star)) = 0\}$. Similarly, the equality constraint $h(\bar{X}) = 0$ is equivalent to the condition $\xi(h(\bar{X})) = 0$, for all Hermitian linear functionals $\xi : C^*(X^\star) \to \mathbb{C}$, and the set of active constraints associated to the positivity of the state $\psi^\star$ is $\{\psi(w) \geq 0 : w \geq 0, \psi^\star(w) = 0\}$.

Bearing the last two paragraphs in mind, the non-commuting analog of the classical Lagrangian (A.1) is:

$$\begin{aligned}\mathcal{L} = \psi(f(\bar{X})) &- \int_{\substack{\psi^\star(w)=0,\\ w\geq 0}} M(w)\mathrm{d}w\psi\,(w) + \alpha(1-\psi(1)) \\ &- \sum_i \int_{\substack{\omega(g_i(X^\star))=0,\\ \omega \text{ state}}} \nu_i(\omega)\mathrm{d}\omega\,\omega(g_i(\bar{X})) - \sum_j \int \Theta_j(\xi)\mathrm{d}\xi\,\xi(h_j(\bar{X})),\end{aligned} \tag{A.2}$$

where the Hermitian operator variables $\bar{X} = (\bar{X}_1, \ldots, \bar{X}_n)$ are elements of $C^*(X^\star)$ and $\psi : C^*(X^\star) \to \mathbb{C}$ represents our state variable. The multipliers α, $M(w)\mathrm{d}w$, $\{\nu_i(\omega)\mathrm{d}\omega\}_i$, $\{\Theta_j(\xi)\mathrm{d}\xi\}_j$ respectively denote a real variable and measure-type variables over the set of positive semidefinite $w \in C^*(X^\star)$, the set of states ω and the set of Hermitian linear functionals ξ.

Notice that integration over the set of states (or functionals, or positive semidefinite elements) of a C^*-algebra might not be well defined if the latter is infinite dimensional. This lack of mathematical rigor is all right: we are not aiming to *prove* optimality conditions, just to guess their form.

For simplicity, for each i, we next define a new multiplier μ_i, of the form

$$\mu_i := \int_{\substack{\omega(g_i(X^\star))=0,\\ \omega \text{ state}}} \nu_i(\omega)\mathrm{d}\omega\,\omega. \tag{A.3}$$

We do likewise with the integration over w, i.e., we define

$$M := \int_{\substack{\psi^\star(w)=0,\\ w\geq 0}} M(w)\mathrm{d}w\,w. \tag{A.4}$$

By (A.3), (A.4), it follows that, for each i, μ_i is a positive linear functional and that M is positive semidefinite. Moreover,

$$\begin{aligned}\mu_i(g_i(X^\star)) = 0, \quad i = 1,\ldots,m, \\ \psi^\star(M) = 0.\end{aligned} \tag{A.5}$$

These conditions are a non-commutative analog of *complementary slackness* [20].

Analogously, we absorb the integrals over ξ by defining the Hermitian linear functionals

$$\lambda_j := \int \Theta_j(\xi) \mathrm{d}\xi \, \xi. \tag{A.6}$$

Substituting in (A.2), this expression is simplified to:

$$\mathcal{L} = \psi(f(\bar{X})) - \psi(M) + \alpha(1 - \psi(1)) - \sum_i \mu_i(g_i(\bar{X})) - \sum_j \lambda_j(h_j(\bar{X})). \tag{A.7}$$

We next arrive at candidate first-order optimality conditions for Problem (2.5) by varying the problem state variable ψ and the operator variables $\bar{X}$ in Eq. (A.7), all the while imposing the stationarity of $\mathcal{L}$. That way, we will obtain two sets of independent constraints: the state optimality conditions (3.2) and strong ncKKT (3.25).

1. State optimality conditions

Varying the state $\psi : C^*(X^\star) \to \mathbb{C}$ in (A.7) from the optimal $\psi^\star$ to $\psi^\star + \delta\psi$ leads to the condition

$$f(X^\star) - \alpha\mathbb{I} = M \geq 0. \tag{A.8}$$

On the other hand, complementary slackness implies that $\psi^\star(M) = 0$, i.e., the optimal state $\psi^\star$ is a ground state of $f(X^\star) - \alpha$ and $E_0(f(X^\star) - \alpha) = 0$. From $\psi^\star(f(X^\star)) = p^\star$, we have that $\alpha = p^\star$.

Call $H = f(X^\star)$. By Proposition 3.5, we have

$$\psi^\star(H\bullet) = \psi^\star(\bullet H) = p^\star \psi^\star(\bullet). \tag{A.9}$$

This implies that

$$\psi^\star([H, q]) = 0, \quad \forall q \in C^*(X^\star). \tag{A.10}$$

In addition, assuming (w.l.o.g.[8]) that $\psi^\star$ is cyclic, the condition $f(X^\star) - p^\star \geq 0$ is equivalent to

$$\psi^\star(q^*(f(X^\star) - p^\star)q) \geq 0, \quad \forall q \in C^*(X^\star). \tag{A.11}$$

Due to (A.9), the second term of the left-hand side of the above equation can be written as $\frac{1}{2}\psi^\star(\{f(X^\star), q^*q\})$. In addition, by continuity it is enough to demand Eqs. (A.10) and (A.11) to hold for $q \in \mathcal{A}(X^\star)$. Putting everything together, we arrive at Eq. (3.1).

2. Strong ncKKT

We next derive the strong ncKKT conditions. Let $p \in C^*(X^\star)^n$ be a tuple of Hermitian operators, and let $\epsilon \in \mathbb{R}^+$. Setting $\bar{X} = X^\star + \epsilon p$ in Eq. (A.7) and demanding stationarity of $\mathcal{L}$ at order ϵ, we arrive at the condition:

$$\psi^\star\left(\nabla_x f(p)\Big|_{x=X^\star}\right) - \sum_i \mu_i\left(\nabla_x g_i(p)\Big|_{x=X^\star}\right) - \sum_j \lambda_j\left(\nabla_x h_j(p)\Big|_{x=X^\star}\right) = 0, \quad \forall p \in C^*(X^\star)^n. \tag{A.12}$$

Again by continuity, it is sufficient to demand the equation above to hold for all $p \in \mathcal{A}(X^\star)^n$.

Together with complementary slackness (A.5), we arrive at the definition of strong ncKKT in Section 3.

---

[8] If $\psi^\star$ is not cyclic, then we can apply the GNS construction [40, 41] and find another solution of Problem (2.5) with the same moments, such that the new state is cyclic.

Appendix B: Re-derivation of the first-order optimality conditions of [1] through Theorem 4.1

Consider the following problem:

$$\begin{aligned}\inf_{\mathcal{H},\tau,E} \tau &\left(\sum_{x,y,a,b} W(a,b,x,y)E_{a|x}E_{b|y}\right)\\ \text{s.t. } &\tau \text{ tracial state},\\ &E_{a|x} - E_{a|x}^2 = 0, \quad \forall a,x\\ &\sum_a E_{a|x} - 1 = 0, \quad \forall x,\end{aligned} \tag{B.1}$$

where $x = 1,\dots,n$, $a = 1,\dots,d$. Here, by tracial state, we mean a state τ with the property $\tau(ab) = \tau(ba)$, for all $a, b \in B(\mathcal{H})$. This problem is considered in [1] in the context of quantum non-locality. As shown in [91], optimizations over tracial states can also be relaxed through hierarchies of SDPs. Since the associated quadratic module is Archimedean, such hierarchies converge and optimal solutions for (B.1) exist. Moreover, the optimal state can be assumed faithful, i.e., $\tau(aa^*) = 0$ iff $a = 0$, for $a \in B(\mathcal{H})$.

Let $(\mathcal{H}^\star, \tau^\star, E^\star)$ be one of those solutions. By Theorem 4.1, it follows that, for any $k \in \mathbb{N}$, there exist linear functionals $\{\lambda^k_{a|x} : \mathbb{C}\langle e\rangle \to \mathbb{C}\}$, $\{\lambda^k_x : \mathbb{C}\langle e\rangle \to \mathbb{C}\}$ such that

$$\begin{aligned}&\tau^\star\left(\sum_{x,y,a,b} W(a,b,x,y)(p_{a|x}(E^\star)E^\star_{b|y} + E^\star_{a|x}p_{b|y}(E^\star))\right)\\ &= \sum_{a,x}\lambda^k_{a|x}\left(p_{a|x} - \{p_{a|x}, E_{a|x}\}\right) + \sum_x \lambda^k_x\left(\sum_a p_{a|x}\right),\end{aligned} \tag{B.2}$$

for any tuple of polynomials $p = (p_{a|x})_{a|x}$ of degree $2(k-1)$.

Now, for some $z \in \{1,\dots,n\}$, $s \in \mathbb{C}\langle e\rangle$, choose $p_{a|z} = i[s, e_{a|z}]$, for all a and $p_{a|x} = 0$, for all $x \neq z$. Substituting in (B.2), we have that

$$\tau^\star\left(i\left[\sum_{y,a,b} W(a,b,z,y)[s(E^\star), E^\star_{a|z}]E^\star_{b|y} + \sum_{x,a,b} W(a,b,x,z)E^\star_{a|x}[s(E^\star), E^\star_{b|z}]\right]\right) = 0. \tag{B.3}$$

Invoking the tracial properties of the state, this can be rewritten as $\tau^\star\left(s(E^\star)q_z(E^\star)\right) = 0$, with

$$q_z(E^\star) := \sum_{y,a,b} W(a,b,z,y)[E^\star_{a|z}, E^\star_{b|y}] + \sum_{x,a,b} W(a,b,x,z)[E^\star_{b|z}, E^\star_{a|x}]. \tag{B.4}$$

For $k \geq 2$, we can choose $s = q_z^*$, and thus $\tau^\star\left(q_z(E^\star)^* q_z(E^\star)\right) = 0$. Since by assumption the state $\tau^\star$ is faithful, we have that

$$q_z(E^\star) = 0, \tag{B.5}$$

for $z = 1,\dots,n$. This is the first-order optimality optimality condition derived in [1, Prop. 7.1].

Appendix C: NPO for quantum nonlocality

What follows is the original NPO formulation to compute the maximum quantum value of a Bell functional. We have added the partial operator KKT conditions derived in Section 6.2,

together with the state optimality conditions (3.2).

$$
\begin{aligned}
c^\star = &\min \sigma\left(\frac{1}{2}\sum_{a,b,x,y} C(a,b,x,y)\{E_{a|x}, F_{b|y}\}\right)\\
\text{s.t. } &\sigma(ss^*) \geq 0, \quad \forall s \in \mathbb{C}\langle x\rangle,\\
&\sigma\Big(s\big((E_{a|x})^2 - E_{a|x}\big)s'\Big) = 0, \quad \forall a, x, \forall s, s' \in \mathbb{C}\langle x\rangle\\
&\sigma\Big(s\Big(\sum_a E_{a|x} - 1\Big)s'\Big) = 0, \quad \forall x, \forall s, s' \in \mathbb{C}\langle x\rangle\\
&\sigma\Big(s\big((F_{b|y})^2 - F_{b|y}\big)s'\Big) = 0, \quad \forall b, y, \forall s, s' \in \mathbb{C}\langle x\rangle\\
&\sigma\Big(s\Big(\sum_b F_{b|y} - 1\Big)s'\Big) = 0, \quad \forall y, \forall s, s' \in \mathbb{C}\langle x\rangle\\
&\sigma\big(s[E_{a|x}, F_{b|y}]s'\big) = 0, \quad \forall a, b, x, y, \forall s, s' \in \mathbb{C}\langle x\rangle\\
&\mu^A_{a|x}(ss^*) \geq 0, \quad \forall a, x, \forall s \in \mathcal{P}_E,\\
&\mu^A_{a|x}\Big(s\big((E_{a'|x'})^2 - E_{a'|x'}\big)s'\Big) = 0, \quad \forall a, a', x, x', \forall s, s' \in \mathcal{P}_E\\
&\mu^A_{a|x}\Big(s\Big(\sum_{a'} E_{a'|x'} - 1\Big)s'\Big) = 0, \quad \forall a, x, x', \forall s, s' \in \mathcal{P}_E,\\
&\lambda^A_x\Big(s\big((E_{a|x'})^2 - E_{a|x'}\big)s'\Big) = 0, \quad \forall a, x, x', \forall s, s' \in \mathcal{P}_E\\
&\lambda^A_x\Big(s\Big(\sum_a E_{a|x} - 1\Big)s'\Big) = 0, \quad \forall x, \forall s, s' \in \mathcal{P}_E,\\
&\mu^A_{a|x}(E_{a|x}) = 0, \quad \forall a, x,\\
&\sigma\left(\frac{1}{2}\sum_{b,y} C(a,b,x,y)\{p, F_{b|y}\}\right) = \mu^A_{a|x}(p) + \lambda^A_x(p), \quad \forall a, x, \forall p \in \mathcal{P}_E.\\
&\mu^B_{b|y}(ss^*) \geq 0, \quad \forall b, y, \forall s \in \mathcal{P}_F\\
&\mu^B_{b|y}\Big(s\big(F_{b'|y'})^2 - F_{b'|y'}\big)s'\Big) = 0, \quad \forall b, b', y, y', \forall s, s' \in \mathcal{P}_F\\
&\mu^B_{b|y}\Big(s\Big(\sum_{b'} E_{b'|y'} - 1\Big)s'\Big) = 0, \quad \forall b, y, y', \forall s, s' \in \mathcal{P}_F\\
&\lambda^B_y\Big(s\big(F_{b|y'})^2 - F_{b|y'}\big)s'\Big) = 0, \quad \forall b, y, y', \forall s, s' \in \mathcal{P}_F\\
&\lambda^B_y\Big(s\Big(\sum_b F_{b|y} - 1\Big)s'\Big) = 0, \quad \forall y, \forall s, s' \in \mathcal{P}_F\\
&\mu^B_{b|y}(F_{b|y}) = 0, \quad \forall b, y,\\
&\sigma\left(\frac{1}{2}\sum_{a,x} C(a,b,x,y)\{E_{a|x}, p\}\right) = \mu^B_{b|y}(p) + \lambda^B_y(p), \quad \forall b, y, \forall p \in \mathcal{P}_F,\\
&\sigma\left(\left[\frac{1}{2}\sum_{a,b,x,y} C(a,b,x,y)\{E_{a|x}, F_{b|y}\}, p\right]\right) = 0,\\
&\sigma\left(p^*\left(\frac{1}{2}\sum_{a,b,x,y} C(a,b,x,y)\{E_{a|x}, F_{b|y}\}\right)p - \frac{1}{2}\left\{\frac{1}{2}\sum_{a,b,x,y} C(a,b,x,y)\{E_{a|x}, F_{b|y}\}, p^*p\right\}\right) \geq 0, \quad \forall p \in \mathbb{C}\langle x\rangle.
\end{aligned}
$$