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, Universidad de Valladolid, 47011 Valladolid, Spain*

[2] *Faculty of Mathematics and Physics, University of Ljubljana & 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] *MTA Atomki Lendület Quantum Correlations Research Group, 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 phrasing the general NPO problem in Lagrangian form, we heuristically derive, via small variations on the problem variables, state and operator optimality conditions, both of which can be enforced by adding new positive semidefinite constraints to the SDP hierarchies. State optimality conditions are satisfied by all Archimedean (that is, bounded) NPO problems, and 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 non-commutative operator optimality holds for all Archimedean NPO problems; stronger versions require the problem constraints to satisfy some qualification criterion, just like in the classical case. 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.

CONTENTS

I. 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 [1, 2], the electronic energy of atoms and molecules [3–5], the ground state energies of spin systems [6–9], or the ground state behavior of fermions at finite density [10].

From the work of Pironio *et al.* [11], (see also Refs. [1, 2, 12–14]), we know that all NPO problems involving bounded operators can be solved through hierarchies of semidefinite programs (SDPs) [15, 16] 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, while 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 [17, 18]. 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 [11]. Optimality constraints on the solutions of specific NPO problems have already been considered in the literature [19, 20], but not in the context of deriving new or improving existing numerical methods. Motivation aside, our contribution differs from these earlier works 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 Section III B, 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 bounded NPO problems, justifying the corresponding operator optimality conditions (the non-commutative Karush-Kuhn-Tucker conditions, or ncKKT) requires more work. In this context, we show that: (a) essential ncKKT, the most relaxed variant of the operator optimality conditions, holds in all Archimedean NPO problems; (b) normed ncKKT, a substantially strengthened version, holds in all NPO problems that satisfy the non-commutative analog of the Mangasarian–Fromovitz conditions [21]; (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, we also provide necessary conditions that guarantee that either normed or strong ncKKT 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.

The structure of this paper is as follows: In Sec. II, we recall the class of non-commutative optimization problems that we consider in this paper, and present their corresponding hierarchies of SDP relaxations. In Sec. III, using heuristic arguments, 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 Sec. III B. Sufficient criteria for the validity of the different forms of operator optimality are presented in Sec. IV. Sec. V investigates when it is legitimate to enforce the new optimality conditions partially. In Sec. VI, 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 (Sec. VI A) and the maximum violation of bipartite Bell inequalities (Sec. VI B). We then present our conclusions.

While conducting this research, we found that Fawzi *et al.* [22] had independently arrived at the state optimality conditions (42). In their interesting preprint, 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 VI A of this paper.

II. NPO PROBLEMS

In this work, we will be interested in polynomials of n non-commuting Hermitian variables $x = (x_1, \dots, x_n)$. Any such polynomial $p(x)$ is called symmetric or Hermitian if $p(x) = p(x)^*$. A non-commutative polynomial optimization (NPO) problem [11, 14] is the natural analog of a polynomial optimization problem [23–25].

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

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

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

Computing the maximal quantum violation of a Bell inequality [26–28] or the energy of a many-body quantum system [29] are examples of NPO problems.

Call $\mathcal{P}$ the space of all polynomials of the non-commuting variables $x_1, \dots, x_n$, i.e., the unital $*$-algebra freely generated by $x_1, \dots, x_n$. Problem (1) can be relaxed to

$$\begin{aligned} p^\star := &\min_{\sigma:\mathcal{P}\to\mathbb{C}} \sigma(f) \\ \text{s.t. } & \sigma(1) = 1,\ \sigma(pp^*) \geq 0, \quad \forall p \in \mathcal{P}, \\ & \sigma(pg_ip^*) \geq 0, \quad \forall p \in \mathcal{P},\ i = 1, \dots, m, \\ & \sigma(ph_jq) = 0, \quad \forall p, q \in \mathcal{P},\ j = 1, \dots, m', \end{aligned} \tag{2}$$

Clearly, Problem (2) is a relaxation of (1). In the presence of a boundedness assumption (such as the Archimedean condition, see Definition 3) the two problems are equivalent as a consequence of the Gelfand–Naimark–Segal (GNS) construction [30, 31]: given a linear functional $\sigma^\star$ minimizing (2), the GNS construction builds a Hilbert space $\mathcal{H}^\star$, bounded operators $X^\star$ (this is where the Archimedean condition enters) satisfying the constraints of Problem (1), and a unit vector $\phi^\star \in \mathcal{H}^\star$ such that

$$\langle \phi^\star | p(X^\star) | \phi^\star \rangle = \sigma^\star(p(x)), \quad \forall p \in \mathcal{P}. \tag{3}$$

Defining $\psi^\star(\bullet) := \langle \phi^\star | \bullet | \phi^\star \rangle$, we thus have that $(\mathcal{H}^\star, X^\star, \psi^\star)$ is a solution of Problem (1) with the same objective value as $\sigma^\star$. In view of this observation, we call a solution $\sigma^\star$ of Problem (2) *bounded* if its GNS construction generates bounded operators $X_1^\star, \dots, X_n^\star$.

Very conveniently, Problem (2) can be relaxed through hierarchies of semidefinite programs (SDP) [1, 2, 11]. Let $\mathcal{P}_k$ be the space of polynomials on x of degree at most k. A straightforward relaxation of Problem (2) is thus:

$$\begin{aligned} p^k := &\min_{\sigma^k:\mathcal{P}_{2k}\to\mathbb{C}} \sigma^k(f) \\ \text{s.t. } & \sigma^k(1) = 1,\ \sigma^k(pp^*) \geq 0, \quad \forall p \in \mathcal{P}_k, \\ & \sigma^k(s^* g_i s) \geq 0, \quad \forall s \in \mathcal{P},\ \deg(s) \leq k - \left\lceil \frac{\deg(g_i)}{2} \right\rceil,\ i = 1, \dots, m, \\ & \sigma^k(s h_j s') = 0, \quad \forall s, s' \in \mathcal{P},\ \deg(s) + \deg(s') \leq 2k - \deg(h_j),\ j = 1, \dots, m'. \end{aligned} \tag{4}$$

The relaxation (4) can be cast as a semidefinite program [15, 32] with $|\mathcal{P}_{2k}|$ 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 k ($k - \left\lceil \frac{\deg(g_l)}{2} \right\rceil$). Then, the matrices

$$\begin{aligned} \left(M^k(\sigma^k)\right)_{ab} &:= \sigma^k(o_a^* o_b), \\ \left(M_l^k(\sigma^k)\right)_{ab} &:= \sigma^k((o_a^l)^* g_l o_b^l), \end{aligned} \tag{5}$$

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 [11]. Enforcing the first two lines of constraints in Problem (4) 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 that the identity holds for all monomials s, s' satisfying the degree constraint.

The dual problem of (4) is of the form:

$$\begin{aligned} q^k := &\max \theta \\ \text{s.t. } &\exists\{s_j\}_j \subset \mathcal{P}_k,\ \{s_{il}\}_{il} \subset \mathcal{P}:\ 2\deg(s_{il}) \leq 2k - \deg(g_i), \\ &\exists\{s_{jl}^+\}_{jl},\ \{s_{jl}^-\}_{jl} \subset \mathcal{P}:\ \deg(s_{jl}^+) + \deg(s_{jl}^-) \leq 2k - \deg(h_j), \\ &f - \theta = \sum_l s_l^* s_l + \sum_{il} s_{il}^* g_i s_{il} + \sum_{jl} s_{jl}^+ h_j (s_{jl}^-)^* + \sum_{jl} s_{jl}^- h_j (s_{jl}^+)^*. \end{aligned} \tag{6}$$

This problem is also an SDP, and the right-hand side of the last line is a *weighted sum of squares (SOS) decomposition* [12].

Definition 2. *Given Problem* (2) *and a Hermitian polynomial* p*, we say that* p *admits an SOS decomposition if there exist polynomials* $\{s_j\}_j, \{s_{il}\}_{il}, \{s_{jl}^+\}_{jl}, \{s_{jl}^-\}_{jl}$ *such that*

$$p = \sum_j s_j^* s_j + \sum_{il} s_{il}^* g_i s_{il} + \sum_{jl} s_{jl}^+ h_j (s_{jl}^-)^* + \sum_{jl} s_{jl}^- h_j (s_{jl}^+)^*. \tag{7}$$

If p admits an SOS decomposition and the tuple of operators $\bar{X} \in B(\mathcal{H})^{\times n}$ satisfies the constraints of Problem (1), then $p(\bar{X})$ must be a positive semidefinite operator. Problem (6) can thus be interpreted as finding the maximum real number θ such that $f - \theta$ is an SOS (under some restrictions on the degrees of the polynomials in the decomposition).

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 $p^\star$ of Problem (2). Clearly, $p^1 \leq p^2 \leq \cdots \leq p^\star$, and similarly for the q's. It is sufficient that (2) satisfies the *Archimedean property* for these hierarchies to be complete, in the sense that $\lim_{n\to\infty} p^n = \lim_{n\to\infty} q^n = p^\star$ [11].

Definition 3 (The Archimedean property)**.** *Problem* (1) *or Problem* (2) *is* Archimedean *if there exists* $K \in \mathbb{R}^+$*, such that the polynomial*

$$K - \sum_i x_i^2 \tag{8}$$

admits an SOS decomposition.

The Archimedean property implies that all feasible operators in Problem (2) must be bounded. In particular, under the Archimedean assumption, Problems (1) and (2) are equivalent. Conversely, if the feasible set is bounded, a relation of the form of (8) can be added to the inequality constraints without changing the problem.

III. FIRST-ORDER OPTIMALITY CONDITIONS

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

$$\begin{aligned} p^\star := &\min f(x) \\ \text{s.t. } &g_i(x) \geq 0, \quad i = 1, \ldots, m, \\ &h_j(x) = 0, \quad j = 1, \ldots, m', \end{aligned} \tag{9}$$

where $x = (x_1, \ldots, x_n)$ is a vector of real variables, and f, g_i, h_j are real-valued functions thereof. Given a function $s(x)$, call $\partial_x s$ its gradient, i.e., $\partial_x s = \left(\frac{\partial s(x)}{\partial x_1}, \ldots, \frac{\partial s(x)}{\partial x_n}\right)$. In this commutative scenario, the Karush–Kuhn–Tucker (KKT) conditions read:

$$\begin{aligned} &\exists\{\mu_i\}_i \subset \mathbb{R}^+,\ \{\lambda_j\}_j \subset \mathbb{R}, \\ \text{such that } &\partial_x f\Big|_{x=x^\star} = \sum_{i\in\mathbb{A}(x^\star)} \mu_i \partial_x g_i\Big|_{x=x^\star} + \sum_j \lambda_j \partial_x h_j\Big|_{x=x^\star}, \\ &\mu_i g_i(x^\star) = 0, \quad \forall i \in \mathbb{A}(x^\star), \end{aligned} \tag{10}$$

where $x^\star$ is a solution of (9). Here, the set of *active constraints* $\mathbb{A}(x)$ denotes the set of indices $i \in \{1, \ldots, m\}$ for which $g_i(x) = 0$.

The optimal solutions of many classical optimization problems are known to satisfy the KKT conditions, also known as first-order optimality conditions [17, 18]. This has led some authors to enforce these conditions implicitly to numerically solve polynomial optimization problems [33, 34].

The KKT conditions can be non-rigorously derived by considering infinitesimal variations over $x^\star$ of the Lagrangian functional

$$\mathcal{L}_c = f(x) - \sum_{i \in \mathbb{A}(x^\star)} \mu_i g_i(x) - \sum_j \lambda_j h_j(x), \tag{11}$$

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 (10). To find them, we start by writing a Lagrangian for Problem (1).

Let $(\mathcal{H}^\star, X^\star, \psi^\star)$ be any bounded solution of Problem (1). 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 $\mathcal{A}(X^\star)$, 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 $\mathcal{A}(X^\star)$.

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 $\omega : \mathcal{A}(X^\star) \to \mathbb{C}$. 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 $\xi : \mathcal{A}(X^\star) \to \mathbb{C}$, 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 (11) is:

$$\begin{aligned}\mathcal{L} = {} & \psi(f(\bar{X})) - \int_{\substack{\psi^\star(w)=0,\\ w \geq 0}} M(w) \mathrm{d}w \psi(w) + \alpha(1 - \psi(1)) \\ & - \sum_i \int_{\substack{\omega(g_i(X^\star))=0,\\ \omega \text{ state}}} \nu_i(\omega) \mathrm{d}\omega\, \omega(g_i(\bar{X})) - \sum_j \int \Theta_j(\xi) \mathrm{d}\xi\, \xi(h_j(\bar{X})),\end{aligned} \tag{12}$$

where the Hermitian operator variables $\bar{X} = (\bar{X}_1, \ldots, \bar{X}_n)$ are elements of $\mathcal{A}(X^\star)$ and $\psi : \mathcal{A}(X^\star) \to \mathbb{C}$ 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 $w \in \mathcal{A}(X^\star)$, 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 fine for the time being: we are not aiming to *prove* optimality conditions yet, just to guess their form.

For simplicity, for each i, we next define a new multiplier μ_i, of the form

$$\mu_i := \int_{\substack{\omega(g_i(X^\star))=0,\\ \omega \text{ state}}} \nu_i(\omega) \mathrm{d}\omega\, \omega. \tag{13}$$

We do likewise with the integration over w, i.e., we define

$$M := \int_{\substack{\psi^\star(w)=0,\\ w \geq 0}} M(w) \mathrm{d}w\, w. \tag{14}$$

By (13), (14), it follows that, for each i, μ_i is a positive linear functional and that M is positive semidefinite. Moreover,

$$\begin{aligned}&\mu_i(g_i(X^\star)) = 0, \quad i = 1, \ldots, m, \\ &\psi^\star(M) = 0.\end{aligned} \tag{15}$$

These conditions are a non-commutative analog of *complementary slackness* [21].

Analogously, we absorb the integrals over ξ by defining the Hermitian linear functionals

$$\lambda_j := \int \Theta_j(\xi) \mathrm{d}\xi\, \xi. \tag{16}$$

Substituting in (12), this expression is simplified to:

$$\mathcal{L} = \psi(f(\bar{X})) - \psi(M) + \alpha(1 - \psi(1)) - \sum_i \mu_i(g_i(\bar{X})) - \sum_j \lambda_j(h_j(\bar{X})). \tag{17}$$

We will arrive at candidate first-order optimality conditions for Problem (1) by varying the problem state variable ψ and the operator variables $\bar{X}$ in Eq. (17), all the while imposing the stationarity of $\mathcal{L}$. That way, we will obtain two sets of independent constraints: state optimality and operator optimality conditions. The latter will be dubbed *non-commutative KKT conditions*, or ncKKT.

A. Operator optimality conditions

To study how the Lagrangian (17) changes when we vary the operator variables of Problem (1), we must first propose an analog of gradients for non-commutative functions. Closely related notions have been studied in free function theory [35] and non-commutative real algebraic geometry [36].

Definition 4. *The gradient of a (non-commutative) polynomial $p(x)$ is a polynomial of the original non-commuting variables x and their 'variations' $\bar{x} = (\bar{x}_1, \ldots, \bar{x}_n)$, linear in $\bar{x}$. This polynomial, denoted as $\nabla_x p(\bar{x})$, is obtained from $p(x)$ by evaluating $p(x + \epsilon\bar{x})$ and keeping only the terms linear in ϵ. Informally:*

$$\nabla_x p(\bar{x}) := \lim_{\epsilon \to 0} \frac{p(x+\epsilon\bar{x}) - p(x)}{\epsilon}. \tag{18}$$

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

With this definition, it is easy to express the effect of a variation of the operator variables $\bar{X}$ in Eq. (17). Let $p \in \mathcal{A}(X^\star)^{\times n}$ be a tuple of Hermitian operators, and let $\epsilon \in \mathbb{R}^+$. Setting $\bar{X} = X^\star + \epsilon p$ in Eq. (17) and demanding stationarity of $\mathcal{L}$ at order ϵ, we arrive at the condition:

$$\psi^\star\left(\nabla_x f(p)\Big|_{x=X^\star}\right) - \sum_i \mu_i\left(\nabla_x g_i(p)\Big|_{x=X^\star}\right) - \sum_j \lambda_j\left(\nabla_x h_j(p)\Big|_{x=X^\star}\right) = 0, \quad \forall p \in \mathcal{A}(X^\star)^{\times n}. \tag{19}$$

Complementary slackness (15) and Eq. (19) are the basis of a straightforward generalization of the KKT conditions.

Definition 5 (Strong ncKKT)**.** *We say that the NPO Problem (1) satisfies the* strong non-commutative Karush–Kuhn–Tucker conditions *if, for any bounded solution $(\mathcal{H}^\star, X^\star, \psi^\star)$ of Problem (1), 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*

$$\mu_i(g_i(X^\star)) = 0, \quad i = 1, \ldots, m, \tag{20a}$$

$$\psi^\star\left(\nabla_x f(p)\Big|_{x=X^\star}\right) - \sum_i \mu_i\left(\nabla_x g_i(p)\Big|_{x=X^\star}\right) - \sum_j \lambda_j\left(\nabla_x h_j(p)\Big|_{x=X^\star}\right) = 0, \quad \forall p \in \mathcal{A}(X^\star)^{\times n}. \tag{20b}$$

The problem with the above definition is that we do not know of any way of enforcing these conditions fully in any NPO problem, unless the solution $(\mathcal{H}^\star, X^\star, \psi^\star)$ is known to satisfy further operator relations, or average-value constraints of the form $\psi^\star(p(X^\star)) \geq 0$, for some $p \in \mathcal{P}$. Replacing $\mathcal{A}(X^\star)$ by $\mathcal{P}$ in the definition above, we arrive at a much more convenient set of constraints, phrased in terms of the solutions of Problem (2), rather than Problem (1).

Definition 6 (Weak ncKKT)**.** *We say that the NPO Problem (2) satisfies the* weak non-commutative Karush–Kuhn–Tucker conditions *if, for any solution $\sigma^\star$ of Problem (2), there exist positive linear functionals $\{\mu_i : \mathcal{P} \to \mathbb{C}\}_{i=1}^m$, compatible with the constraints $\{g_i(x) \geq 0\}_i \cup \{h_j(x) = 0\}_j$, and Hermitian linear functionals $\{\lambda_j : \mathcal{P} \to \mathbb{C}\}_{j=1}^{m'}$, compatible with the constraints $\{h_j(x) = 0\}_j$, such that*

$$\mu_i(g_i) = 0, \quad i = 1, \ldots, m, \tag{21a}$$

$$\sigma^\star\left(\nabla_x f(p)\right) - \sum_i \mu_i\left(\nabla_x g_i(p)\right) - \sum_j \lambda_j\left(\nabla_x h_j(p)\right) = 0, \quad \forall p \in \mathcal{P}^{\times n}. \tag{21b}$$

Above, we introduced notation that we will use for simplicity. First, we will call a Hermitian linear functional $\mu : \mathcal{P} \to \mathbb{C}$ *positive* if $\mu(pp^*) \geq 0$ for all $p \in \mathcal{P}$. Second, if for some Hermitian $g, h \in \mathcal{P}$, the functional μ satisfies the condition

$$\mu(pgp^*) \geq 0, \quad \forall p \in \mathcal{P}, \tag{22}$$

or

$$\mu(s^+ h s^-) = 0, \quad \forall s^+, s^- \in \mathcal{P}, \tag{23}$$

we will respectively write that μ is *compatible with the constraint* $g(x) \geq 0$ *or* $h(x) = 0$.

If we were guaranteed that the weak ncKKT conditions held for Problem (2), then we could add some further constraints to our SDP relaxation (4), namely:

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

$$
\begin{aligned}
\exists\, \mu_i^k : \mathcal{P}_{2k} \to \mathbb{C}, \quad & \mu_i^k(pp^*) \geq 0, && \forall p \in \mathcal{P}_k, \\
& \mu_i^k(pg_lp^*) \geq 0, && \forall p \in \mathcal{P},\ \deg(p) \leq k - \left\lceil \frac{\deg(g_l)}{2} \right\rceil,\ l = 1, \ldots, m, \\
& \mu_i^k(s^+ h_j s^-) = 0, && \forall s^+, s^- \in \mathcal{P},\ \deg(s^+) + \deg(s^-) \leq 2k - \deg(h_j),\ j = 1, \ldots, m', \\
& \mu_i^k(g_i) = 0,
\end{aligned} \tag{24a}
$$

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

$$
\begin{aligned}
\exists\, \lambda_j^k : \mathcal{P}_{2k} \to \mathbb{C}, \quad & \lambda_j^k(p + p^*) \in \mathbb{R}, && \forall p \in \mathcal{P}_{2k}, \\
& \lambda_j^k(s^+ h_l s^-) = 0, && \forall s^+, s^- \in \mathcal{P},\ \deg(s^+) + \deg(s^-) \leq 2k - \deg(h_j),\ l = 1, \ldots, m',
\end{aligned} \tag{24b}
$$

and

$$
\sigma^k(\nabla_x f(t(x))) - \sum_i \mu_i^k(\nabla_x g_i(t(x))) + \sum_j \lambda_j^k(\nabla_x h_j(t(x))) = 0,
$$
$$
\forall t \in \mathcal{P}^{\times n},\ \deg\left(\nabla_x f\right),\ \deg\left(\nabla_x g_i\right),\ \deg\left(\nabla_x h_j\right) \leq 2k - \deg(t), \tag{24c}
$$

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 t, $\deg(t) = \max_k \deg(t_k)$. Note that, in order to enforce the last condition, it is enough to consider tuples of monomials t with just one non-zero entry.

At this point, the reader might wonder if we lose something when we relax the strong variant of the ncKKT conditions. To answer this question, consider the following commutative problem:

$$
\begin{aligned}
&\min\ x \\
&\text{s.t. } x^2 = 0.
\end{aligned} \tag{25}
$$

The solution is $x^\star = 0$. The commutative KKT conditions would demand that there exists $\lambda \in \mathbb{R}$ such that

$$
0 = 2\lambda \times 0 = \lambda \frac{dx^2}{dx}\Big|_{x=0} = \frac{dx}{dx}\Big|_{x=0} = 1. \tag{26}
$$

This is obviously not the case. However, viewed as an NPO problem, Eq. (25) reads:

$$
\begin{aligned}
&\min\ \psi(X) \\
&\text{s.t. } X^2 = 0.
\end{aligned} \tag{27}
$$

The solution of the above is $\mathcal{H}^\star = \mathbb{C}$, $X^\star = 0$ and $\psi^\star = 1$ (the only state in dimension 1). Defining $\sigma^\star(p) := \psi^\star(p(X^\star))$, both weak and strong ncKKT conditions then imply:

$$
\begin{aligned}
\lambda(\{x, p\}) &= \sigma^\star(p) \quad \forall p \in \mathcal{P}. \\
\lambda(s^+ x^2 s^-) &= 0, \quad \forall s^+, s^- \in \mathcal{P}.
\end{aligned} \tag{28}
$$

The above is equivalent to

$$
\begin{aligned}
\lambda(x) &= \frac{1}{2}, \\
\lambda(x^n) &= 0, \quad \forall n \geq 2.
\end{aligned} \tag{29}
$$

These relations do not lead to any contradiction if we just require λ to be a linear functional (i.e., if we just demand weak ncKKT). They are contradictory, though, if we expect λ to be of the form $\lambda(p) = \tilde{\lambda}(p(X^\star))$. Problem (27) is thus an instance of an NPO problem that satisfies the weak variant of the ncKKT conditions, but not the strong one.

Can we strengthen the weak variant of ncKKT to rule out examples like Problem (27) in a manner that we can still enforce? A close examination reveals that Eq. (29) cannot hold if we demand λ to be a bounded functional under the seminorm [1]

$$
\|p\|_{\text{SOS}} := \inf\{\nu \in \mathbb{R}^+ : \nu^2 - pp^* \ \ \text{SOS}\}. \tag{30}
$$

Certainly, the condition $x^2 = 0$ implies that $\|x\|_{\text{SOS}} = 0$, and so $\lambda(x) = 0$, for all bounded functionals λ. Unsatisfactory relations such as (28) would thus be banned if we could somehow enforce that the linear functionals $\{\lambda_j\}_j$ appearing in Eq. (21b) were bounded. In this regard, the Jordan decomposition [37, Theorem 3.3.10] shows that any Hermitian linear functional of bounded norm can be expressed as the difference between two positive linear functionals. This observation suggests the following new definition.

---

[1] More precisely, we demand λ to be bounded under the norm $\|\|_{SOS}^*$ induced by $\|\|_{SOS}$, i.e., $\|\lambda\|_{SOS}^* := \sup\{|\lambda(p)| : p \in \mathcal{P}, \|p\|_{SOS} \leq 1\}$.

Definition 7 (Normed ncKKT)**.** *We say that the NPO Problem* (2) *satisfies the* normed ncKKT conditions *if, for any solution $\sigma^\star$ of Problem* (2)*, it holds that*

$$\forall i,j,\ \exists \text{ positive } \mu_i, \lambda_j^{\pm}: \mathcal{P}\to\mathbb{C}, \text{compatible with } \{g_k(x)\geq 0\}_k \cup \{h_l(x)=0\}_l, \tag{31a}$$

$$\text{such that } \mu_i(g_i)=0, \quad i=1,\ldots,m, \tag{31b}$$

$$\sigma^\star\left(\nabla_x f(t)\right) - \sum_i \mu_i\left(\nabla_x g_i(t)\right) - \sum_j \lambda_j\left(\nabla_x h_j(t)\right) = 0, \quad \forall t\in\mathcal{P}^{\times n}, \tag{31c}$$

$$\text{with } \lambda_j := \lambda_j^+ - \lambda_j^-, \quad j=1,\ldots,m'. \tag{31d}$$

From the discussion above, it is easy to see that Problem (27) does not satisfy these normed ncKKT conditions. Indeed, the requirement that the 2×2 matrices $M^{\pm}_{kl} := \lambda^{\pm}(p_k p_l^*)$, with $p_1 = 1, p_2 = x$, be positive semidefinite implies that the off-diagonal elements $\lambda^{\pm}(x)$ vanish, and so $\lambda(x) = 0$.

It is also clear that, to enforce the constraints (31), one just needs to add to Eq. (24), for each $j = 1,\ldots m'$, the positive semidefinite constraints:

$$\exists \lambda_j^{\pm,k}: \mathcal{P}_{2k}\to\mathbb{C}, \quad \lambda_j^{\pm,k}(pp^*)\geq 0, \qquad \forall p\in\mathcal{P}_k, \tag{32a}$$

$$\lambda_j^{\pm,k}(pg_lp^*)\geq 0, \qquad \forall p\in\mathcal{P},\ \deg(p)\leq k-\left\lceil\frac{\deg(g_l)}{2}\right\rceil,\ l=1,\ldots,m, \tag{32b}$$

$$\lambda_j^{\pm,k}(sh_ks')=0, \qquad \forall s,s'\in\mathcal{P},\ \deg(s)+\deg(s')\leq 2k-\deg(h_k), k=1,\ldots,m', \tag{32c}$$

$$\lambda_j^k = \lambda_j^{+,k}-\lambda_j^{-,k}. \tag{32d}$$

In spite of their name, the weak ncKKT conditions are not so weak that they always hold. Consider the following problem:

$$\begin{aligned} &\min\ \sigma(x)\\ &\text{s.t.}\ \ -x^2\geq 0. \end{aligned} \tag{33}$$

The solution is, obviously, $\mathcal{H}^\star = \mathbb{C}, X^\star = 0$ and $\sigma^\star = 1$. In this case, the weak ncKKT conditions would demand the existence of a positive linear functional μ, compatible with the constraint $-x^2 \geq 0$ and such that

$$-\mu(\{x,p\}) = \sigma^\star(p), \quad \forall p\in\mathcal{P}. \tag{34}$$

Taking $p = 1$, the relation above implies that $\mu(x) = -\frac{1}{2}$. However, $\mu(x^2) = 0$ (it is non-negative due to the positivity of μ and smaller than or equal to 0 due to the constraint $-x^2 \geq 0$). The 2×2 matrix

$$\begin{pmatrix} \mu(1\cdot 1) & \mu(x\cdot 1)\\ \mu(1\cdot x) & \mu(x\cdot x)\end{pmatrix} = \begin{pmatrix} \mu(1) & -\frac{1}{2}\\ -\frac{1}{2} & 0\end{pmatrix} \tag{35}$$

cannot be made positive semidefinite, no matter the choice of $\mu(1)$. Thus, said positive functional does not exist.

Hence, not all NPO problems satisfy the weak ncKKT conditions. However, in Section IV A we will see that all Archimedean NPO problems do satisfy a relaxed variant thereof, which we call the *essential* ncKKT conditions.

Definition 8 (Essential ncKKT)**.** *An NPO Problem* (2) *satisfies the* essential ncKKT conditions *if, for any minimizer $\sigma^\star$ of Problem* (2) *and for all $k\in\mathbb{N}$, $\{\epsilon_i\}_i\subset\mathbb{R}^+$ there exist Hermitian linear functionals $\mu_i^k:\mathcal{P}_{2k}\to\mathbb{C}$, $i=1,\ldots,m$, $\lambda_j^k:\mathcal{P}_{2k}\to\mathbb{C}$, $j=1,\ldots,m'$, such that*

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

$$\begin{aligned} &\mu_i^k(pp^*)+\epsilon_i\|p\|_2^2\geq 0, && \forall p\in\mathcal{P},\ \deg(p)\leq k,\\ &\mu_i^k(pg_lp^*)+\epsilon_i\|p\|_2^2\geq 0, && \forall p\in\mathcal{P},\ \deg(p)\leq k-\left\lceil\frac{\deg(g_l)}{2}\right\rceil,\ l=1,\ldots,m,\\ &\mu_i^k(s^+h_js^-)=0, && \forall s^+,s^-\in\mathcal{P},\ \deg(s^+)+\deg(s^-)\leq 2k-\deg(h_l),\ j=1,\ldots,m',\\ &\mu_i^k(sg_i)=\mu_i^k(g_is)=0, && \forall s\in\mathcal{P},\ \deg(s)\leq 2k-\deg(g_i), \end{aligned} \tag{36a}$$

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

$$\lambda_j^k(s^+h_ls^-)=0, \qquad \forall s^+,s^-\in\mathcal{P},\ \deg(s^+)+\deg(s^-)\leq 2k-\deg(h_l),\ l=1,\ldots,m' \tag{36b}$$

and

$$\sigma^\star\left(\nabla_x f(t(x))\right) - \sum_i \mu_i^k\left(\nabla_x g_i(t(x))\right) - \sum_j \lambda_j^k\left(\nabla_x h_j(t(x))\right) = 0,$$
$$\forall t \in \mathcal{P}^{\times n}, \deg\left(\nabla_x f\right), \deg\left(\nabla_x g_i\right), \deg\left(\nabla_x h_j\right) \le 2k - \deg(t), \quad \text{(36c)}$$

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

Remark 9. Rather than demanding $\mu_i(g_i) = 0$ (as we did in Eq. (24a)), in Eq. (36a) we demand the stronger condition $\mu_i(sg_i) = \mu_i(g_i s) = 0$ for all $s \in \mathcal{P}$. The explanation is as follows: if the 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 need to impose the condition by hand.

Remark 10. Enforcing the first two lines of Eq. (36a) amounts to demanding that the k^{th}-order moment and localizing matrices (see Eq. (5) for a definition) $M^k(\mu_i), M_l^k(\mu_i)$ satisfy

$$M^k(\mu_i) + \epsilon_i \mathbb{I} \ge 0, \; M_l^k(\mu_i) + \epsilon_i \mathbb{I} \ge 0, \quad l = 1, \ldots, m. \tag{37}$$

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 (4).

To summarize: all Archimedean NPO problems satisfy the essential ncKKT conditions (Section IV A), but only some of them satisfy any of the three other variants. The inclusion relations between the sets of Archimedean problems satisfying either condition are thus **All** = **Essential** $\supsetneq$ **Weak** $\supsetneq$ **Normed** $\supseteq$ **Strong**.

This situation is analogous to the classical case, where the KKT conditions do not hold generally, but can be shown to hold if they fulfill a *constraint qualification* (see [21, Chapter 12]). In Sections IV B, IV C we will see that, in fact, some such classical criteria can be generalized to the non-commutative realm, so as to ensure that either the normed or the strong form of ncKKT holds in a given NPO problem.

B. State optimality conditions

Varying the state $\psi : \mathcal{A}(X^\star) \to \mathbb{C}$ in (17) from the optimal $\psi^\star$ to $\psi^\star + \delta\psi$ leads to the condition

$$f(X^\star) - \alpha\mathbb{I} = M \ge 0. \tag{38}$$

On the other hand, complementary slackness implies that $\psi^\star(M) = 0$. This suggests that the optimal state $\psi^\star$ has support on the space of eigenvectors of $f(X^\star)$ with minimum eigenvalue $\alpha = p^\star$. In that case, calling $H = f(X^\star)$, we have

$$\psi^\star(H\bullet) = \psi^\star(\bullet H) = p^\star\psi^\star(\bullet). \tag{39}$$

This implies that $\psi^\star([H, \bullet]) = 0$. In addition, assuming (w.l.o.g.[2]) that $\psi^\star$ is cyclic, the condition $f(X^\star) - p^\star \ge 0$ is equivalent to

$$\psi^\star(q^*(f(X^\star) - p^\star)q) \ge 0, \quad \forall q \in \mathcal{A}(X^\star). \tag{40}$$

Due to (39), the second term of the left-hand side of the above equation can be written as $\frac{1}{2}\psi^\star(\{f(X^\star), q^*q\})$. Putting everything together, we have that, for variations of the form $\delta\psi(\bullet) = \psi(q^* \bullet q) - \psi^\star(q^*q)\psi(\bullet)$, the optimality of $\psi^\star$ implies

$$\psi^\star([f(X^\star), p]) = 0, \quad \forall p \in \mathcal{A}(X^\star), \tag{41a}$$
$$\psi^\star\left(p^* f(X^\star) p - \frac{1}{2}\{f(X^\star), p^*p\}\right) \ge 0, \quad \forall p \in A(X^\star). \tag{41b}$$

It is convenient to express the above constraint in terms of an abstract state $\sigma^\star$, the solution of Problem (2), which acts on polynomials of x rather than on the elements of $\mathcal{A}(X^\star)$. Thus, we arrive at the *state optimality conditions*:

$$\sigma^\star([f, p]) = 0, \quad \forall p \in \mathcal{P}, \tag{42a}$$
$$\sigma^\star\left(p^* f p - \frac{1}{2}\{f, p^*p\}\right) \ge 0, \quad \forall p \in \mathcal{P}. \tag{42b}$$

---

[2] If $\psi^\star$ is not cyclic, then we can apply the GNS construction [30, 31] and find another solution of Problem (1) with the same moments, such that the new state is cyclic.

We obtained conditions (42) through a heuristic argument. However, they can be rigorously shown to hold. Indeed, let Problem (2) be Archimedean and let $\sigma^\star$ be one of its solutions. Then we can apply the GNS construction (as in [11]) to obtain a solution $(\mathcal{H}^\star, X^\star, \psi^\star)$ of Problem (1) such that $\sigma^\star(p) = \psi^\star(p(X^\star))$ holds for all $p \in \mathcal{P}$. Moreover, the state $\psi^\star$ can be chosen to be normal, i.e., there exists a vector $|\phi^\star\rangle \in \mathcal{H}^\star$ such that $\psi^\star(\bullet) = \langle\phi^\star| \bullet |\phi^\star\rangle$. Now, $|\phi^\star\rangle$ must be an eigenstate of the operator $f(X^\star)$ with eigenvalue equal to the bottom of the spectrum of $f(X^\star)$. Otherwise, there would exist a unit vector $|\phi\rangle \in \mathcal{H}^\star$ with $\langle\phi|f(X^\star)|\phi\rangle < \langle\phi^\star|f(X^\star)|\phi^\star\rangle = p^\star$. This would contradict the assumption that $(\mathcal{H}^\star, X^\star, \psi^\star)$ is a minimizer of Problem (1).

The state optimality conditions (42) can therefore be assumed in any Archimedean NPO:

Proposition 11. *If Problem (2) is Archimedean, then its optimizer $\sigma^\star$ satisfies the state optimality conditions (42).*

Conditions (42) allow us to incorporate new constraints to non-commutative optimization problems. Given a Hermitian operator H, define $E_0(H)$ as the bottom of the spectrum of H, and let $\mathrm{Gr}(H)$ denote the set of so-called *ground states* such that $\sigma(H - E_0(H)) = 0$. Then, the state optimality conditions (42) allow us to solve optimization problems of the form

$$\begin{aligned} p^\star := \min_{\mathcal{H},X,\psi} \; & \psi(f(X)) \\ \text{s.t. } & g_i(X) \geq 0, && i = 1,\ldots,m, \\ & h_j(X) = 0, && j = 1,\ldots,m', \\ & \sigma \in \mathrm{Gr}(b_k(X)), && k = 1,\ldots,m''. \end{aligned} \tag{43}$$

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

IV. NON-COMMUTATIVE CONSTRAINT QUALIFICATION

A. Essential KKT conditions

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

Theorem 12. *Any* bounded *solution $\sigma^\star$ of NPO Problem (2) satisfies Eq. (36). In particular, any Archimedean NPO problem satisfies the essential ncKKT conditions.*

Remark 13. Setting $\epsilon_i = 0$ for all i, essential ncKKT reduces to the SDP constraints conditions (24) implied by weak ncKKT on the SDP relaxation (4). The reason why Theorem 12 demands non-zero ϵ_i is to prevent that $\mu_i^k(p) = \infty$ for some polynomials $p \in \mathcal{P}_{2k}$. Think, e.g., of Problem (33): 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 12, we first define an important ideal:

Definition 14. *Let $\mathcal{J}$ be the ideal generated by $\{h_j\}_j$. It consists of all polynomials of the form*

$$\sum_{j,l} v_{jl}(x) h_j(x) w_{jl}(x). \tag{44}$$

We call $\mathcal{J}_k \subset \mathcal{J}$ the subspace obtained by imposing $deg(v_{jl}) + deg(w_{jl}) \leq k - \deg h_j$ in the decomposition (44) above.

We will also need the following technical result.

Lemma 15. *Let the n-tuple of symmetric polynomials $q(x)$ satisfy*

$$\nabla_x h_j(q(x)) \in \mathcal{J}, \quad j = 1,\ldots,m', \tag{45}$$

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

$$h_j(X^\star) = 0, \quad j = 1,\ldots,m'. \tag{46}$$

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

$$\begin{aligned} & h_j(X(t)) = 0, \quad j = 1,\ldots,m', \; t \in [-\epsilon, \epsilon], \\ & X(0) = X^\star, \\ & \left.\frac{dX(t)}{dt}\right|_{t=0} = q(X^\star). \end{aligned} \tag{47}$$

If, in addition, $g_i(X^\star) \geq 0$ for all i, and, for some $\{s_i \in \mathcal{P}\}_i$, the polynomials

$$s_i(x)g_i(x) + g_i(x)s_i(x)^* + \nabla_x g_i(q(x)), \quad i = 1, \ldots, m, \tag{48}$$

are SOS, then the trajectory $X(t)$ also satisfies:

$$g_i(X(t)) \geq 0, \quad i = 1, \ldots, m, \tag{49}$$

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

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

$$\begin{aligned} \frac{dX(t)}{dt} &= q(X), \\ X(0) &= X^\star. \end{aligned} \tag{50}$$

Since $X_1^\star, \ldots, X_n^\star$ are bounded, we can apply the Cauchy-Kovalevskaya theorem 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 (47).

Being polynomials of a tuple of analytic operators, $\{h_j(X(t))\}_j$ are also analytic in the region $\{t \in \mathbb{C} : |t| \leq \epsilon\}$. Due to relation (45), we have that

$$\begin{aligned} \frac{dh_j(X(t))}{dt} &= \nabla_x h_j \left(\frac{dX(t)}{dt} \right) = \nabla_x h_j(q(X(t))) \\ &= \sum_{j',l} r^+_{jj'l}(X(t)) h_{j'}(X(t)) r^-_{jj'l}(X(t)). \end{aligned} \tag{51}$$

Now, define $H_j(t) := h_j(X(t))$. That way, we arrive at the system of ODEs:

$$\begin{aligned} \frac{dH_j(t)}{dt} &= \sum_{j',l} r^+_{jj'l}(X(t)) H_{j'}(t) r^-_{jj'l}(X(t)), \quad j = 1, \ldots, m', \\ H_j(0) &= h_j(X(0)) = h_j(X^\star) = 0, \quad j = 1, \ldots, m'. \end{aligned} \tag{52}$$

Since all the operators are bounded, this equation can be solved through any standard numerical method, e.g.: Euler's explicit method. Take $\Delta > 0$ and consider the following time discretization: $t \in \{k\Delta : k = 0, 1, 2, \ldots\}$. From the equation above, we obtain the recursion relation

$$H_j(k+1; \Delta) = H_j(k; \Delta) + \Delta \sum_{j',l} r^+_{jj'l}(X(k\Delta)) H_{j'}(k; \Delta) r^-_{jj'l}(X(k\Delta)), \quad j = 1, \ldots, m'. \tag{53}$$

Starting from the point $H_j(0; \Delta) = 0$ for all j, this recursion relation will always give us $H_j(k; \Delta) = 0$ for all k. Taking the limit $\Delta \to 0$, we end up with

$$0 = H_j(t) = h_j(X(t)), \tag{54}$$

for $t \geq 0$. An analogous recursion relation shows that the above equation also holds for t for negative times. In sum, the trajectory $\{X(t) : t \in [-\epsilon, \epsilon]\}$ satisfies Eq. (47).

Similarly, let us assume that Eq. (49) holds. Define the variable $G_i(t) := \omega_i(t) g_i(X(t)) \omega_i(t)^*$, where $\omega_i(t)$ is the solution of the differential equation:

$$\frac{d\omega_i(t)}{dt} = \omega_i(t) s_i(X(t)), \quad \omega_i(0) = \mathbb{I}. \tag{55}$$

For short times, $\omega(t)$ is close to the identity, and thus invertible.

The function $G_i(t)$ is also analytic in $t \in [-\epsilon', \epsilon']$, for some $\epsilon' > 0, \epsilon' \leq \epsilon$. Taking differentials, we find

$$\begin{aligned} \frac{dG_i(t)}{dt} &= \omega(t)(s_i(X(t)) g_i(X(t)) + g_i(X(t)) s_i^*(X(t)) + \nabla_x g_i(q(X(t))) \omega_i(t)^* \\ &\overset{(48)}{=} \omega_i(t) \left(\sum_l s^*_{il}(X(t)) s_{il}(X(t)) + \sum_{i'l} s^*_{ii'l}(X(t)) g_{i'}(X(t)) s_{ii'l}(X(t)) + \sum_{jl} s^+_{ijl}(X(t)) h_j(X(t)) s^-_{ijl}(X(t)) \right) \omega_i(t)^* \\ &\overset{H(t)=0}{=} \omega_i(t) \left(\sum_l s^*_{il}(X(t)) s_{il}(X(t)) + \sum_{i'l} s^*_{ii'l}(X(t)) \omega_{i'}(t)^{-1} G_{i'}(t) (\omega_{i'}(t)^{-1})^* s_{ii'l}(X(t)) \right) \omega_i(t)^*. \end{aligned} \tag{56}$$

Again, we try to solve this system of equations on $\{G_i(t)\}_i$ with the Euler explicit method. The recursion relation is

$$G_i(k+1;\Delta) = G_i(k;\Delta) + \Delta\omega_i(\Delta k)\left(\sum_l s^*_{il}(X(\Delta k))s_{il}(X(\Delta k)) + \sum_{i'l} s^*_{ii'l}(X(\Delta k))\omega_{i'}(\Delta k)^{-1}G_{i'}(k;\Delta))(\omega_{i'}(\Delta k)^{-1})^* s_{ii'l}(X(\Delta k))\right)\omega_i(\Delta k)^*. \tag{57}$$

Clearly, starting from positive semidefinite operators

$$G_i(0;\Delta) = g_i(X^\star) \geq 0, \quad i = 1,\dots,m, \tag{58}$$

it holds that $G_i(k;\Delta) \geq 0$, for all k. Taking the limit $\Delta \to 0$, we find that $\omega_i(t)g_i(X(t))\omega_i(t)^* \geq 0$, for $t \in [0,\epsilon']$. Since $\omega_i(t)$ is invertible, this implies that $g_i(X(t)) \geq 0$ and thus Eq. (49) holds. □

Proof of Theorem 12. 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). For fixed $k \in \mathbb{N}$, consider the semidefinite program

$$\begin{aligned}
\mathbb{P} := \min_{\epsilon,\mu,\lambda}\ & \epsilon \\
\text{s.t. }\ & \epsilon \geq 0, \\
& \mu^k_i(pp^*) + \epsilon\|p\|_2^2 \geq 0, \quad \forall p \in \mathcal{P},\ \deg(p) \leq k, && \text{(59a)}\\
& \mu^k_i(pg_lp^*) + \epsilon\|p\|_2^2 \geq 0, \quad \forall p \in \mathcal{P},\ \deg(p) \leq k - \left\lceil \frac{\deg(g_l)}{2}\right\rceil,\ i = 1,\dots,m,\ l = 1,\dots,m, && \text{(59b)}\\
& \mu^k_i(s^+h_js^-) = 0, \quad \forall s^+, s^- \in \mathcal{P},\ \deg(s^+) + \deg(s^-) \leq 2k - \deg(h_j),\ j = 1,\dots,m, && \text{(59c)}\\
& \mu^k_i(sg_i) = \mu^k_i(g_is) = 0, \quad \forall s \in \mathcal{P},\ \deg(s) \leq 2k - \deg(g_i) \quad i = 1,\dots,m, && \text{(59d)}\\
& \lambda^k_j(s^+h_js^-) = 0, \quad \forall s^+, s^- \in \mathcal{P},\ \deg(s^+) + \deg(s^-) \leq 2k - \deg(h_j),\ j = 1,\dots,m', && \text{(59e)}\\
& \sigma^\star\left(\nabla_x f(p(x))\right) - \sum_i \mu^k_i\left(\nabla_x g_i(p(x))\right) - \sum_j \lambda^k_j\left(\nabla_x h_j(p(x))\right) = 0, \quad \forall p \in \mathcal{P}^{\times n}, && \text{(59f)}\\
& \deg\left(\nabla_x f\right), \deg\left(\nabla_x g_i\right), \deg\left(\nabla_x h_j\right) \leq 2k - \deg(p). && \text{(59g)}
\end{aligned}$$

This problem has feasible points. To see why, note that there exist Hermitian λ,μ satisfying Eqs. (59c)–(59f) iff, for all $p \in \mathcal{P}^{\times n}$, $\{s_i\}_i \subset \mathcal{P}$, the conditions

$$\begin{aligned}
&\deg\left(\nabla_x f\right), \deg\left(\nabla_x g_i\right), \deg\left(\nabla_x h_j\right) \leq 2k - \deg(p),\\
&\deg(s_i) \leq 2k - \deg(g_i), i = 1,\dots,m,
\end{aligned} \tag{60}$$

and

$$\begin{aligned}
&\nabla_x g_i(p) - s_ig_i - g_is_i^* \in \mathcal{J}_{2k}, i = 1,\dots,m,\\
&\nabla_x h_j(p) \in \mathcal{J}_{2k}, j = 1,\dots,m',
\end{aligned} \tag{61}$$

imply that

$$\sigma^\star(\nabla_x f(p)) = 0. \tag{62}$$

Indeed, if that were not the case, then the second and third terms of Eq. (59f) would be zero, no matter the values of λ,μ, while the first term would not. Conversely, if condition (62) holds whenever Eqs. (60), (61) do, then, by basic linear algebra, there exists a solution λ,μ of the linear system of equations (59c)–(59f).

Let $(\mathcal{H}^\star, X^\star, \psi^\star)$ be the result of applying the GNS construction on $\sigma^\star$, and suppose that $p \in \mathcal{P}^{\times t}$ satisfies Eqs. (60) and (61). Then, $X^\star, q := \pm p$ satisfy the conditions of Lemma 15, and thus there exist two feasible operator trajectories $\{X^\pm(t) : t \in [0,\delta]\} \subset B(\mathcal{H}^\star)$ such that

$$X^\pm(0) = X^\star,\ \left.\frac{dX^\pm}{dt}\right|_{t=0} = \pm p(X^\star). \tag{63}$$

Hence,

$$\pm\sigma^\star(\nabla_x f(p)) = \sigma^\star(\nabla_x f(\pm p)) = \psi^\star\left(\nabla_x f(\pm p(X^\star))\Big|_{X=X^\star}\right) = \left.\frac{d\psi^\star\left(f(X^\pm(t))\right)}{dt}\right|_{t=0} \geq 0, \tag{64}$$

where the last inequality follows from the fact that $X = X^\star$ minimizes the expression $\psi^\star\left(f(X)\right)$ over tuples $X \in B(\mathcal{H}^\star)^{\times n}$ of feasible operators. Thus, Eq. (62) holds whenever p, s satisfy Eqs. (60), (61).

It follows that there exist μ, λ satisfying conditions (59c)–(59f). For each i, the corresponding moment and localizing matrices of μ_i might not be positive semidefinite, but, having finite entries, one can find $\epsilon > 0$ such that conditions (59a), (59b) hold. We therefore conclude that SDP $\mathbb{P}$ is feasible.

The dual of $\mathbb{P}$ is the following SDP:

$$\begin{aligned}
\mathbb{P}^* := \max_{q,s,Z} & -\sigma^\star(\nabla_x f(q)) \\
\text{s.t. } & \nabla_x g_i(q) + s_i g_i + g_i s_i^* - \sum_{a,b} (Z_i)_{ab} o_a g_i o_b^* - \sum_{l,a,b} (Z_{i,l})_{ab} o_a^l g_l (o_b^l)^* \in \mathcal{J}_{2k}, \\
& \nabla_x h_j(q) \in \mathcal{J}_{2k}, \quad j = 1, \ldots, m', \\
& Z_i \geq 0, \; Z_{i,l} \geq 0, \quad i, k = 1, \ldots, m, \\
& \sum_{i,l} \mathrm{tr}(Z_{i,l}) \leq 1.
\end{aligned} \tag{65}$$

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 $\mathbb{P}^*$ 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 program $\mathbb{P}^*$ 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 Problem $\mathbb{P}^*$. By definition, $q, X^\star$ satisfy the conditions of Lemma 15. Thus, there exists $\delta > 0$ and a trajectory of feasible operators $(X(t) : t \in [0, \delta])$ such that

$$X(0) = X^\star, \; \left.\frac{dX}{dt}\right|_{t=0} = q(X^\star). \tag{66}$$

Hence,

$$-\sigma^\star(\nabla_x f(q)) = -\left.\frac{d\psi^\star\left(f(X(t))\right)}{dt}\right|_{t=0} \leq 0. \tag{67}$$

The solution of Problem $\mathbb{P}^*$ is therefore zero.

Now, SDP $\mathbb{P}$ 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 $\mathbb{P}$, $\mathbb{P}^*$ coincide. This implies that one can find feasible points of Problem $\mathbb{P}$ for any $\epsilon > 0$. Hence, conditions (36) hold for arbitrary $\{\epsilon_i\}_i \subset \mathbb{R}^+$. □

1. *Setting ϵ_i to zero*

In Remark 13, the presence of $\epsilon_i > 0$ in Theorem 12 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. (36). This is exactly what the next theorem shows.

Theorem 16. *Consider an NPO problem that satisfies the Archimedean condition. For some set of indices $I \subset \{1, \ldots, m\}$, let there exist $r \in \mathbb{R}^+$ and a tuple of Hermitian polynomials $q \in \mathcal{P}^{\times n}$ such that*

$$\begin{aligned}
& \nabla_x g_i(q) + s_i g_i + g_i s_i^* - r \; SOS, \quad \forall i \in I, \\
& \nabla_x g_i(q) + s_i g_i + g_i s_i^* \; SOS, \quad \forall i \notin I, \\
& \nabla_x h_j(q) \in \mathcal{J}, \quad j = 1, \ldots, m'.
\end{aligned} \tag{68}$$

for some polynomials $\{s_i : i = 1, \ldots, m\}$. Then, there exist positive linear functionals $\{\mu_i : \mathcal{P} \to \mathbb{C}\}_{i \in I}$, compatible with the constraints $\{g_i(x) \geq 0\}_i \cup \{h_j(x) = 0\}_j$, such that, for any $k \in \mathbb{N}$ and $\{\epsilon_i : i \notin I\} \subset \mathbb{R}^+$, *there exist Hermitian functionals* $\{\mu_i^k : \mathcal{P}_{2k} \to \mathbb{C}\}_{i \notin I}$, $\{\lambda_j^k : \mathcal{P}_{2k} \to \mathbb{C}\}_j$ *satisfying*

$$
\begin{aligned}
&\mu_i^k(pp^*) + \epsilon_i \|p\|_2^2 \geq 0, \quad \forall p \in \mathcal{P},\ \deg(p) \leq k,\ \forall i \notin I\\
&\mu_i^k(pg_l p^*) + \epsilon_i \|p\|_2^2 \geq 0, \quad \forall p \in \mathcal{P},\ \deg(p) \leq k - \left\lceil \frac{\deg(g_l)}{2} \right\rceil,\ l = 1, \dots, m,\ \forall i \notin I\\
&\mu_i^k(s^+ h_j s^-) = 0, \quad \forall s^+, s^- \in \mathcal{P},\ \deg(s^+) + \deg(s^-) \leq 2k - \deg(h_j),\ j = 1, \dots, m,\ \forall i \notin I,\\
&\mu_i^k(sg_i) = \mu_i^k(g_i s) = 0, \quad \forall s \in \mathcal{P},\ \deg(s) \leq 2k - \deg(g_i),\ \forall i \notin I,\\
&\lambda_j^k(s^+ h_j s^-) = 0, \quad \forall s^+, s^- \in \mathcal{P},\ \deg(s^+) + \deg(s^-) \leq 2k - \deg(h_j),\ j = 1, \dots, m',\\
&\sum_{i \notin I} \mu_i^k \left(\nabla_x g_i(t(x))\right) + \sum_j \lambda_j^k \left(\nabla_x h_j(t(x))\right) = \sigma^\star \left(\nabla_x f(t(x))\right) - \sum_{i \in I} \mu_i \left(\nabla_x g_i(t(x))\right), \quad \forall t \in \mathcal{P}^{\times n},\\
&\deg\left(\nabla_x f\right), \deg\left(\nabla_x g_i\right), \deg\left(\nabla_x h_j\right) \leq 2k - \deg(t).
\end{aligned}
\tag{69}
$$

In particular, the essential ncKKT conditions (36) hold with $\epsilon_i = 0$ *for all* $i \in I$*.*

Remark 17. Note that, if weak ncKKT holds, then one can use Eq. (68) to upper bound the norm of the Lagrange multipliers $\{\mu_i : i \in I\}$. Indeed, take $p = q$ in Eq. (21b). By Eq. (68) it then holds that

$$
\sigma^\star(\nabla_x f(q)) = \sum_i \mu_i(\nabla_x g_i(q)) = \sum_i \mu_i(\nabla_x g_i(q) + s_i g_i + g_i s_i^*) \geq r \sum_i \mu_i(1). \tag{70}
$$

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 16, we will need the following simple lemma.

Lemma 18. *Consider the system of linear equations in the variables* $y \in \mathbb{R}^s$ *given by*

$$
A \cdot y = b, \tag{71}
$$

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* $b' \in \operatorname{im}(A) \subset \mathbb{R}^r$ *and any solution* y *of (71), there exists a solution* y' *of the system*

$$
A \cdot y' = b', \tag{72}
$$

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

Proof. If system (71) is solvable, then any solution y thereof can be expressed as

$$
y = A^+ \cdot b + \bar{y}, \tag{73}
$$

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. (72), as long as Eq. (72) is solvable. Moreover,

$$
\|y - y'\|_2 \leq \|A^+\| \|b - b'\|_2. \tag{74}
$$

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

Proof of Theorem 16. By Theorem 12, 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. (36). 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 (68) implies that we can choose the functionals $\{\mu_i^k : i \in I\}$ to have bounded entries.

First, by virtue of the first three lines of Eq. (36), we have that, for any polynomial $p \in \mathcal{P}_{2k}$ admitting an SOS decomposition (7) with the degree constraints

$$
\deg(s_j) \leq k,\ 2\deg(s_{il}) \leq 2k - \deg(g_i),\ \deg(s_{jl}^+) + \deg(s_{jl}^-) \leq 2k - \deg(h_j), \quad \forall i, j, l, \tag{75}
$$

it holds that

$$
\mu_i^k(p) + \epsilon_i \left(\sum_j \|s_j\|_2^2 + \sum_l \|s_{il}\|_2^2 \right) \geq 0. \tag{76}
$$

Take $p = q$ (the polynomial tuple in Eq. (68)). Provided that k is large enough so that the degree conditions (75) of the SOS decompositions in Eq. (68) hold, this implies that there exist constants $\{K_i\}_i \subset \mathbb{R}^+$, independent of $\{\epsilon_i\}_i$, such that

$$
\sigma^\star(\nabla_x f(q)) + \sum_i \epsilon_i K_i = \sum_i \mu_i^k(\nabla_x g_i(q)) + \sum_i \epsilon_i K_i \geq r \sum_{i \in I} \mu_i^k(1). \tag{77}
$$

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. (76) then implies that, for any $p \in \mathcal{P}$, there exists a constant C_p, independent of $\{\epsilon_i\}_i$, such that, for k large enough,

$$K_p \mu_i^k(1) \pm \mu_i^k(p) + \epsilon_i C_p \geq 0. \tag{78}$$

If k is also large enough to accommodate the SOS decompositions in Eq. (68), we conclude by Eq. (77) 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

$$|\mu_i^k(p)| \leq C, \quad \forall i \in I, \tag{79}$$

for all polynomials p of the form $p = m+m^*$ or $p = i(m-m^*)$, where m is a monomial of degree smaller than or equal to $2k'$. Obviously, the restrictions $\lambda^{k'}, \mu^{k'}$ of said Lagrange multipliers λ^k, μ^k to the smaller domain $\mathcal{P}_{2k'}$ also satisfy conditions (36), with the replacement $k \to k'$. Moreover, by the argument above, for all $p \in \mathcal{P}_{2k'}$, $i \in I$, $\mu_i^{k'}(p)$ 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 12, for each $k \in \mathbb{N}$ there exists feasible $\mu^k, \lambda^k : \mathcal{P}_{2k} \to \mathbb{C}$ satisfying Eq. (36) for $\epsilon_i = \frac{1}{k}$, with $\{\mu_i^k : i \in I\}$ bounded. Consider now the sequence of functionals $(\{\mu_i^k : \mathcal{P} \to \mathbb{C}\}_{i\in I})_k$ (we extend μ_i^k from $\mathcal{P}_{2k}$ to $\mathcal{P}$ by adding zeros). Now, let $i \in I$. Since, for all monomials $m \in \mathcal{P}$, 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 $m \in \mathcal{P}$. By the Banach-Alaoglu theorem [38, 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 $(k_s)_s$, call $\{\tilde{\mu}_i : \mathcal{P} \to \mathbb{C}\}_{i\in I}$ 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 $i \in I$,

$$\lim_{s\to\infty} \mu_i^{k_s}\Big|_{\mathcal{P}_{2k}} = \mu_i\Big|_{\mathcal{P}_{2k}}. \tag{80}$$

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

$$M^k(\mu_i) + \delta\mathbb{I} \geq 0, \; M_l^k(\mu_i) + \delta\mathbb{I} \geq 0. \tag{81}$$

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 $s \in \mathbb{N}$ such that $k \leq k_s$, there exist $\{\mu_i^{k,s}\}_{i\notin I}$, $\{\lambda_j^{k,s}\}_j$ satisfying the first five lines of Eqs. (69) and such that

$$\sum_{i\notin I} \mu_i^{k,s}(\nabla_x g_i(p)) + \sum_j \lambda_j^{k,s}(\nabla_x h_j(p)) = \sigma^\star(\nabla_x f(p)) - \sum_{i\in I} \mu_i^{k_s}(\nabla_x g_i(p)), \tag{82}$$

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

$$\begin{aligned}
&\sum_{i\notin I} \mu_i^k(\nabla_x g_i(p)) + \sum_j \lambda_j(\nabla_x h_j(p)) = c^s(p), \quad \forall p, \\
&\mu_i^k(s^+ h_j s^-) = 0, \quad \forall s^+, s^- \in \mathcal{P}, \; \deg(s^+) + \deg(s^-) \leq 2k - \deg(h_j), \; j = 1, \ldots, m, \; i \notin I \\
&\mu_i^k(s g_i) = \mu_i^k(g_i s) = 0, \quad \forall s \in \mathcal{P}, \; \deg(s) \leq 2k - \deg(g_i), \; i \notin I, \\
&\lambda_j^k(s^+ h_j s^-) = 0, \quad \forall s^+, s^- \in \mathcal{P}, \; \deg(s^+) + \deg(s^-) \leq 2k - \deg(h_j), \; j = 1, \ldots, m',
\end{aligned} \tag{83}$$

where p, s, s^+, s^- denote any polynomial of the form $m+m^*$ or $i(m-m^*)$ and m is a monomial of degree small enough so that all polynomials involved have degree $2k$ or lower. The "constant vector" $c^s(p)$ is given by

$$c^s(p) := \sigma^\star(\nabla_x f(p)) - \sum_{i\in I} \mu_i^{k_s}(\nabla_x g_i(p)). \tag{84}$$

Since system (83) is solvable for all s, then it is also solvable in the limit $s \to \infty$ (because the image of the corresponding matrix of coefficients, being finite-dimensional, is a closed subspace). By Lemma 18, for high enough s, there exists a solution $\{\mu_i^k : i \notin I\}, \{\lambda_j^k\}_j$ of the system

$$\begin{aligned}
&\sum_{i\notin I} \mu_i^k(\nabla_x g_i(p)) + \sum_j \lambda_j^k(\nabla_x h_j(p)) = \lim_{s\to\infty} c^s(p) = \sigma^\star(\nabla_x f(p)) - \sum_{i\in I} \mu_i(\nabla_x g_i(p)), \quad \forall p,\\
&\mu_i^k(s^+ h_j s^-) = 0, \quad \forall s^+, s^- \in \mathcal{P},\ \deg(s^+) + \deg(s^-) \leq 2k - \deg(h_j),\ j = 1, \ldots, m,\ i \notin I\\
&\mu_i^k(s g_i) = \mu_i^k(g_i s) = 0, \forall s \in \mathcal{P},\ \deg(s) \leq 2k - \deg(g_i),\ i \notin I,\\
&\lambda_j^k(s^+ h_j s^-) = 0, \quad \forall s^+, s^- \in \mathcal{P},\ \deg(s^+) + \deg(s^-) \leq 2k - \deg(h_j),\ j = 1, \ldots, m',
\end{aligned} \tag{85}$$

such that $\{\mu_i^k : i \notin I\}, \{\lambda_j^k\}_j$ is arbitrarily close to $\{\mu_i^{k,s} : i \notin I\}, \{\lambda_j^{k,s}\}_j$. As s increases, the moment and localizing matrices of $\{\mu_i^k : i \notin I\}$ will therefore differ from those of $\{\mu_i^{k,s} : i \notin I\}$ by an arbitrarily small amount. Thus, adding them $\frac{1}{l_s} + \kappa_s$ times the identity, where κ_s tends to zero as s increases, will give positive semidefinite matrices.

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. (69) 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. (69) 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. □

B. Normed KKT conditions

In this section, we will investigate tractable conditions under which the normed ncKKT conditions from Definition 7 hold. To find them, we will generalize known sufficient criteria for the classical case.

In this regard, Mangasarian-Fromovitz constraint qualification MFCQ [21] stipulates that any commutative problem of the form (9) satisfies the (commutative) KKT conditions if

$$\{\partial_x h_j(x^\star)\}_j, \tag{86}$$

are linearly independent vectors and there exists $z \in \mathbb{R}^n$ such that

$$\begin{aligned}
&\langle \partial_x g_i(x^\star) | z\rangle > 0, \quad \forall i \in \mathbb{A}(x^\star),\\
&\langle \partial_x h_j(x^\star) | z\rangle = 0, \quad j = 1, \ldots, m',
\end{aligned} \tag{87}$$

where $\mathbb{A}(x)$ denotes the set of active inequality constraints.

In Section IV B 1, we generalize MFCQ to non-commutative problems and prove that it suffices to guarantee that the corresponding NPO problem satisfies the normed ncKKT conditions. Before we proceed, though, we need to generalize the notion of gradient linear independence for sets of non-commutative polynomial equality constraints. This is the subject of the next section.

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

$$\langle \partial_x h_j | v_k(x)\rangle = \delta_{j,k}, \quad j, k = 1, \ldots, m'. \tag{88}$$

Now, define the matrices

$$\begin{aligned}
&\hat{P}_j(x) := |v_j(x)\rangle\langle \partial_x h_j|, \quad j = 1, \ldots, m',\\
&\hat{P}_0(x) := \mathbb{I} - \sum_{k=1}^{m'} \hat{P}_k(x).
\end{aligned} \tag{89}$$

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

$$\begin{aligned}
&\sum_k \hat{P}_k(x)|z\rangle = |z\rangle,\\
&\langle \partial_x h_j | \hat{P}_k(x)|z\rangle = 0, \quad \forall j \neq k,\\
&\langle \partial_x h_j | z\rangle = 0 \ \rightarrow\ \langle \hat{P}_j(x)|z\rangle = 0.
\end{aligned} \tag{90}$$

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 $h(x)$ is a linear map from n-tuples of polynomials $p = (p_1, \ldots, p_n)$ to a single polynomial $\nabla_x g(p(x))$. A non-commutative analog of relations (90) 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 $\hat{P}_k(x) \cdot z$ played in relations (90). Correspondingly, the tuples $P_0, \ldots, P_{m'}$ should satisfy the following conditions:

$$\sum_k P_k(X,Z) = Z, \tag{91a}$$

$$\nabla_x h_j(P_k(X,Z)) = 0, \quad \forall j \neq k, \tag{91b}$$

$$\nabla_x h_j(Z) = 0 \; \rightarrow \; P_{j+m}(X,Z) = 0. \tag{91c}$$

We will regard the algebraic version of constraints (91a)–(91c) as the non-commutative generalization of gradient linear independence.

Definition 19. *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,*

$$\sum_{j=0}^{m'} P_j(x,p) - p \in \mathcal{J}^{\times n}, \quad \forall p \in \mathcal{P}\times n, \tag{92a}$$

$$\nabla_x h_j(P_k(x,p)) \in \mathcal{J}, \quad \forall k \neq j, \; p \in \mathcal{P}^{\times n}, \tag{92b}$$

$$\exists \beta^+, \beta^- : (P_j(x,p))_k - \sum_l \beta^+_{jkl}(x) \nabla_x h_j(p)(\beta^-_{jkl}(x))^* - \beta^-_{jkl}(x)\nabla_x h_j(p)(\beta^+_{jkl}(x))^* \in \mathcal{J},$$
$$j = 1, \ldots, m', \quad k = 1, \ldots, n, \quad p \in \mathcal{P}^{\times n}. \tag{92c}$$

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

Remark 20. One can replace Eq. (92c) by the weaker, problem-dependent constraints

$$\exists \beta^+, \beta^-, \gamma^+, \gamma^-, s, \text{ such that}$$
$$\nabla_x g_i\left(P_j(x,p)\right) - \sum_l \left(\beta^+_{ijl}(x)\nabla_x h_j(p)(\beta^-_{ijl}(x))^* + \beta^-_{ijl}(x)\nabla_x h_j(p)(\beta^+_{ijl}(x))^*\right) + s_i(p) g_i + g_i s_i(p) \in \mathcal{J},$$
$$\nabla_x f\left(P_j(x,p)\right) - \sum_l \left(\gamma^+_{jl}(x)\nabla_x h_j(p)(\gamma^-_{jl}(x))^* + \gamma^-_{jl}(x)\nabla_x h_j(p)(\gamma^+_{jl}(x))^*\right) \in \mathcal{J},$$
$$j = 1, \ldots, m', \quad k = 1, \ldots, n, \quad p \in \mathcal{P}^{\times n}. \tag{93}$$

Indeed, the proofs of Lemmas 24, 28 and Theorem 27 below follow through with such a modified definition.

Unless the quotient space $\mathcal{P}/\{h_j\}_j$ is finite-dimensional, verifying that conditions (92) hold for all $p \in \mathcal{P}^{\times n}$ is a challenging endeavor. The next proposition provides a practical, sufficient condition to guarantee gradient linear independence.

Proposition 21. *Let $\{h_j\}_j$ be a set of Hermitian polynomials, and let $\{r_k\}_k \subset \mathcal{P}$ be such that $[r_k, x_i] \in \mathcal{J}$ 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*

$$\sum_{j=0}^{m'} P_j(x,z) - z \in (\mathcal{J}^Z)^{\times n}, \tag{94a}$$

$$\nabla_x h_j(P_k(x,z)) \in \mathcal{J}^Z, \quad \forall k \neq j, \tag{94b}$$

$$\exists \beta^+, \beta^- : (P_j(x,z))_k - \sum_l \beta^+_{jkl}(x)\nabla_x h_j(z)(\beta^-_{jkl}(x))^* - \beta^-_{jkl}(x)\nabla_x h_j(z)(\beta^+_{jkl}(x))^* \in \mathcal{J}^Z,$$
$$j = 1, \ldots, m', \quad k = 1, \ldots, n, \tag{94c}$$

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

$$q(x,z) = \sum_{j,l} p^+_{jl}(x,z) h_j(x) p^-_{jl}(x,z) + \sum_{k,l,i} q^+_{ikl}(x)[r_k(x), z_i] q^-_{ikl}(x). \tag{95}$$

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

The proof is obvious.

Example 22. Let $\{h_j(x)=0\}_{j=1}^{m'}$ be of the form

$$h_j(x)=\sum_k \nu_{jk}x_k+b_j, \tag{96}$$

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

$$\begin{aligned}(P_j(x,z))_k &:= v^j_k\sum_l \nu_{jl}z_l, \quad j=1,\ldots,m',\\ P_0(x,z) &:= z-\sum_j P_j(x,z).\end{aligned} \tag{97}$$

The newly defined $\{P_j\}_j$ satisfy Eqs. (94). That is: for linear constraints, the commutative and non-commutative definitions of gradient linear independence coincide.

Example 23. Consider the equality constraints $\{h_j(x)=0\}_{j=1}^n$ for

$$h_j(x):=x_j^2-1, \quad j=1,\ldots,n. \tag{98}$$

This system also satisfies gradient linear independence. Indeed, define

$$\begin{aligned}(P_j(x,z))_k &:= \delta_{jk}\frac{1}{2}(z_j+x_jz_jx_j),\\ (P_0(x,z))_k &:= \frac{1}{2}(z_k-x_kz_kx_k).\end{aligned} \tag{99}$$

Then it can be easily verified that the n-tuples of polynomials $P_0,\ldots,P_n$ satisfy conditions (94).

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 24. *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, $\{\mu_i\}_i$, compatible with the constraints $\{g_i(x)\geq 0\}_i\cup\{h_j(x)=0\}_j$, 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*

$$\begin{aligned}&\lambda^k_j(s^+h_ks^-)=0, \quad j,k=1,\ldots,m',\ \deg(s^+h_ks^-)\leq 2k,\\ &\sum_j\lambda^k_j(\nabla_x h_j(t(x)))=\sigma(\nabla_x f(t(x)))-\sum_i\mu_i(\nabla_x g_i(t(x))), \quad \forall t\in\mathcal{P}^{\times n},\\ &\qquad \deg(t)\leq 2k+1-\deg(f)+1,\ 2k+1-\deg(g_i), \quad i=1,\ldots,m,\end{aligned} \tag{100}$$

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 $\|\lambda^\pm\|_{\text{SOS}}\leq O(K)$.

Proof. By Eq. (92b), if we set $t=P_0(x,p)$ in Eq. (100), we arrive at

$$\sigma(\nabla_x f(P_0(x,p)))-\sum_i\mu_i(\nabla_x g_i(P_0(x,p)))=0, \tag{101}$$

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 $p\in\mathcal{P}$.

Now, consider the linear functional $\lambda_j:\mathcal{P}\to\mathbb{C}$, defined by

$$\begin{aligned}\lambda_j(p) := \sigma&\left(\nabla_x f\left(\left(\sum_l\beta^+_{jkl}p(\beta^-_{jkl})^*+\beta^-_{jkl}p(\beta^+_{jkl})^*\right)_k\right)\right)\\ &-\sum_i\mu_i\left(\nabla_x g_i\left(\left(\sum_l\beta^+_{jkl}p(\beta^-_{jkl})^*+\beta^-_{jkl}p(\beta^+_{jkl})^*\right)_k\right)\right),\end{aligned} \tag{102}$$

where the polynomials $\{\beta^\pm_{jkl}\}_{jkl}$ are the ones appearing in Eq. (92c).

Note that, for $r,t\in\mathcal{P}$ and any positive functional ω, the functional

$$\lambda(p):=\omega(rpt^*+tpr^*) \tag{103}$$

equals the difference between two positive functionals, namely,

$$\lambda(p) = \omega^+(p) - \omega^-(p), \tag{104}$$

for

$$\omega^{\pm}(p) := \frac{1}{2}\omega((r \pm t)p(r \pm t)^*). \tag{105}$$

Now, we expand the arguments of the functionals on the right-hand side of (102), as polynomials of the form $\sum_i s_i p t_i^* + t_i p s_i^*$. Since $\sigma^\star$, $\{\mu_i\}$ are bounded positive functionals, we find that the right-hand side of Eq. (102) can be decomposed as a finite sum of differences of positive functionals. Thus,

$$\lambda_j = \lambda_j^+ - \lambda_j^-, \tag{106}$$

where the positive functionals $\lambda_j^{\pm}$ inherit from σ, $\{\mu_i\}_i$ the property of being compatible with the constraints $\{g_i(x) \geq 0\}_i \cup \{h_j(x) = 0\}_j$. The values $\lambda_j^{\pm}(1)$ correspond to expressions of the form

$$\lambda_j^{\pm}(1) = \sigma(s^{\pm}) + \sum_i \mu_i(s_i^{\pm}), \tag{107}$$

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 (31c). Taking $t = P_0(x,p)$ in (31c) and invoking Eq. (92b) and Eq. (101) we have that Eq. (31c) is satisfied as long as t is of the form $t = P_0(x,p)$.

Next, set $t = P_j(x,p)$ in Eq. (31c), for some $j \in \{1, \ldots, m'\}$, $p \in \mathcal{P}^{\times n}$. By Eq. (92b), all terms of the form $\lambda_k(\nabla_x h_k(P_j(x,p)))$ with $k \neq j$, drop. Furthermore, by Eqs. (92a), (92b), we have that

$$\lambda_j\left(\nabla_x h_j(p)\right) = \sum_k \lambda_j\left(\nabla_x h_j(P_k(x,p))\right) = \lambda_j\left(\nabla_x h_j(P_j(x,p))\right). \tag{108}$$

It follows that

$$\begin{aligned}
\lambda_j(\nabla_x h_j(P_j(x,p))) &= \lambda_j(\nabla_x h_j(p)) \\
&= \sigma\left(\nabla_x f\left(\sum_l \beta_{jkl}^+ \nabla_x h_j(p) \beta_{jkl}^-\right)\right) - \sum_i \mu_i\left(\nabla_x g_i\left(\sum_l \beta_{jkl}^+ \nabla_x h_j(p) \beta_{jkl}^-\right)\right) \\
&= \sigma(\nabla_x f(P_j(x,p))) - \sum_i \mu_i(\nabla_x g_i(P_j(x,p))),
\end{aligned} \tag{109}$$

where the last equality is a consequence of Eq. (92c). Hence, Eq. (31c) holds for $t = P_j(x,p)$, for $j = 0, ..m'$.

By Eq. (92a), any tuple of polynomials p can be expressed as a sum of terms of the form $\{P_k(x,p)\}_k$ (modulo elements of $\mathcal{J}$), and so Eq. (31c) holds in general. □

2. *Non-commutative Mangasarian-Fromovitz Constraint Qualification*

We are ready to generalize Mangasarian-Fromovitz constraint qualification.

Definition 25 (ncMFCQ)**.** *Consider an NPO Problem* (2)*. 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 $q \in \mathcal{P}^{\times n}$ such that*

$$\nabla_x g_i(q) + s_i g_i + g_i s_i^* - r \quad SOS, \quad i = 1, \ldots, m, \tag{110a}$$

$$\nabla_x h_j(q) \in \mathcal{J}, \quad j = 1, \ldots, m'. \tag{110b}$$

for some polynomials $\{s_i : i = 1, \ldots, m\} \subset \mathcal{P}$.

Example 26. 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:

$$g_i(x) = 1 + \tilde{g}_i(x) \geq 0, i = 1, \ldots, m, \tag{111}$$

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. (110b) holds. On the other hand, $\nabla_x g_i(x)(q) = -r_i \tilde{g}_i(x)$, and so

$$r_i g_i(x) + \nabla_x g_i(x)(q) - r_i = 0. \tag{112}$$

That is, Eq. (110a) 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 Magasarian-Fromovitz constraint qualification [21].

Theorem 27. *Consider an NPO Problem (2) that satisfies both the Archimedean condition and ncMFCQ. Then, the problem satisfies the normed ncKKT conditions.*

Proof. Eq. (110a) and (110b) represent the conditions of Theorem 16 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

$$\begin{aligned}
&\lambda_j^k(s^+ h_k s^-) = 0, \quad j, k = 1, \ldots, m', \ \deg(s^+ h_k s^-) \leq 2k, \\
&\sum_j \lambda_j^k(\nabla_x h_j(t(x))) = \sigma^\star(\nabla_x f(t(x))) - \sum_i \mu_i(\nabla_x g_i(t(x))), \quad \forall t \in \mathcal{P}^{\times n},
\end{aligned} \tag{113}$$

where the degree of t is appropriately bounded, has a solution λ^k. The conditions of Lemma 24 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 28. *Assume that Problem (2) 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. (31). That is, we can find $K \in \mathbb{R}^+$ such that*

$$\mu_i(1), \lambda_j^\pm(1) \leq K, \quad \forall i, j. \tag{114}$$

Proof. Take $t = q$ in Eq. (31c). Then we have that

$$\begin{aligned}
\sigma^\star\left(\nabla_x f(q(x))\right) &\overset{(110b)}{=} \sum_i \mu_i\left(\nabla_x g_i(q(x))\right) \\
&\overset{(110a)}{\geq} r \sum_i \mu_i(1) - \sum_i \mu_i(s_i g_i + g_i s_i^*) \\
&\overset{(21a)}{=} r \sum_i \mu_i(1).
\end{aligned} \tag{115}$$

In turn, provided that the original NPO problem is Archimedean, there exists $\eta \in \mathbb{R}^+$ such that

$$\eta - \nabla_x f(q(x)) \tag{116}$$

admits an SOS decomposition. This implies that the left-hand side of Eq. (115) is upper bounded by K. Thus, the SOS seminorm of each of the multipliers $\{\mu_i\}_i$ is bounded by $\frac{\eta}{r}$.

The set of linear functionals $\{\lambda_j\}_j$, restricted to the set of polynomials of degree $2k$ or smaller, is a solution of Eq. (100). The conditions of Lemma 24 are met, and thus there exist bounded linear functionals $\{\tilde{\lambda}_j^\pm : \mathcal{P} \to \mathbb{C}\}_j$, with norms bounded by $O(\eta)$, compatible with the problem constraints and satisfying Eq. (31c). □

C. Strong KKT conditions

It is observed in practice that many natural NPO problems admit an exact SOS resolution, see, e.g., [11]. Namely, for some $k \in \mathbb{N}$, the k^{th} level of the hierarchy of SDP relaxations (6) has a (feasible) maximizer achieving the exact solution of the problem. The next theorem shows that, in such a predicament, strong ncKKT holds.

Theorem 29. *Consider the NPO Problem (2), and let $f - p^\star$ be SOS. Then, any bounded solution $(\mathcal{H}^\star, X^\star, \psi^\star)$ of Problem (1) satisfies Eqs. (20). Therefore, if Problem (2) is Archimedean, it satisfies strong ncKKT.*

Proof. Let $(\mathcal{H}^\star, X^\star, \psi^\star)$ be a bounded minimizer of Problem (1). For simplicity, in the following we use its abstract, functional form $\sigma^\star : \mathcal{P} \to \mathbb{C}$, namely:

$$\sigma^\star(p) = \psi^\star(p(X^\star)). \tag{117}$$

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

$$f - p^\star = \sum_l s_l s_l^* + \sum_{i,l} s_{il} g_i s_{il}^* + \sum_{j,l} s^+_{jl} h_j (s^-_{jl})^* + s^-_{jl} h_j (s^+_{jl})^*. \tag{118}$$

In addition,

$$\sigma^\star(f - p^\star) = 0. \tag{119}$$

It follows that

$$\begin{aligned} &\sigma^\star(s_l s_l^*) = 0, \quad \forall l, \\ &\sigma^\star(s_{il} g_i s_{il}^*) = 0, \quad \forall i, l. \end{aligned} \tag{120}$$

These relations, in turn, imply that, for any $q \in \mathcal{P}$,

$$\sigma^\star(s_l q) = \sigma^\star(q s_l^*) = 0, \quad \forall l, \tag{121a}$$

$$\sigma^\star(s_{il} g_i q) = \sigma^\star(q g_i s_{il}^*) = 0, \quad \forall i, l. \tag{121b}$$

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

$$\begin{pmatrix} \sigma^\star(qq^*) & \sigma^\star(q s_l^*) \\ \sigma^\star(s_l q^*) & \sigma^\star(s_l s_l^*) \end{pmatrix}. \tag{122}$$

The second one, from the positive semidefiniteness of

$$\begin{pmatrix} \sigma^\star(q g_i q^*) & \sigma^\star(q g_i s_{il}^*) \\ \sigma^\star(s_{il} g_i q^*) & \sigma^\star(s_{ij} g_i s_{ij}^*) \end{pmatrix}. \tag{123}$$

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 $\pi^\delta : \mathcal{P} \to \mathcal{P}$ 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 $\sigma^\delta(pp^*) \geq 0$ for all $p \in \mathcal{P}$. However, it does not necessarily satisfy feasibility conditions of the form $\sigma^\delta(p g_i p^*) \geq 0$, $\sigma^\delta(s^+ h_j s^-) = 0$.

We apply the state σ^δ on both sides of Eq. (118). Taking into account Eqs. (119), (121), and the chain rule of differentiation, the result is

$$\delta \sigma^\star(\nabla_x f(p(x))) + O(\delta^2) = \delta \sum_i \mu_i(\nabla_x(g_i(p(x))) + \delta \sum_j \lambda_j(\nabla_x(h_j(p(x))) + O(\delta^2), \tag{124}$$

where μ_i denotes the positive linear functional given by

$$\mu_i(q) := \sigma^\star(\sum_l s_{il} q s_{il}^*), \tag{125}$$

and λ_j is the linear functional

$$\lambda_j(q) := \sigma^\star(\sum_l s^+_{jl} q (s^-_{jl})^* + \sum_l s^-_{jl} q (s^+_{jl})^*). \tag{126}$$

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

$$\lambda_j^\pm(q) := \frac{1}{2} \sum_l \sigma^\star((s^+_{jl} \pm s^-_{jl}) q (s^+_{jl} \pm s^-_{jl})^*). \tag{127}$$

Collecting the terms in Eq. (124) that depend linearly on δ, we have that

$$\sigma^\star(\nabla_x f(p(x))) = \sum_i \mu_i(\nabla_x(g_i(p(x))) + \sum_j \lambda_j(\nabla_x(h_j(p(x))). \tag{128}$$

This is condition (20b).

Finally, the positive linear functionals $\{\mu_i\}_i$ satisfy complementary slackness (20a), for

$$\mu_i(g_i) = \sum_l \sigma^\star(s_{il} g_i s_{il}^*) = 0, \quad \forall i, \tag{129}$$

by Eq. (121).

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

$$\mu_i(p) = \tilde{\mu}_i(p(X^\star)),\ \lambda_j^\pm(p) = \tilde{\lambda}_j^\pm(p(X^\star)), \tag{130}$$

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. (117) and the fact that both sets of functionals are defined in terms of $\sigma^\star$. □

The reader might question the practical use of Theorem 29. 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 (1). This polynomial is non-negative, but not positive.

As we pointed out in Section II, 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) do not admit an SOS resolution.

However, some sets of noncommuting 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) [39]. In the following, we provide two families of constraints that are known to generate Archimedean closed quadratic modules.

1. Equality constraints with a faithful finite-dimensional ∗-representation

Proposition 30. *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 $\pi : \mathcal{P} \to B(\mathcal{H}^\star)$, with $\ker(\pi) = \mathcal{J}$. Then, the quadratic module generated by $\{h_j(x) = 0\}_j$ is Archimedean closed.*

Proof. Define $X_k^\star := \pi(x_k)$, denote by $\mathcal{A}(X^\star)$ the unital C^*-algebra generated by $X_1^\star, \ldots, X_n^\star$ and let $p \in \mathcal{P}$ 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 $s \in \mathcal{P}$ such that $p(X^\star)^{1/2} = s(X^\star)$. We thus have that

$$p(X^\star) - s(X^\star)s(X^\star)^* = 0. \tag{131}$$

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

Remark 31. A set of constraints satisfying the conditions of Proposition 30 is the Pauli algebra (170), (171) used to model many-body quantum systems, see Section VI A.

Corollary 32. *Let $\{h_j(x) = 0\}_j$ be a set of constraints satisfying the conditions of Proposition 30. Then, any problem of the form (2) with just equality constraints $\{h_j(x) = 0\}_j$ satisfies strong ncKKT.*

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

$$p\left(\delta Y_1 + (1-\delta) Y_2\right) \leq \delta p(Y_1) + (1-\delta) p(Y_2), \tag{132}$$

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 *[40].*

As shown in [40], $p \in \mathcal{P}$ is matrix convex iff its *Hessian* is *matrix positive*, i.e., if the polynomial

$$\frac{d^2 p(x+th)}{dt^2}\Big|_{t=0} \tag{133}$$

is a sum of squares on the Hermitian variables $x_1, \dots, x_n, h_1, \dots, h_n$. In [40] 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 34. *Let the Archimedean Problem (2) 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*

$$h_j(x) = \sum_k \nu_{jk} x_k - b_j, \tag{134}$$

and linearly independent.

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

$$g_i(q(x)) - r \tag{135}$$

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

$$h_j(q(x)) \in \mathcal{J}, \quad j = 1, \dots, m'. \tag{136}$$

Then, the constraints $\{g_i(x) \geq 0\}_i \cup \{h_j(x) = 0\}_j$ *generate an Archimedean closed quadratic module. Thus, by Theorem 29, Problem (2) satisfies the strong ncKKT conditions.*

In the proof of the theorem we shall make use of Schur complements [41]:

Lemma 35. *The block matrix*

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

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

Proof. This is a special case of [41, 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 34. Let $(\mathcal{H}^\star, \sigma^\star, X^\star)$ be a solution of Problem (1). By (135) and (136), the choice $\hat{X} = q(X^\star)$ satisfies the constraints

$$g_i(\hat{X}) > 0 \quad \text{for all } i, \qquad h_j(\hat{X}) = 0 \quad \text{for all } j. \tag{137}$$

We solve the system of linear equations $\{h_j = 0\}_j$. Without loss of generality, we express the last $n - r$ variables in terms of $x_1, \dots, x_r$. By back substitution, our Problem (1) 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 (7). It thus suffices to consider an Archimedean Problem (1) with $m' = 0$.

By the Helton-McCullough structure theorem for concave noncommutative polynomials [40, Corollary 7.1], each g_i is of the form

$$g_i(x) = c_i + \Lambda_{i0}(x) - \sum_{j=1}^{N} \Lambda_{ij}(x)^* \Lambda_{ij}(x) \tag{138}$$

for some $c_i \in \mathbb{R}$ and homogeneous linear $\Lambda_{ij}(x)$. (By Caratheodory's theorem on convex hulls in finite dimensions [42, Theorem I.2.3], the length N of the sum of squares in (138) can be chosen independently of i.) Such a g_i is the Schur complement of a linear pencil, namely,

$$G_i(x) = \begin{bmatrix} c_i + \Lambda_{i0}(x) & \Lambda_{i1}(x)^* & \cdots & \Lambda_{iN}(x)^* \\ \Lambda_{i1}(x)^* & 1 & & \\ \vdots & & \ddots & \\ \Lambda_{iN}(x) & & & 1 \end{bmatrix} \tag{139}$$

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 35 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

$$\begin{aligned} \langle L(\hat{x})\eta|\eta\rangle &= \langle L(\psi^* \hat{X}\psi)\eta|\eta\rangle = \langle (I_{m(N+1)} \otimes \psi)^* L(\hat{X})(I_{m(N+1)} \otimes \psi)\eta|\eta\rangle \\ &= \langle L(\hat{X})(I_{m(N+1)} \otimes \psi)\eta|(I_{m(N+1)} \otimes \psi)\eta\rangle = \langle L(\hat{X})(\eta \otimes \psi)|(\eta \otimes \psi)\rangle > 0. \end{aligned} \tag{140}$$

Now again by Lemma 35, $g_i(\hat{x}) > 0$ 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 [43, Theorem 1.2] to deduce that $f - p^\star$ has a SOS decomposition. Thus, by Theorem 29, Problem (2) admits the strong ncKKT conditions. □

Corollary 36. *Let the Problem* (2) *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) *such that $g_i(X) > 0$ for all i, then Problem* (2) *satisfies the strong ncKKT conditions.*

Proof. Immediate from the proof of Theorem 34. □

Remark 37. The Convex Positivstellensatz of [43] has minimal degree. Enforcing the strong ncKKT conditions on an NPO problem with matrix convex constraints is therefore superfluous: the first level of relaxation (6) with degree high enough to encode the problem will already achieve convergence.

V. PARTIAL OPERATOR OPTIMALITY CONDITIONS

In some situations, one might not be able to justify all operator optimality conditions, but some subset thereof. This is the case, for instance, when 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. That is,

$$[z_k, y_l] = 0, \quad \forall k, l. \tag{141}$$

As it turns out, if the remaining constraints on z satisfy convex constraints 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. This is formalized in the following two sections.

A. Partial normed ncKKT

Theorem 38. *Consider an Archimedean NPO* (2) *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:*

$$[y_r, z_s] = 0, \quad \forall r, s. \tag{142}$$

(b) *The only remaining constraints on z satisfy ncMFCQ.*

Let $\mathcal{P}_Z$ be the set of polynomials with variables z. Then, there exist positive linear functionals $\{\mu_i : \mathcal{P}_Z^{\times q} \to \mathbb{C}\}_i$, $\{\lambda_i^\pm : \mathcal{P}_Z^{\times n} \to \mathbb{C}\}_j$, both compatible with the constraints $\{\hat{g}_i(z) \geq 0\}_i \cup \{\hat{h}_j(z) = 0\}_j$, such that

$$\mu_i(\hat{g}_i) = 0, \quad i = 1, \ldots, m_Z, \tag{143a}$$

$$\sigma\left(\nabla_z f(p)\right) - \sum_i \mu_i\left(\nabla_z \hat{g}_i(p)\right) - \sum_j \lambda_j \left(\nabla_z \hat{h}_j(p)\right) = 0, \quad \forall p \in \mathcal{P}_Z^{\times q}, \tag{143b}$$

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

Proof. Let $(\mathcal{H}^\star, X^\star, \sigma^\star)$ be a solution of Eq. (1). For fixed $k \in \mathbb{N}$, consider the following SDP:

$$\begin{aligned}
&\min_{\epsilon,\mu,\lambda} \epsilon \\
\text{s.t. } &\epsilon \geq 0, \\
&\mu_i^k(pp^*) + \epsilon\|p\|_2^2 \geq 0, \quad \forall p \in \mathcal{P}_Z,\ \deg(p) \leq k,\ i = 1,.,,,m_Z, \\
&\mu_i^k(p\hat{g}_l p^*) + \epsilon\|p\|_2^2 \geq 0, \quad \forall p \in \mathcal{P}_Z,\ \deg(p) \leq k - \left\lceil \frac{\deg(\hat{g}_l)}{2} \right\rceil,\ i,l = 1,\ldots,m_Z, \\
&\mu_i^k(s^+\hat{h}_j s^-) = 0, \quad \forall s^+, s^- \in \mathcal{P}_Z,\ \deg(s^+) + \deg(s^-) \leq 2k - \deg(\hat{h}_j),\ j = 1,\ldots,m, \qquad (144) \\
&\mu_i^k(sg_i) = \mu_i^k(g_i s) = 0, \quad \forall s \in \mathcal{P},\ \deg(s) \leq 2k - \deg(g_i), i = 1,\ldots,m, \\
&\lambda_j^k(s^+\hat{h}_j s^-) = 0, \quad \forall s^+, s^- \in \mathcal{P}_Z,\ \deg(s^+) + \deg(s^-) \leq 2k - \deg(h_j),\ j = 1,\ldots,m'_Z, \\
&\sigma^\star(f'(p)) - \sum_i \mu_i^k(\nabla_z \hat{g}_i(p)) - \sum_j \lambda_j^k(\nabla_z h_j(p)) = 0, \quad \forall p \in \mathcal{P}_Z^{\times q}, \\
&\deg_z(f'(p)),\ \deg(\nabla_z \hat{g}_i),\ \deg\left(\nabla_z \hat{h}_j\right) \leq 2k - \deg(p), \qquad (145)
\end{aligned}$$

where

$$f'(p) = \lim_{\delta\to 0} \frac{f(Y^\star, Z^\star + \delta p(Z^\star)) - f(Y^\star, Z^\star)}{\delta}, \tag{146}$$

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

Invoking Lemma 15 as in the proof of Theorem 12, we find that this SDP problem has a solution for every $\epsilon > 0$. Next, using relations (110a), (110b), like in the proof of Theorem 16, it is shown that Problem (145) 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 24 to prove that $\{\lambda_j^k\}_j$ can be replaced by linear functionals. □

B. Partial strong ncKKT

Theorem 39. *Consider an NPO problem (2), not necessarily Archimedean, 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:*

$$[y_r, z_s] = 0, \quad \forall r, s. \tag{147}$$

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

$$\begin{aligned}
\hat{g}_i(z) &\geq 0, \quad i = 1,\ldots,m_Z, \\
\hat{h}_j(z) &= 0, \quad j = 1,\ldots,m'_Z,
\end{aligned} \tag{148}$$

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

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

$$\hat{g}_i(Q(z)) - r \tag{149}$$

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

$$\hat{h}_j(Q(z)) = \sum_{l,j'} s_{jj'l}(z)\hat{h}_{j'}(z)s'_{jj'l}(z), \tag{150}$$

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

(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*

$$\begin{aligned}
\left.\frac{d^2 f(y, z+th)}{dt^2}\right|_{t=0} &= \sum_{l,k} \zeta_{lk}^+(y,z,h)[y_k, z_l]\zeta_{lk}^-(y,z,h) + \eta_{lk}^+(y,z,h)[y_k, h_l]\eta_{lk}^-(y,z,h) \\
&+ \sum_l \theta_l(z,h,y)\theta_l^*(z,h,y) + \sum_i \theta_{il}(z,h,y)\tilde{g}_i(y)\theta_{il}^*(z,h,y) + \sum_j \iota_{jl}^+(z,h,y)\tilde{h}_j(y)\iota_{jl}^-(z,h,y),
\end{aligned} \tag{151}$$

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

Let $(\mathcal{H}^\star, X^\star, \sigma^\star)$ be any bounded solution of Problem (1), with $X^\star = (Y^\star, Z^\star)$, and denote by $\mathcal{A}$ the C^*-algebra generated by $Z_1^\star, \ldots, Z_q^\star$. Then, there exist positive linear functionals $\{\mu_i : \mathcal{A} \to \mathbb{C}\}_i$ and bounded Hermitian linear functionals $\{\lambda_i : \mathcal{A} \to \mathbb{C}\}_j$ such that

$$\mu_i(\hat{g}_i(Z^\star)) = 0, \quad i = 1, \ldots, m_Z, \tag{152a}$$

$$\sigma^\star \left(\nabla_z f(p)\right) - \sum_i \mu_i \left(\nabla_z \hat{g}_i(p)\right) - \sum_j \lambda_j \left(\nabla_z \hat{h}_j(p)\right) = 0, \quad \forall p \in \mathcal{A}^{\times q}. \tag{152b}$$

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

Lemma 40. *Let $\mathcal{A}$ be a C^*-algebra, call $\mathcal{A}_h$ its set of Hermitian elements. Let $Z = (Z_1, \ldots, Z_q)$ be a set of Hermitian operator variables and let $\tilde{f} : \mathcal{A}^{\times q} \to \mathbb{C}$ be a Hermitian convex function[3]. Given some* concave *non-commutative polynomials $\{\hat{g}_i\}$ and linear affine polynomials $\{\hat{h}_i\}$, consider the optimization problem*

$$\begin{aligned} \min_{Z \in \mathcal{A}^{\times q}} \quad & \hat{f}(Z) \\ s.t. \quad & \hat{g}_i(Z) \geq 0, \quad i = 1, \ldots, m_Z, \\ & \hat{h}_j(Z) = 0, \quad j = 1, \ldots, m'_Z. \end{aligned} \tag{153}$$

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

Further assume that Problem (153) admits a strictly feasible point, i.e., there exists a feasible tuple $\hat{Z} \in \mathcal{A}^{\times q}$ such that

$$\begin{aligned} & \hat{g}_i(\hat{Z}) > 0, \quad i = 1, \ldots, m_Z, \\ & \hat{h}_j(\hat{Z}) = 0, \quad j = 1, \ldots, m'_Z. \end{aligned} \tag{154}$$

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

$$\mu_i(\hat{g}_i(Z^\star)) = 0, \quad i = 1, \ldots, m_Z, \tag{155}$$

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

$$\min_{Z \in \mathcal{A}_h^{\times q}} \mathcal{L}(Z; \mu, \lambda), \tag{156}$$

with

$$\mathcal{L}(Z; \mu, \lambda) := \hat{f}(Z) - \sum_i \mu_i(\hat{g}_i(Z)) - \sum_j \lambda_j(\hat{h}_j(Z)). \tag{157}$$

Proof. It suffices to follow the classical proof of the Slater criterion for strong duality (cf. [16, §4.2]). Given $Z^\star$, we define the sets:

$$\begin{aligned} A :=& \{(r, S, T) : r \in \mathbb{R},\, S \in \mathcal{A}_h^{\times m_Z},\, T \in \mathcal{A}_h^{\times m'_Z}, \\ & \exists Z_1, \ldots, Z_n \in \mathcal{A}_h, \hat{f}(Z) \leq r,\, -\hat{g}_i(Z) \leq S_i,\, \hat{h}_j(Z) = T_j,\, \forall i, j\}, \\ B :=& \{(\nu, 0, 0) : \nu < \hat{f}(Z^\star)\}. \end{aligned} \tag{158}$$

Clearly, $A \cap B = \emptyset$. Also, both sets are convex. Since they live in a real normed space (namely, $\mathbb{R} \times \mathcal{A}^{\times m_Z + m'_Z}$) and B is open, the Hahn-Banach separation theorem [38, Theorem V.4(a)] implies that there exists a separating linear functional (ϕ, μ, λ) and $\alpha \in \mathbb{R}$ such that

$$\phi r + \sum_i \mu_i(S_i) + \sum_j \lambda_j(T_j) \geq \alpha, \quad \forall (r, S, T) \in A, \tag{159a}$$

$$\phi \nu \leq \alpha, \quad \forall (\nu, 0, 0) \in B. \tag{159b}$$

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. (159a) 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 $\mathcal{A}$.

---

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

Notice as well that we can choose ν to be arbitrarily small in Eq. (159b). It follows that $\phi \geq 0$. We next prove that $\phi > 0$. Suppose, on the contrary, that $\phi = 0$. From Eqs. (159a), (159b) we have that

$$\phi \hat{f}(Z) - \sum_i \mu_i(g_i(Z)) - \sum_j \lambda_j(h_i(Z)) \geq \alpha \geq \phi \hat{f}(Z^\star), \quad \forall Z \in \mathcal{A}^{\times p}. \tag{160}$$

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

$$\phi(\hat{f}(\hat{Z}) - f(Z^\star)) \geq \sum_i \mu_i(g_i(\hat{Z})). \tag{161}$$

Thus, if $\phi = 0$, $\mu_i(\hat{g}_i(\hat{Z})) = 0$, 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. (159a) implies that

$$\sum_j \lambda_j(h_j(Z)) \geq \alpha, \quad \forall Z. \tag{162}$$

This can just 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. (160) by ϕ, we have that

$$\hat{f}(Z) - \sum_i \tilde{\mu}_i(\hat{g}_i(Z)) - \sum_j \tilde{\lambda}_j(\hat{h}_j(Z)) \geq \hat{f}(Z^\star), \quad \forall Z, \tag{163}$$

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. (163). We arrive at:

$$f(Z^\star) - \sum_i \tilde{\mu}_i(\hat{g}_i(Z^\star)) \geq f(Z^\star). \tag{164}$$

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 (155). In that case,

$$\mathcal{L}(Z^\star, \tilde{\mu}, \tilde{\lambda}) = \hat{f}(Z^\star), \tag{165}$$

and so, by the above equation and (163), $Z^\star$ is a global solution of the unconstrained problem (156). □

Proof of Theorem 39. Let $(\mathcal{H}^\star, \sigma^\star, X^\star)$ be a bounded solution of Problem (1), 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 (147), it follows that the solution $p^\star$ of Problem (2) satisfies

$$\begin{aligned} p^\star = &\min_{Z \in \mathcal{A}^{\times q}} \hat{f}(Z) \\ \text{s.t. } &\hat{g}_i(Z) \geq 0, \quad i = 1, \ldots, m_Z, \\ &\hat{h}_j(Z) = 0, \quad j = 1, \ldots, m'_Z, \end{aligned} \tag{166}$$

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

$$\hat{f}(Z) = \sigma^\star(f(Y^\star, Z)). \tag{167}$$

Moreover, one of the minimizers of (166) is $Z = Z^\star$. This function is convex by virtue of Eq. (151), which ensures that its Hessian is non-negative [40].

In addition, by Eqs. (149), (150), 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 40 and conclude that $Z^\star$ is the solution of the unconstrained Problem (156), 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 $\mathcal{A}^{\times q}$

$$Z(t) := Z^\star + tp(Z^\star). \tag{168}$$

Since $Z^\star$ is a minimizer of Problem (156), it follows that

$$0 = \frac{d\mathcal{L}(Z(t); \mu, \lambda)}{dt}\Big|_{t=0} = \sigma^\star\left(\nabla_z f(Y^\star, z)(p(Z^\star))\right) - \sum_i \mu_i^k \left(\nabla_z \hat{g}_i(p)\Big|_{Z=Z^\star}\right) - \sum_j \lambda_j \left(\nabla_z \hat{h}_j(p)\Big|_{Z=Z^\star}\right) = 0. \tag{169}$$

Since this relation is valid for arbitrary $p \in \mathcal{P}_Z^{\times q}$, we arrive at the statement of the theorem. □

VI. APPLICATIONS

A. Many-body quantum systems

A consequence of the state optimality condition (42) 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*:

$$\begin{aligned} &(\sigma_x^j)^2 = (\sigma_y^j)^2 = (\sigma_z^j)^2 = 1, \\ &\sigma_x^j\sigma_y^j - i\sigma_z^j = \sigma_y^j\sigma_z^j - i\sigma_x^j = \sigma_z^j\sigma_x^j - i\sigma_y^j = 0. \end{aligned} \tag{170}$$

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

$$[\sigma_a^j, \sigma_b^k] = 0, \quad a, b \in \{x, y, z\},\ j \neq k. \tag{171}$$

The set of constraints (170), (171) admits (up to unitary equivalence) a unique irreducible operator representation $\pi : \mathcal{P} \to B(\mathbb{C}^2)^{\otimes n}$, with

$$\begin{aligned} \pi(\sigma_x^j) &:= \mathbb{I}_2^{\otimes j-1} \otimes \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \otimes \mathbb{I}_2^{\otimes n-j}, \\ \pi(\sigma_y^j) &:= \mathbb{I}_2^{\otimes j-1} \otimes \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \otimes \mathbb{I}_2^{\otimes n-j}, \\ \pi(\sigma_z^j) &:= \mathbb{I}_2^{\otimes j-1} \otimes \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \otimes \mathbb{I}_2^{\otimes n-j}. \end{aligned} \tag{172}$$

It is easy to see that π satisfies the conditions of Proposition 30.

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

$$H(\sigma) = \sum_{j>k}^{n} P_{jk}(\sigma^j, \sigma^k), \tag{173}$$

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) [44]. Quantum chemists [3–5] (and, more recently, condensed matter physicists [9, 45–47]) use NPO to lower bound $E_0(H)$. In essence, they relax the problem

$$\begin{aligned} E_0(H) = &\min \rho(H) \\ &\text{s.t. } (\sigma_x^j)^2 = (\sigma_y^j)^2 = (\sigma_z^j)^2 = 1, \quad j = 1, \ldots, n, \\ &\quad \sigma_x^j\sigma_y^j - i\sigma_z^j = \sigma_y^j\sigma_z^j - i\sigma_x^j = \sigma_z^j\sigma_x^j - i\sigma_y^j = 0, \quad j = 1, \ldots, n, \\ &\quad [\sigma_a^j, \sigma_b^k] = 0, \quad a, b \in \{x, y, z\},\ j \neq k \end{aligned} \tag{174}$$

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

$$M^{\pm} := \mp \min\{\mp\langle\psi|\pi(\sum_j \sigma_z^j)|\psi\rangle : \langle\psi| \bullet |\psi\rangle \in \mathrm{Gr}(\pi(H))\}. \tag{175}$$

For instance, Wang *et al.* [47] 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:

$$\begin{aligned} \bar{M}^{\pm} = &\mp \min\{\mp\rho(\sum_j \sigma_z^j)\} \\ &\text{s.t. } (\sigma_x^j)^2 = (\sigma_y^j)^2 = (\sigma_z^j)^2 = 1, \quad j = 1, \ldots, n, \\ &\quad \sigma_x^j\sigma_y^j - i\sigma_z^j = \sigma_y^j\sigma_z^j - i\sigma_x^j = \sigma_z^j\sigma_x^j - i\sigma_y^j = 0, \quad j = 1, \ldots, n, \\ &\quad [\sigma_a^j, \sigma_b^k] = 0, \quad a, b \in \{x, y, z\},\ j \neq k, \\ &\quad \rho(H) \leq E_0^+. \end{aligned} \tag{176}$$

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.* [47] 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 [47] 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 [48–50]). The results of [47] are not that good for 2D systems, namely, qubit systems with a Hamiltonian of the form (182) below. For such systems, current variational tools are very imprecise [51].

The state optimality condition (42) allows us to formulate Problem (175) as the following NPO:

$$
\begin{aligned}
m^{\pm} = &\mp \min \rho(\mp \sum_j \sigma_z^j) \\
\text{s.t. } &(\sigma_x^j)^2 = (\sigma_y^j)^2 = (\sigma_z^j)^2 = 1, \quad j = 1, \dots, n, \\
&\sigma_x^j \sigma_y^j - i\sigma_z^j = \sigma_y^j \sigma_z^j - i\sigma_x^j = \sigma_z^j \sigma_x^j - i\sigma_y^j = 0, \quad j = 1, \dots, n, \\
&[\sigma_a^j, \sigma_b^k] = 0, \quad a, b = x, y, z; \quad j \neq k, \\
&\rho \in \mathrm{Gr}(H).
\end{aligned}
\tag{177}
$$

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

$$\sigma([H, p]) = 0 \tag{178}$$

and

$$\sigma\left(p^* H p - \frac{1}{2}\{H, p^* p\}\right) \geq 0. \tag{179}$$

The advantage of the formulation (177) with respect to (176) is that it does not require any upper bound on $E_0(H)$. Problem (177) is thus appropriate to tackle 2D and 3D systems, and even spin glasses [52].

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 [53], the modeller YALMIP [54], and the solver MOSEK [55]. 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. [47], leads to dramatic improvements in performance.

In the 1D case, the Hamiltonian reads

$$H = \frac{1}{4} \sum_{i=0}^{n-1} \sum_{a \in \{x,y,z\}} \sigma_a^i \sigma_a^{i \oplus 1}, \tag{180}$$

where addition $\oplus$ is modulo n.

Results for the energy are shown in Table I. Here, the lower bounds are much tighter than the upper bounds. This is because when calculating the lower bound the state optimality condition (42) is only a tightening of the SDP: without it, the SDP hierarchy (4) would converge anyway to the ground state energy. In calculating the upper bound, however, the state optimality condition (42) is doing all the work, as without it the SDP would converge to the system's maximum energy.

The magnetization is given by

$$M = \sum_{i=0}^{n-1} \sigma_z^i. \tag{181}$$

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 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 II.

In the 2D case the Hamiltonian reads

$$H = \frac{1}{4} \sum_{i,j=0}^{L-1} \sum_{a \in \{x,y,z\}} \sigma_a^{i,j} (\sigma_a^{i \oplus 1, j} + \sigma_a^{i, j \oplus 1}), \tag{182}$$

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

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 III, and for the magnetization in Table IV.

Problem (177) 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:

$$H = \sum_{j=-\infty}^{\infty} P(\sigma^j, \sigma^{j+1}). \tag{183}$$

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

$$\pi_R(\sigma_a^j) = \sigma_a^{j+1}, \quad \pi_L(\sigma_a^j) = \sigma_a^{j-1}. \tag{184}$$

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

$$\begin{aligned}
e_0^n := &\min \rho(P(\sigma^1, \sigma^2))\\
\text{s.t. } &(\sigma_x^j)^2 = (\sigma_y^j)^2 = (\sigma_z^j)^2 = 1, \quad j = 1, \ldots, n,\\
&\sigma_x^j\sigma_y^j - i\sigma_z^j = \sigma_y^j\sigma_z^j - i\sigma_x^j = \sigma_z^j\sigma_x^j - i\sigma_y^j = 0, \quad j = 1, \ldots, n,\\
&[\sigma_a^j, \sigma_b^k] = 0, \quad a, b \in \{x, y, z\}, \ j \neq k\\
&\rho(p) = \rho(\pi_L(p)) = \rho(\pi_R(p)), \quad \text{for } \ p, \pi_L(p), \pi_R(p) \in \mathcal{P}(\sigma^1, \ldots, \sigma^n).
\end{aligned} \tag{185}$$

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 (178) [45, 56]. Note that, if $p \in \mathcal{P}(\sigma^2, \ldots, \sigma^{n-1})$ has degree k, then the above commutator has degree $k + 1$ and only involves the variables $\sigma^1, \ldots, \sigma^n$.

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 III. Ground state energy per site of 2D Heisenberg model. For $L = 3$ we used all nearest-neighbour 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 (42). 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 IV. Magnetization per site of 2D Heisenberg model. For $L = 3$ we used all nearest-neighbour 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 (42). For higher L we used degree up to 2.

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 (179), for it also allows us to bound whatever local property of the system, such as the magnetization. For $1D$ Hamiltonians (183), if p has degree k and depends on the variables $\sigma^2, \ldots, \sigma^{n-1}$, the polynomial in Eq. (179) will be of degree $2k+1$ and only depend on $\sigma^1, \ldots, \sigma^n$. 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

$$H = \frac{1}{4} \sum_{i=-\infty}^{\infty} \sum_{a \in \{x,y,z\}} \sigma_a^i \sigma_a^{i+1}. \tag{186}$$

Results for the energy are shown in Table V.

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 V. 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-neighbour monomials of degree up to 4.

B. The curious case of quantum Bell inequalities

Consider a quantum bipartite Bell experiment [26, 27], where two separate parties conduct measurements on an entangled quantum state. The first party, Alice, conducts measurement x and obtains outcome a. The second party, Bob, respectively calls y, b his measurement setting and outcome. 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

$$C(P) := \sum_{a,b,x,y} C(a, b, x, y) P(a, b|x, y) \tag{187}$$

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

$$\begin{aligned}
c^\star = &\min \sigma\left(\frac{1}{2}\sum_{a,b,x,y} C(a,b,x,y)\{E_{a|x}, F_{b|y}\}\right)\\
\text{s.t. } & E_{a|x} \geq 0, \quad \forall a, x,\\
& \sum_a E_{a|x} - \mathbb{I} = 0, \quad \forall x,\\
& F_{b|y} \geq 0, \quad \forall b, y,\\
& \sum_b F_{b|y} - \mathbb{I} = 0, \quad \forall y,\\
& [E_{a|x}, F_{b|y}] = 0, \quad \forall a, b, x, y.
\end{aligned} \tag{188}$$

As we can appreciate, taking the partition $X = (E, F)$, the NPO satisfies the conditions of Theorem 39, 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 (188) can be chosen such that the non-commuting variables are all projectors [57]. That is,

$$\begin{aligned}
(E^\star_{a|x})^2 = E^\star_{a|x}, \quad \forall a, x,\\
(F^\star_{b|y})^2 = F^\star_{b|y}, \quad \forall b, y.
\end{aligned} \tag{189}$$

Next, we apply Theorem 39 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 (189), 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:

$$\mu^A_{a|x}(ss^*) \geq 0, \quad \forall a, x, \forall s \in \mathcal{P}_E, \tag{190a}$$

$$\mu^A_{a|x}\Big(s\big((E_{a'|x'})^2 - E_{a'|x'}\big)s'\Big) = 0, \quad \forall a, a', x, x',\ \forall s, s' \in \mathcal{P}_E \tag{190b}$$

$$\mu^A_{a|x}\Big(s\Big(\sum_{a'} E_{a'|x'} - 1\Big)s'\Big) = 0, \quad \forall a, x, x',\ \forall s, s' \in \mathcal{P}_E, \tag{190c}$$

$$\lambda^A_x\Big(s\big((E_{a|x'})^2 - E_{a|x'}\big)s'\Big) = 0, \quad \forall a, x, x',\ \forall s, s' \in \mathcal{P}_E \tag{190d}$$

$$\lambda^A_x\Big(s\Big(\sum_a E_{a|x} - 1\Big)s'\Big) = 0, \quad \forall x,\ \forall s, s' \in \mathcal{P}_E, \tag{190e}$$

$$\mu^A_{a|x}(E_{a|x}) = 0, \quad \forall a, x, \tag{190f}$$

$$\sigma\left(\frac{1}{2}\sum_{b,y} C(a,b,x,y)\{p, F_{b|y}\}\right) = \mu^A_{a|x}(p) + \lambda^A_x(p), \quad \forall a, x,\ \forall p \in \mathcal{P}_E. \tag{190g}$$

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 A.

1. *Only two outcomes*

When a, b can only take two values, it is customary to rewrite Problem (188) in terms of 'dichotomic operators' A_x, B_y. The problem to solve is thus

$$\begin{aligned}
&\min \sigma(H)\\
\text{s.t. } & \frac{1 - A_x}{2} \geq 0, \ \frac{1 + A_x}{2} \geq 0, \quad \forall x,\\
& \frac{1 - B_y}{2} \geq 0, \ \frac{1 + B_y}{2} \geq 0, \quad \forall y,\\
& [A_x, B_y] = 0, \quad \forall x, y
\end{aligned} \tag{191}$$

where H is the Bell polynomial

$$\frac{1}{2}\sum_{x,y} c_{xy}\{A_x, B_y\} + \sum_x d_x A_x + \sum_y e_y B_y. \tag{192}$$

To simplify notation, we define the polynomials

$$\mathcal{F}_x := \sum_y \frac{1}{2} c_{xy} B_y + \frac{1}{2} d_x \mathbb{I}, \tag{193a}$$

$$\mathcal{G}_y := \sum_x \frac{1}{2} c_{xy} A_x + \frac{1}{2} e_y \mathbb{I}, \tag{193b}$$

which allow us to express H as

$$H = \sum_x \{\mathcal{F}_x, A_x\} + \sum_y e_y B_y = \sum_y \{\mathcal{G}_y, B_y\} + \sum_x d_x A_x. \tag{194}$$

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

$$(A_x^\star)^2 = (B_y^\star)^2 = 1. \tag{195}$$

Thus, once more we can apply Theorem 39 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:

$$\mu_x^\pm(s(A_{x'}^2 - 1)s') = 0, \quad \forall x, x', \ \forall s, s' \in \mathcal{A}, \tag{196a}$$

$$\mu_x^\pm\left(\frac{1 \pm A_x}{2}\right) = 0, \quad \forall x, \tag{196b}$$

$$\sigma\left(\{p, \mathcal{F}_x\}\right) = \mu_x^+(p) - \mu_x^-(p), \quad \forall x, \ \forall p \in \mathcal{A}. \tag{196c}$$

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. (196b) and the positivity of $\mu_x^\pm$, an analogous argument to the one used to derive Eqs. (121) shows that

$$\mu_x^\pm\left(E_x^\pm p\right) = \mu_x^\pm\left(p E_x^\pm\right) = 0 \quad \forall p \in \mathcal{A}. \tag{197}$$

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

$$\mu_x^\pm\left(\{A_x, q\}\right) = \mp\mu_x^\pm\left(q + A_x q A_x\right) = 0. \tag{198}$$

Thus, if we set $p = -\{A_x, q\}$ in Eq. (196c), we arrive that

$$-\sigma\left(\{\{A_x, q\}, \mathcal{F}_x\}\right) = \mu^+(q + A_x q A_x) + \mu^-(q + A_x q A_x). \tag{199}$$

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

$$-\sigma\left(\{\{A_x, ss^*\}, \mathcal{F}_x\}\right) = \mu^+(ss^* + A_x ss^* A_x) + \mu^-(ss^* + A_x ss^* A_x) \geq 0, \quad \forall s \in \mathcal{P}_E. \tag{200}$$

Setting $p = [A_x, q]$ in Eq. (196c) and using Eq. (197), we obtain another useful constraint:

$$\sigma\left(\{[A_x, q], \mathcal{F}_x\}\right) = 0, \quad \forall q \in \mathcal{A}. \tag{201}$$

Constraints (200), (201) are, respectively, extra positivity and linear conditions that one can apply to the already existing variables of the 'quantum NPO' (188).

2. *Numerical implementation*

In order to implement numerically the constraints (200) and (201), together with the analogous constraints for Bob and the state optimality conditions (42), 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

$$\sigma\left(\mathcal{F}_x[A_x, m_i^A]\right) = 0, \tag{202a}$$

$$\sigma\left(\mathcal{G}_y[B_y, m_i^B]\right) = 0, \tag{202b}$$

$$\sigma\left([H, m_i]\right) = 0, \tag{202c}$$

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 (200) are equivalent to the positive semidefiniteness of the matrices $\{\alpha^x\}_x, \{\beta^y\}_y, \gamma$, with elements given by

$$\alpha^x_{ij} := -\sigma\left(\mathcal{F}_x\{A_x, m_i^{A^*} m_j^A\}\right), \tag{203a}$$

$$\beta^y_{ij} := -\sigma\left(\mathcal{G}_y\{B_y, m_i^{B^*} m_j^B\}\right), \tag{203b}$$

$$\gamma_{ij} := \sigma\left(m_i^* H m_j - \frac{1}{2}\{H, m_i^* m_j\}\right). \tag{203c}$$

Note that, when dealing with Bell inequalities, it is more usual to formulate the problem as a maximization instead of a minimization [1]. One can adapt the KKT constraints for maximization by simply flipping the sign of the positivity conditions (203).

In order for the interior point algorithm to work reliably, it is 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 [58]. Although the vanilla NPA hierarchy is always strictly feasible [59], this is in general not true when additional constraints are enforced [60]. 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. (202c) to rewrite Eq. (203c) as

$$\gamma_{ij} = \sigma(m_i^*[H, m_j]). \tag{204}$$

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 (205), 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 with Bell inequalities in the 2222, 3322, and 4422 scenarios. All calculations were done using the toolkit for non-commutative polynomial optimization Moment [53], the modeller YALMIP [54], and the arbitrary-precision solver SDPA-GMP [61]. 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 [14, 62–65] or, in physicists' slang, that it has a *rank loop* [2]. We remark under the corresponding table whether it holds.

We start with a tilted version of the CHSH inequality [66], with an additional $\tau(A_0 + B_0)$ term [67, 68]. In full correlation notation the coefficients table is given by

$$\left(\begin{array}{c|cc} 0 & \tau & 0 \\ \hline \tau & 1 & 1 \\ 0 & 1 & -1 \end{array}\right). \tag{205}$$

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 VI and VII. 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 VI. Results for the tilted CHSH inequality with $\tau = 0.95$. For comparison, the best known lower bound is 3.9001 6389 9372. With the KKT constraints we get a rank loop at level 3, and without at level 4.

Our next example is the well-studied I3322 inequality [69–71]. In full correlation notation the coefficients table is given by

$$\frac{1}{4}\left(\begin{array}{c|ccc} 0 & -1 & -1 & 0 \\ \hline -1 & -1 & -1 & -1 \\ -1 & -1 & -1 & 1 \\ 0 & -1 & 1 & 0 \end{array}\right). \tag{206}$$

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 VII. Results for the tilted CHSH inequality with $\tau = 0.99$. For comparison, the best known lower bound is 3.9800 0132 8893. With the KKT constraints we find a rank loop at level 4; without, at level 7.

The results are shown in Table VIII. Its maximal violation is conjectured to occur only for an infinite-dimensional system [70], 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 (203), 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 VIII. 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 [72], that had a gap between the best known lower bound and the best known upper bound (see Table IV of Ref. [73]). In full correlation notation the coefficients table is given by

$$\frac{1}{4}\left(\begin{array}{r|rrrr} -12 & -1 & -1 & -2 & 4 \\ \hline -1 & -1 & 1 & 1 & 2 \\ -1 & 1 & -1 & 1 & 2 \\ -2 & 1 & 1 & -2 & 2 \\ 4 & 2 & 2 & 2 & -2 \end{array}\right). \tag{207}$$

The results are shown in Table IX. For increased efficiency the calculations here were done without the positivity conditions (203), 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 IX. 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 6, 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 (205), where we got numerical problems.

VII. CONCLUSION

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

The state optimality conditions (Eq. (42)) and essential ncKKT, 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 [24] under the KKT constraints, proven in [33], 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 [47]. It is intriguing whether the state optimality conditions can be integrated within renormalization flow techniques, like those proposed in [46]. 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. acknowledges support by the Spanish Ministry of Science and Innovation (MCIN) with funding from the European Union Next Generation EU (PRTRC17.I1) and the Department of Education of Castilla y León (JCyL) through the QCAYLE project, as well as MCIN projects PID2020-113406GB-I00 and RED2022-134301-T. M.A and A.G. acknowledge funding from the FWF stand-alone project P 35509-N. 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, J1-2453, N1-0217, J1-3004. T.V. acknowledges the support of the EU (QuantERA eDICT) and the National Research, Development and Innovation Office NKFIH (No. 2019-2.1.7-ERA-NET-2020-00003).

[1] M. Navascués, S. Pironio, and A. Acín, Bounding the set of quantum correlations, Physical Review Letters **98**, 010401 (2007), arXiv:quant-ph/0607119.

[2] 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].

[3] 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).

[4] D. A. Mazziotti, Realization of quantum chemistry without wave functions through first-order semidefinite programming, Physical Review Letters **93**, 213001 (2004).

[5] 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].

[6] 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].

[7] 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].

[8] A. Haim, R. Kueng, and G. Refael, Variational-correlations approach to quantum many-body problems (2020), arXiv:2001.06510 [cond-mat.str-el].

[9] 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].

[10] S. Lawrence, Semidefinite programs at finite fermion density, Physical Review D **107**, 094511 (2023), arXiv:2211.08874 [hep-lat].

[11] 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].

[12] J. W. Helton and S. McCullough, A Positivstellensatz for non-commutative polynomials, Transactions of the American Mathematical Society **356**, 3721 (2004).

[13] 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].

[14] S. Burgdorf, I. Klep, and J. Povh, *Optimization of polynomials in non-commuting variables*, Vol. 2 (Springer, 2016).

[15] L. Vandenberghe and S. Boyd, Semidefinite programming, SIAM Review **38**, 49 (1996).
[16] 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).
[17] W. Karush, Minima of functions of several variables with inequalities as side constraints, M. Sc. Dissertation. Dept. of Mathematics, Univ. of Chicago (1939).
[18] H. Kuhn and A. 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.
[19] C. J. Hillar, Advances on the Bessis–Moussa–Villani trace conjecture, Linear Algebra and its Applications **426**, 130 (2007), arXiv:math/0507166.
[20] J. W. Helton, H. Mousavi, S. S. Nezhadi, V. I. Paulsen, and T. B. Russell, Synchronous values of games (2023), arXiv:2109.14741 [quant-ph].
[21] J. Nocedal and S. Wright, *Numerical Optimization* (Springer New York, 2006).
[22] H. Fawzi, O. Fawzi, and S. O. Scalet, Certified algorithms for equilibrium states of local quantum Hamiltonians (2023), arXiv:2311.18706 [quant-ph].
[23] M. Putinar, Positive polynomials on compact semi-algebraic sets, Indiana University Mathematics Journal **42**, 969 (1993).
[24] J. B. Lasserre, Global optimization with polynomials and the problem of moments, SIAM Journal on Optimization **11**, 796 (2001).
[25] 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.
[26] J. S. Bell, On the Einstein Podolsky Rosen paradox, Physics Physique Fizika **1**, 195–200 (1964).
[27] B. S. Tsirel'son, Quantum analogues of the Bell inequalities. the case of two spatially separated domains, Journal of Soviet Mathematics **36**, 557 (1987).
[28] 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].
[29] P. W. Anderson, *Basic Notions of Condensed Matter Physics*, edited by P. W. Anderson (CRC Press, 2018).
[30] 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).
[31] I. Segal, Irreducible representations of operator algebras, Bulletin of the American Mathematical Society **53**, 73 (1947).
[32] M. F. Anjos and J. B. Lasserre, *Handbook on semidefinite, conic and polynomial optimization*, Vol. 166 (Springer Science & Business Media, 2011).
[33] J. Nie, An exact Jacobian SDP relaxation for polynomial optimization, Mathematical Programming **137**, 225 (2013), arXiv:1006.2418 [math.OC].
[34] 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].
[35] D. S. Kaliuzhnyi-Verbovetskyi and V. Vinnikov, *Foundations of free noncommutative function theory*, Math. Surv. Monogr., Vol. 199 (Providence, RI: American Mathematical Society (AMS), 2014).
[36] 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.
[37] G. J. Murphy, *C*–Algebras and Operator Theory* (Academic Press, San Diego, 1990).
[38] M. Reed and B. Simon, *Methods of modern mathematical physics. I: Functional analysis.* (1980).
[39] N. Ozawa, About the Connes embedding conjecture, Japanese Journal of Mathematics **8**, 147–183 (2013), arXiv:1212.1700 [math.OA].
[40] J. W. Helton and S. McCullough, Convex noncommutative polynomials have degree two or less, SIAM Journal on Matrix Analysis and Applications **25**, 1124 (2004).
[41] F. Zhang, *The Schur Complement and Its Applications*, Numerical Methods and Algorithms (Springer US, 2006).
[42] A. Barvinok, *A Course in Convexity*, Graduate studies in mathematics (American Mathematical Society, 2002).
[43] 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].
[44] 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.
[45] X. Han, Quantum many-body bootstrap (2020), arXiv:2006.06002 [cond-mat.str-el].
[46] I. Kull, N. Schuch, B. Dive, and M. Navascués, Lower bounding ground-state energies of local Hamiltonians through the renormalization group (2022), arXiv:2212.03014 [quant-ph].
[47] J. Wang, J. Surace, I. Frérot, B. Legat, M.-O. Renou, V. Magron, and A. Acín, Certifying ground-state properties of many-body systems (2023), arXiv:2310.05844 [quant-ph].
[48] 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].
[49] 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].
[50] F. Verstraete and J. I. Cirac, Matrix product states represent ground states faithfully, Physical Review B **73**, 094423 (2006), arXiv:cond-mat/0505140.
[51] 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.
[52] 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).
[53] A. Garner and M. Araújo, Moment, `https://github.com/ajpgarner/moment` (2023).
[54] 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/.

[55] MOSEK ApS, *The MOSEK Optimization Suite 10.1.11* (2023), https://docs.mosek.com/latest/intro/index.html.

[56] X. Han, S. A. Hartnoll, J. Kruthoff, *et al.*, Bootstrapping matrix quantum mechanics, Physical Review Letters **125**, 041601 (2020).

[57] 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].

[58] 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].

[59] A. Tavakoli, A. Pozas-Kerstjens, P. Brown, and M. Araújo, Semidefinite programming relaxations for quantum correlations (2023), arXiv:2307.02551 [quant-ph].

[60] M. Araújo, Comment on 'Geometry of the quantum set on no-signaling faces', Physical Review A **107**, 036201 (2023), arXiv:2302.03529 [quant-ph].

[61] 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.

[62] 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.

[63] J. Nie, Certifying convergence of Lasserre's hierarchy via flat truncation, Mathematical Programming **142**, 485 (2013), arXiv:1106.2384 [math.OC].

[64] 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.

[65] 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].

[66] 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).

[67] P. H. Eberhard, Background level and counter efficiencies required for a loophole-free Einstein-Podolsky-Rosen experiment, Physical Review A **47**, R747 (1993).

[68] 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].

[69] 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.

[70] 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].

[71] D. Rosset, SymDPoly: symmetry-adapted moment relaxations for noncommutative polynomial optimization (2018), arXiv:1808.09598.

[72] 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].

[73] K. F. Pál and T. Vértesi, Quantum bounds on Bell inequalities, Physical Review A **79**, 022120 (2009), arXiv:0810.1615 [quant-ph].

Appendix A: 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 VI B, together with the state optimality conditions (42).

$$
\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 \mathcal{P}, \\
& \sigma\Big(s\big((E_{a|x})^2 - E_{a|x}\big)s'\Big) = 0, \quad \forall a, x, \forall s, s' \in \mathcal{P} \\
& \sigma\Big(s\Big(\sum_a E_{a|x} - 1\Big)s'\Big) = 0, \quad \forall x, \forall s, s' \in \mathcal{P} \\
& \sigma\Big(s\big((F_{b|y})^2 - F_{b|y}\big)s'\Big) = 0, \quad \forall b, y, \forall s, s' \in \mathcal{P} \\
& \sigma\Big(s\Big(\sum_b F_{b|y} - 1\Big)s'\Big) = 0, \quad \forall y, \forall s, s' \in \mathcal{P} \\
& \sigma\big(s[E_{a|x}, F_{b|y}]s'\big) = 0, \quad \forall a, b, x, y, \forall s, s' \in \mathcal{P} \\
& \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 \mathcal{P}.
\end{aligned}
$$