Karush-Kuhn-Tucker 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. In this work, we introduce a non-commutative analog of the Karush-Kuhn-Tucker (KKT) optimality conditions, which are satisfied by many classical optimization problems. In the non-commutative setting, the KKT conditions amount to adding new SDP constraints to standard SDP hierarchies, with the effect of boosting their speed of convergence. The new optimality conditions also allow enforcing a new type of constraints in NPO problems: namely, restricting the optimization over states to the set of common ground states of an arbitrary number of operators. Like in the classical case, some necessary conditions or constraint qualifications are needed to ensure that the KKT conditions hold in an NPO problem. We provide three: the existence of a sum of weighted squares resolution of the problem and the non-commutative analogs of Linear Independence Constraint Qualification and Mangasarian-Fromovitz Constraint Qualification. We also present sufficient conditions to justify enforcing the KKT conditions partially. We test the power of the non-commutative KKT conditions by computing local properties of ground states of many-body spin systems and the maximum quantum violation of Bell inequalities.

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 equalities 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. Not surprisingly, in the last decades, NPO has found application in quantum information theory, quantum chemistry and condensed matter physics. 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 behaviour 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, which differs from the original method by adding a collection of extra constraints. 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 the KKT conditions to the non-commutative setting, where they 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.

The 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 IV, 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 non-commutative KKT optimality conditions hold. While the state optimality conditions are easily seen to hold in all NPO problems, justifying the corresponding operator optimality conditions (or non-commutative KKT conditions) requires more work. In this context, we show that, if the solution of the NPO problem is achieved at a finite level of the original SDP hierarchy, then the NPO problem admits the non-commutative KKT conditions. We also generalize two well-known classical criteria: Linear Independence Constraint Qualification (LICQ) and Mangasarian-Fromovitz Constraint Qualification (MFCQ) [21].

Classical LICQ requires that the gradients of the active constraints are linearly independent at the solution of the problem. We generalize this result, showing that a non-commutative analog of LICQ *essentially* implies the KKT conditions, i.e., an infinitesimal modification of the KKT conditions holds. This modification is necessary to prevent the occurrence of variables with infinite value in the corresponding semidefinite programs.

Classical MFCQ demands the gradients of the equality constraints to be independent at the solution and also the existence of a vector with a positive overlap with the gradients of all the active inequality constraints. Contrarily to LICQ, MFCQ allows one to bound the value of the Lagrange multipliers. We find that the non-commutative analog of MFCQ retains this 'boundedness' property. This allows us to prove that, under non-commutative MFCQ, the KKT conditions hold, without the need to introduce any infinitesimal changes.

Since the NPO formulation of quantum nonlocality does not seem to satisfy either criterion, we also provide a necessary condition that guarantees that a partial version of the KKT conditions holds. 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 are convex for each of these parts, then a relaxed form of the KKT condition holds. This allows us to enforce 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, we propose a generalization of the KKT conditions for the non-commutative framework, which will allow us to incorporate extra constraints into our optimization problems. In Sec. IV we prove the general necessity of one such constraint, the so-called state optimality conditions. We show that these conditions, on their own, can be used to extract ground state properties of many-body quantum systems. In the next two sections, we proceed to prove analytically when it is legitimate to enforce the new optimality conditions fully (Sec. V) or partially (Sec. VI). In Sec. VII we illustrate the power of KKT conditions in NPO by computing the maximum violation of bipartite Bell inequalities. We then present our conclusions.

While conducting this research, we found that Fawzi *et al.* [22] had independently arrived at the state optimality conditions (11d). In their interesting pre-print, 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 IV 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 if and only 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:*

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

where the minimization takes place over all Hilbert spaces $\mathcal{H}$, normalized states $\sigma : B(\mathcal{H}) \to \mathbb{C}$ 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 many-body quantum system [29] are examples of NPO problems.

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

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

In the following, we denote by $\sigma^\star$ the minimizer of this problem. 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 $\sigma^\star$ as above, 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 vector $\psi^\star\in\mathcal{H}^\star$ such that

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

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:

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

The relaxation (4) can be cast as a semidefinite program [15, 32] with $|\mathcal{P}^{2k}|$ free complex variables. Its dual problem is of the form:

$$\begin{aligned} q^k := &\max\theta \\ \text{s.t. } & \exists\{s_j\}_j\subset\mathcal{P}^k,\ \{s_{il}\}_{il}\subset\mathcal{P}:\ 2\deg(s_{il})\leq 2k-\deg(g_i),\ \{s^+_{jl}\}_{jl},\ \{s^-_{jl}\}_{jl}\subset\mathcal{P}:\ \deg(s^+_{jl})+\deg(s^-_{jl})\leq 2k-\deg(h_j), \\ & f-\theta=\sum_l s_l^*s_l+\sum_{il}s^*_{il}g_is_{il}+\sum_{jl}s^+_{jl}h_js^-_{jl}. \end{aligned} \tag{5}$$

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 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*

$$p=\sum_j s_j^*s_j+\sum_{il}s^*_{il}g_is_{il}+\sum_{jl}s^+_{jl}h_js^-_{jl}. \tag{6}$$

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 (5) 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). What is clear is that $p^1\leq p^2\leq\cdots\leq p^\star$, and similarly for the q's.

Under which circumstances are these hierarchies complete, in the sense that $\lim_{n\to\infty}p^n=\lim_{n\to\infty}q^n=p^\star$? In Ref. [11], it is proven that it is enough that Problem (2) satisfies the *Archimedean property.*

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

$$K-\sum_i x_i^2 \tag{7}$$

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 in bounded, a relation of the form of (7) can be added to the inequality constraints without changing the problem. In the following we shall thus assume that the Archimedean property holds for all problems we consider.

III. KKT CONDITIONS

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

$$\begin{aligned} p^\star := &\min f(x) \\ &\text{s.t. } g_i(x) \geq 0, \quad i = 1, \ldots, m, \\ &\quad\; h_j(x) = 0, \quad j = 1, \ldots, m', \end{aligned} \tag{8}$$

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:

$$\begin{aligned} &\exists \{\mu_i\}_i \subset \mathbb{R}^+, \ \{\lambda_j\}_j \subset \mathbb{R}, \\ \text{such that } &\partial_x f = \sum_{i \in \mathbb{A}} \mu_i \partial_x g_i + \sum_j \lambda_j \partial_x h_j, \\ &\mu_i g_i(x) = 0, \quad \forall i \in \mathbb{A}. \end{aligned} \tag{9}$$

Here, the set of *active constraints* $\mathbb{A}$ 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 [17, 18]. This has led some authors to enforce these conditions implicitly to numerically solve polynomial optimization problems [33, 34].

In this work, we wish to do the same for non-commutative optimization problems. To formulate the KKT conditions for problems of the form (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' $\bar{x} = (\bar{x}_1, \ldots, \bar{x}_n)$. 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 on ϵ. Informally:*

$$\nabla_x p(\bar{x}) := \lim_{\epsilon \to 0} \frac{p(x + \epsilon \bar{x}) - p(x)}{\epsilon}. \tag{10}$$

With this definition, we propose the following generalization of the KKT conditions for Problem (2):

Definition 5. *We say that the NPO Problem (2) satisfies the non-commutative Karush-Kuhn-Tucker (KKT) conditions if, by adding the constraints*

$$\begin{aligned} \exists \mu_i : \mathcal{P} \to \mathbb{C}, \ &\mu_i(pp^*) \geq 0, \quad \forall p \in \mathcal{P}, \ i = 1, \ldots, m, \\ &\mu_i(p g_l p^*) \geq 0, \quad \forall p \in \mathcal{P}, \ i = 1, \ldots, m, \ l = 1, \ldots, m, \\ &\mu_i(s^+ h_j s^-) = 0, \quad \forall s^+, s^- \in \mathcal{P}, \ j = 1, \ldots, m, \\ \exists \lambda_j : \mathcal{P} \to \mathbb{C}, \ &\lambda_j(p + p^*) \in \mathbb{R}, \quad \forall p \in \mathcal{P}, \ j = 1, \ldots, m', \\ &\lambda_j(s^+ h_j s^-) = 0, \quad \forall s^+, s^- \in \mathcal{P}, \ j = 1, \ldots, m', \end{aligned} \tag{11a}$$

$$\mu_i(g_i) = 0, \quad i = 1, \ldots, m, \tag{11b}$$

$$\sigma\left(\nabla_x f(p)\right) - \sum_i \mu_i \left(\nabla_x g_i(p)\right) - \sum_j \lambda_j \left(\nabla_x h_j(p)\right) = 0, \quad \forall p \in \mathcal{P}, \tag{11c}$$

$$\sigma([f, p]) = 0, \sigma\left(p^* f p - \frac{1}{2}\{f, p^* p\}\right) \geq 0, \quad \forall p \in \mathcal{P}, \tag{11d}$$

the solution of Problem (2) does not change.

The notion of 'active constraints' in the definition above corresponds to the demand that the 'state variables' $\{\mu_i\}_i$ satisfy $\mu_i(g_i) = 0$ for all i. The logic is that an operator inequality constraint of the form $A(X) \geq 0$ can be interpreted as $\langle\phi|A(X)|\phi\rangle \geq 0$ for all ϕ. From this perspective, the active constraints of $A(X) \geq 0$ are $\{\mu(A(X)) : \mu \geq 0,\ \mu(A(X)) = 0\}$.

Conditions (11a), (11b), (11c) and (11d) are heuristically derived from the Lagrangian

$$\mathcal{L} = \sigma(f(\bar{X})) - \sigma(M) + \alpha(1 - \sigma(1)) - \sum_i \mu_i(g_i(\bar{X})) - \sum_j \lambda_j(h_j(\bar{X})), \tag{12}$$

where the operator variables $\bar{X} = (\bar{X}_1, \ldots, \bar{X}_n)$ act on $\mathcal{H}^\star$, the Hilbert space where the solution of Problem (1) lives, and the multipliers α, M, $\{\mu_i\}_i$, $\{\lambda_j\}_j$ respectively denote a real variable, a positive semidefinite operator, a set of non-normalized states and a set of linear functionals on $B(\mathcal{H}^\star)$.

Eqs. (11a), (11b), (11c), dubbed *operator optimality conditions*, appear when we consider variations on the optimal operators $X_1^\star, \ldots, X_n^\star$. They respectively correspond to the classical conditions of dual feasibility, complementary slackness and stationarity. Condition (11d), which we call *state optimality*, is a bit more mysterious, and requires some explanation.

Varying the state $\sigma : B(\mathcal{H}^\star) \to \mathbb{C}$ in (12) from the optimal $\sigma^\star$ to $\sigma^\star + \delta\sigma$ easily leads to the condition

$$f(X^\star) - \alpha\mathbb{I} = M \geq 0. \tag{13}$$

On the other hand, complementary slackness implies that $\sigma^\star(M) = 0$. This suggests that the optimal state $\sigma^\star$ is an eigenvector of $f(X^\star)$ with minimum eigenvalue $\alpha = p^\star$. In that case, letting $H = f(X^\star)$, we have

$$\sigma^\star(H\bullet) = \sigma^\star(\bullet H) = p^\star\sigma^\star(\bullet). \tag{14}$$

This implies that $\sigma^\star([H, \bullet]) = 0$. In addition, the condition $f(X^\star) - p^\star \geq 0$ is equivalent to

$$\sigma^\star(q^*(f(X^\star) - p^\star)q) \geq 0, \quad \forall q \in B(\mathcal{H}^\star). \tag{15}$$

Thanks to relation (14), the second term of the left-hand side of the above equation can be written as $\frac{1}{2}\sigma^\star(\{f(X^\star), q^*q\})$. Putting everything together, we have that, for variations of the form $\delta\sigma(\bullet) = \sigma(q^* \bullet q) - \sigma^\star(q^*q)\sigma(\bullet)$, the state optimality conditions imply eq. (11d).

If we were guaranteed that, for the NPO Problem (2), the non-commutative KKT conditions held, then we could add some further constraints to our SDP relaxation (4), namely:

$$\begin{aligned}
&\exists\mu_i^k : \mathcal{P}^{2k} \to \mathbb{C},\, \mu_i^k(pp^*) \geq 0, \quad \forall p \in \mathcal{P}^k,\ i = 1, \ldots, m,\\
&\qquad\qquad \mu_i^k(pg_lp^*) \geq 0, \quad \forall p \in \mathcal{P},\ \deg(p) \leq k - \left\lceil \frac{\deg(g_l)}{2} \right\rceil,\ i = 1, \ldots, m,\ l = 1, \ldots, m,\\
&\qquad\qquad \mu_i^k(s^+h_js^-) = 0, \quad \forall s^+, s^- \in \mathcal{P},\ \deg(s^+) + \deg(s^-) \leq 2k - \deg(h_j),\ j = 1, \ldots, m,\\
&\qquad\qquad \mu_i^k(g_i) = 0, \quad i = 1, \ldots, m,\\
&\exists\lambda_j^k : \mathcal{P}^{2k} \to \mathbb{C},\, \lambda_j^k(p + p^*) \in \mathbb{R}, \quad \forall p \in \mathcal{P}^{2k},\ j = 1, \ldots, m',\\
&\qquad\qquad \lambda_j^k(s^+h_js^-) = 0, \quad \forall s^+, s^- \in \mathcal{P},\ \deg(s^+) + \deg(s^-) \leq 2k - \deg(h_j),\ j = 1, \ldots, m',\\
&\sigma^k\left(\nabla_x f(t(x))\right) - \sum_i \mu_i^k\left(\nabla_x g_i(t(x))\right) + \sum_j \lambda_j^k\left(\nabla_x h_j(t(x))\right) = 0, \quad \forall t \in \mathcal{P}^{\times n},\\
&\qquad\qquad \deg\left(\nabla_x f\right), \deg\left(\nabla_x g_i\right), \deg\left(\nabla_x h_j\right) \leq 2k - \deg(t),\\
&\qquad\qquad \sigma^k([f, p]) = 0, \quad \forall p \in \mathcal{P},\ \deg(f) + \deg(p) \leq 2k,\\
&\qquad\qquad \sigma^k(p^*fp - \frac{1}{2}\{f, p^*p\}) \geq 0, \quad \forall p \in \mathcal{P},\ 2\deg(p) + \deg(f) \leq 2k,
\end{aligned} \tag{16}$$

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, $\deg(t) = \max_k \deg(t_k)$.

To implement such conditions for all polynomials, it suffices to enforce them for a basis 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

$$\begin{aligned}
\left(M^k(\mu_i)\right)_{ab} &:= \mu_i(o_a^* o_b),\\
\left(M_l^k(\mu_i)\right)_{ab} &:= \mu_i((o_a^l)^* g_l o_b^l),
\end{aligned} \tag{17}$$

are respectively called the k^{th}*-order moment matrix of* μ_i and the k^{th} *order localizing matrix of* μ_i *for constraint* g_l [11]. Enforcing the first two lines of (16) boils down to demanding the moment matrix $M^k(\mu_i)$ and the localizing matrices $\{M_l^k(\mu_i)\}_l$ be positive semidefinite.

Similarly, condition $\sigma^k([f,p]) = 0$ becomes $\sigma^k([f,o]) = 0$ for all monomials o of appropriate degree, and condition $\sigma^k(p^* f p - \frac{1}{2}\{f, p^* p\}) \geq 0$ is equivalent the positive semidefiniteness of the matrix γ with elements

$$\gamma_{ij} := \sigma^k(o_i^* f o_j - \frac{1}{2}\{f, o_i^* o_j\}), \tag{18}$$

for a basis $\{o_j\}_j$ of monomials of degree $k - \left\lceil \frac{\deg(f)}{2} \right\rceil$.

Enforcing linear conditions over n-tuples of polinomials $t \in \mathcal{P}^{\times n}$, such as the stationarity condition in (16), just requires taking t to be of the form $t_k = \delta_{k,l} o$, with o being a monomial.

As we will see, in some situations, the extra conditions (16) boost the speed of convergence of the SDP hierarchy (4). The purpose of the next two sections is to elucidate when it is legitimate to enforce them fully or, at least, partially.

IV. STATE OPTIMALITY CONDITIONS

The state optimality conditions (11d) are trivially equivalent to demanding that $\sigma^\star$ is an eigenvector of $f(X^\star)$ with eigenvalue equal to the bottom of the spectrum of $f(X^\star)$. As it turns out, whenever the considered NPO problem is Archimedean, there exists a minimizer $(\mathcal{H}^\star, \sigma^\star, X^\star)$ for the objective function, with $\sigma^\star(\bullet) = \langle\psi^\star| \bullet |\psi^\star\rangle$, for some $|\psi^\star\rangle \in \mathcal{H}^\star$ [11].

It follows that $|\psi^\star\rangle$ 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 state $\psi \in \mathcal{H}^\star$ with $\langle\psi|f(X^\star)|\psi\rangle < \langle\psi^\star|f(X^\star)|\psi^\star\rangle = p^\star$. This would contradict the assumption that $(\mathcal{H}^\star, \sigma^\star, X^\star)$ is a minimizer of the problem.

The state optimality conditions (11d) can therefore be assumed in any Archimedean NPO.

Proposition 6. *If Problem (2) is Archimedean, then its optimizer $\sigma^\star$ satisfies the state optimality conditions* (11d).

In addition, conditions (11d) 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. Let $\mathrm{Gr}(H)$ denote the set of states σ such that $\sigma(H - E_0(H)) = 0$. Then, the state optimality conditions (11d) allow us to solve optimization problems of the form

$$\begin{aligned} p^\star := & \min_{\mathcal{H}, X, \sigma} \sigma(f(X)) \\ \text{s.t. } & g_i(X) \geq 0, \quad i = 1, \ldots, m, \\ & h_i(X) = 0, \quad i = 1, \ldots, m', \\ & \sigma \in \mathrm{Gr}(b_k(X)), \quad k = 1, \ldots, m''. \end{aligned} \tag{19}$$

A. Application: many-body quantum systems

The computation of the properties of condensed matter systems at zero temperature also 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*:

$$\begin{aligned} & (\sigma_x^j)^2 = (\sigma_y^j)^2 = (\sigma_z^j)^2 = 1, \\ & \sigma_x^j \sigma_y^j - i\sigma_z^j = \sigma_y^j \sigma_z^j - i\sigma_x^j = \sigma_z^j \sigma_x^j - i\sigma_y^j = 0. \end{aligned} \tag{20}$$

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

$$[\sigma_a^j, \sigma_b^k] = 0, \quad a, b \in \{x, y, z\},\ j \neq k. \tag{21}$$

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

$$H(\sigma) = \sum_{j>k}^{n} P_{jk}(\sigma^j, \sigma^k), \tag{22}$$

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

For large n, computing $E_0(H)$ is Quantum-Merlin-Arthur-hard (**QMA**-hard) [37]. Quantum chemists [3–5] (and, more recently, condensed matter physicists [9, 38–40]) use NPO to lower bound $E_0(H)$. In essence, they relax the problem

$$\begin{aligned}E_0(H) = &\min \rho(H)\\ &\text{s.t. } (\sigma_x^j)^2 = (\sigma_y^j)^2 = (\sigma_z^j)^2 = 1, \quad j = 1, \ldots, n,\\ &\qquad \sigma_x^j\sigma_y^j - i\sigma_z^j = \sigma_y^j\sigma_z^j - i\sigma_x^j = \sigma_z^j\sigma_x^j - i\sigma_y^j = 0, \quad j = 1, \ldots, n,\\ &\qquad [\sigma_a^j, \sigma_b^k] = 0, \quad a, b \in \{x, y, z\},\ j \neq k\end{aligned} \tag{23}$$

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 between $[M^-, M^+]$, with

$$M^\pm := \mp \min\{\mp\langle\psi|(\sum_j \sigma_z^j)|\psi\rangle : \langle\psi|\bullet|\psi\rangle \in \mathrm{Gr}(H)\}. \tag{24}$$

For instance, Wang *et al.* [40] 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:

$$\begin{aligned}\bar{M}^\pm = &\mp \min\{\mp\rho(\sum_j \sigma_z^j)\}\\ &\text{s.t. } (\sigma_x^j)^2 = (\sigma_y^j)^2 = (\sigma_z^j)^2 = 1, \quad j = 1, \ldots, n,\\ &\qquad \sigma_x^j\sigma_y^j - i\sigma_z^j = \sigma_y^j\sigma_z^j - i\sigma_x^j = \sigma_z^j\sigma_x^j - i\sigma_y^j = 0, \quad j = 1, \ldots, n,\\ &\qquad [\sigma_a^j, \sigma_b^k] = 0, \quad a, b \in \{x, y, z\},\ j \neq k,\\ &\qquad \rho(H) \leq E_0^+.\end{aligned} \tag{25}$$

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.* [40] 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 [40] 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 [41–43]). The results of [40] are not that good for 2D systems, for which current variational tools are very imprecise [44].

The state optimality condition (11d) allows us to formulate Problem (24) as the following NPO:

$$\begin{aligned}m^\pm = &\mp \min \rho(\mp \sum_j \sigma_z^j)\\ &\text{s.t. } (\sigma_x^j)^2 = (\sigma_y^j)^2 = (\sigma_z^j)^2 = 1, \quad j = 1, \ldots, n,\\ &\qquad \sigma_x^j\sigma_y^j - i\sigma_z^j = \sigma_y^j\sigma_z^j - i\sigma_x^j = \sigma_z^j\sigma_x^j - i\sigma_y^j = 0, \quad j = 1, \ldots, n,\\ &\qquad [\sigma_a^j, \sigma_b^k] = 0, \quad a, b = x, y, z; \quad j \neq k,\\ &\qquad \rho \in \mathrm{Gr}(H).\end{aligned} \tag{26}$$

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

$$\sigma([H, p]) = 0 \tag{27}$$

and

$$\sigma\left(p^* H p - \frac{1}{2}\{H, p^* p\}\right) \geq 0. \tag{28}$$

The advantage of the formulation (26) with respect to (25) is that it does not require any upper bound on $E_0(H)$. Problem (26) is thus appropriate to tackle 2D and 3D systems, and even spin glasses [45].

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 via the software Moment [46].

In the 1D case, the Hamiltonian reads

$$H = \frac{1}{4}\sum_{i=0}^{n-1}\sum_{a\in\{x,y,z\}}\sigma_a^i\sigma_a^{i\oplus 1}, \tag{29}$$

where addition 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 (11d) is only a tightening of the SDP: without it the SDP would converge anyway to the ground state energy. In calculating the upper bound, however, the state optimality condition (11d) is doing all the work, as without it the SDP would converge to the maximum energy.

n	Lower bound	Exact value	Upper bound
6	-0.4671	-0.4671	-0.4463
7	-0.4251	-0.4079	-0.4009
8	-0.4564	-0.4564	-0.3973
9	-0.4416	-0.4219	-0.4037
10	-0.4516	-0.4515	-0.3917
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 $n = 13$ we use all nearest-neighbour monomials of degree up to 3, and for higher n all all nearest-neighbour monomials of degree up to 2.

The magnetization is given by

$$M = \sum_{i=0}^{n-1}\sigma_z^i. \tag{30}$$

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

$$H = \frac{1}{4}\sum_{i,j=0}^{L-1}\sum_{a\in\{x,y,z\}}\sigma_a^{i,j}(\sigma_a^{i\oplus 1,j} + \sigma_a^{i,j\oplus 1}), \tag{31}$$

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 (26) 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

n	Lower bound	Lowest exact value
6	-0.0022	0
7	-0.1469	-0.1429
8	-0.1036	0
9	-0.1393	-0.1111
10	-0.1358	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 $n = 13$ we use all nearest-neighbour monomials of degree up to 3, and for higher n all all nearest-neighbour monomials of degree up to 2.

n	Lower bound	Exact value	Upper bound
3^2	-0.4913	-0.4410	-0.2363

Table III. Ground state energy per site of 2D Heisenberg model. We used all nearest-neighbour monomials of degree up to 2.

the form:

$$H = \sum_{j=-\infty}^{\infty} P(\sigma^j, \sigma^{j+1}). \tag{32}$$

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

$$\pi_R(\sigma_a^j) = \sigma_a^{j+1}, \quad \pi_L(\sigma_a^j) = \sigma_a^{j-1}. \tag{33}$$

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

$$\begin{aligned}
e_0^n :=& \min \rho(P(\sigma^j, \sigma^{j+1}))\\
\text{such that } & (\sigma_x^j)^2 = (\sigma_y^j)^2 = (\sigma_z^j)^2 = 1, \quad j = 1, \dots, n,\\
& \sigma_x^j\sigma_y^j - i\sigma_z^j = \sigma_y^j\sigma_z^j - i\sigma_x^j = \sigma_z^j\sigma_x^j - i\sigma_y^j = 0, \quad j = 1, \dots, n,\\
& [\sigma_a^j, \sigma_b^k] = 0, \quad a, b \in \{x, y, z\},\ j \neq k\\
& \rho(p) = \rho(\pi_L(p)) = \rho(\pi_R(p)), \qquad \text{for } p, \pi_L(p), \pi_R(p) \in \mathcal{P}(\sigma^1, \dots, \sigma^n).
\end{aligned} \tag{34}$$

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 (27) [38, 47]. Note that, if $p \in \mathcal{P}(\sigma^2, \dots, \sigma^{n-1})$ has degree k, then the above commutator has degree $k + 1$ and only involves the variables $\sigma^1, \dots, \sigma^n$.

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

Things change dramatically when we add the optimality condition (28), for it also allows us to bound whatever local property of the system, such as the magnetization. For $1D$ Hamiltonians (32), if p has degree k and depends on the variables $\sigma^2, \dots, \sigma^{n-1}$, the polynomial in eq. (28) will be of degree $2k + 1$ and only depend on $\sigma^1, \dots, \sigma^n$. Thus, even though we are working in the thermodynamic limit, the state optimality condition can be evaluated. This is, in fact, the case for any translation-invariant scenario in arbitrarily many spatial dimensions.

For any local property o, the corresponding SDP hierarchies will converge to the exact interval of allowed values for o (at zero temperature). In the absence of numerical tests, though, we cannot say much about the speed of convergence of the hierarchy. How difficult is computing ground state properties of physically compelling Hamiltonians in the thermodynamic limit? In view of our numerical results for finite systems, one would expect to obtain reasonably good bounds, given reasonable computational resources.

n	Lower bound	Lower exact value
3^2	-0.3066	-0.1111

Table IV. Magnetization per site of 2D Heisenberg model. We used all nearest-neighbour monomials of degree up to 2.

V. OPERATOR OPTIMALITY CONDITIONS

Eqs. (11b-11c) are much more complicated, and can be justified just in some settings. This is analogous to the classical case, where the KKT conditions do not hold generally. However, provided that the constraints of the problem fulfill a *constraint qualification*, one can show that the KKT conditions hold (see Chapter 12 in Ref. [21]).

A. SOS constraint qualification

The first constraint qualification we provide demands the existence of an exact SOS resolution of the problem. This is the case if the quadratic module generated by the problem constraints (namely, the set of all SOS polynomials) is Archimedean closed [48].

Theorem 7. *Let the solution of Problem (2) be achieved at a finite level k of the hierarchy of SDP relaxations (5). Then, Problem (2) admits the KKT conditions.*

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

$$f - p^\star = \sum_l s_l s_l^* + \sum_{i,l} s_{il} g_i s_{il}^* + \sum_{j,l} s^+_{jl} h_j s^-_{jl}. \tag{35}$$

In addition, the minimizer $\sigma^\star$ of Problem (2) satisfies

$$\sigma^\star(f - p^\star) = 0. \tag{36}$$

It follows that

$$\begin{aligned} &\sigma^\star(s_l s_l^*) = 0, \quad \forall l, \\ &\sigma^\star(s_{il} g_i s_{il}^*) = 0, \quad \forall i, l. \end{aligned} \tag{37}$$

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

$$\sigma^\star(s_l q) = \sigma^\star(q s_l^*) = 0, \quad \forall l, \tag{38a}$$

$$\sigma^\star(s_{il} g_i q) = \sigma^\star(q g_i s_{il}^*) = 0, \quad \forall i, l. \tag{38b}$$

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

$$\begin{pmatrix} \sigma^\star(qq^*) & \sigma^\star(q s_l^*) \\ \sigma^\star(s_l q^*) & \sigma^\star(s_l s_l^*) \end{pmatrix}. \tag{39}$$

The second one, from the positive semidefiniteness of

$$\begin{pmatrix} \sigma^\star(q g_i q^*) & \sigma^\star(q g_i s_{il}^*) \\ \sigma^\star(s_{il} g_i q^*) & \sigma^\star(s_{ij} g_i s_{ij}^*) \end{pmatrix}. \tag{40}$$

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. (35). Taking into account eqs. (36), (38), and the chain rule of differentiation, the result is

$$\delta\sigma^\star(\nabla_x f(p(x))) + O(\delta^2) = \delta \sum_i \mu_i(\nabla_x(g_i(p(x))) + \delta \sum_j \lambda_j(\nabla_x(h_j(p(x))) + O(\delta^2), \tag{41}$$

where μ_i denotes the non-normalized state given by

$$\mu_i(p) := \sigma^\star(\sum_l s_{il} p s_{il}^*), \tag{42}$$

and λ_j is the linear functional

$$\lambda_j(p) := \sigma^\star(\sum_l s_{il}^+ p s_{il}^-). \tag{43}$$

Note that $\{\mu_i\}_i$, $\{\lambda_j\}_j$ inherit from $\sigma^\star$ the feasibility conditions (11a).

Collecting the terms in eq. (41) that depend linearly on δ, we have that

$$\sigma^\star(\nabla_x f(p(x))) = \sum_i \mu_i(\nabla_x(g_i(p(x))) + \sum_j \lambda_j(\nabla_x(h_j(p(x))). \tag{44}$$

This is condition (11c).

Finally, the states $\{\mu_i\}_i$ satisfy complementary slackness (11b), for

$$\mu_i(g_i) = \sum_l \sigma^\star(s_{il} g_i s_{il}^*) = 0, \quad \forall i, \tag{45}$$

by eq. (38). □

B. Linear independence constraint qualification

1. Classical LICQ and its non-commutative analog

For classical optimization problems, the linear independence of the gradients of the constraints (the so-called linear independence constraint qualification, or LICQ) implies that the KKT conditions must hold. These conditions amount to proving that, for the optimal solution $x^\star$ of the classical problem (8), the gradients of the equality constraints $\{\partial_x h_j\Big|_{x=x^\star} : j = 1, \ldots, m'\}$ and the active inequality constraints $\{\partial_x g_i\Big|_{x=x^\star} : i \in \mathbb{A}(x^\star)\}$ are linearly independent.

To ensure that LICQ holds, it is customary to prove that, for all feasible x, the vectors $\{\partial_x g_i, \partial_x h_j : i \in \mathbb{A}(x), j\}$ are indeed independent. Without loss of generality, let us assume that x is such that the first m'' inequality constraints are active. Then, gradient linear independence is equivalent to the existence of vectors $v_1(x), \ldots, v_{m''+m'}(x)$ such that

$$\begin{aligned} &\langle \partial_x g_i | v_k(x) \rangle = \delta_{i,k}, \quad i = 1, \ldots, m'', \\ &\langle \partial_x h_j | v_k(x) \rangle = \delta_{j+m'',k}, \quad j = 1, \ldots, m, \end{aligned} \tag{46}$$

for $k = 1, \ldots, m' + m''$. Now, define the matrices

$$\begin{aligned} &\hat{P}_i(x) := |v_i(x)\rangle\langle \partial_x g_i|, \quad i = 1, \ldots, m'', \\ &\hat{P}_{j+m''}(x) := |v_{j+m''}(x)\rangle\langle \partial_x h_j|, \quad j = 1, \ldots, m', \\ &\hat{P}_0(x) := \mathbb{I} - \sum_{k=1}^{m'+m''} \hat{P}_k(x). \end{aligned} \tag{47}$$

It is easy to see that they satisfy

$$\begin{aligned} &\sum_k \hat{P}_k(x) \cdot z = z, \quad \forall z \in \mathbb{R}^n, \\ &\partial_x g_i \cdot \hat{P}_k(x) \cdot z = 0, \quad \forall j \neq k, \qquad \partial_x h_j \cdot \hat{P}_k(x) z = 0, \quad \forall j + m'' \neq k, \\ &\partial_x g_i \cdot z = 0 \to \hat{P}_i(x) \cdot z = 0, \quad \partial_x h_j \cdot z = 0 \to \hat{P}_{j+m''}(x) \cdot z = 0. \end{aligned} \tag{48}$$

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

In non-commutative systems, variables form a vector $x = (x_1, \dots, x_n)$ of non-commuting objects, the gradient $\nabla_x g(\bullet)$ of a polynomial $g(x)$ is a linear map from n-tuples of polynomials $p = (p_1, \dots, p_n)$ to a single polynomial $\nabla_x g(p(x))$.

A non-commutative analog of relations (48) would thus demand the existence of $m+m'+1$ n-tuples of polynomials $P_0(x,z), \dots, P_{m+m'}(x,z)$ in the variables $x = (x_1, \dots, x_n)$, $z = (z_1, \dots, z_n)$, 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 (48). Correspondingly, the tuples $P_0, \dots, P_{m+m'}$ should satisfy the following conditions:

$$\sum_k P_k(X,Z) = Z, \tag{49a}$$

$$\mu\left(\nabla_x g_i(P_k(X,Z))\right) = 0, \quad \forall \text{ states } \mu,\ \mu(g_i) = 0,\ \forall i \neq k, \qquad \nabla_x h_j(P_k(X,Z)) = 0, \quad \forall j+m \neq k, \tag{49b}$$

$$\mu(\nabla_x g_i(Z)) = 0, \quad \forall \text{ states } \mu,\ \mu(g_i) = 0,\ \rightarrow\ P_i(X,Z) = 0, \qquad \nabla_x h_j(Z) = 0\ \rightarrow\ P_{j+m}(X,Z) = 0. \tag{49c}$$

Note that the non-commutative version of (48) contains the full m inequality constraints. The reason is that, as we did in the definition of the non-commutative KKT conditions (Def. 5), in eq. (49c) we model the active constraints of an operator inequality of the form $g(x) \geq 0$ through states μ such that $\mu(g(x)) = 0$.

Definition 8. *Let $\mathcal{J}$ be the ideal generated by $\{h_j\}_j$. It consists of all polynomials of the form*

$$\sum_{j,l} v_{jl}(x) h_j(x) w_{jl}(x). \tag{50}$$

Similarly, call $\mathcal{J}^Z$ the set of polynomials $q(x,z)$, linear on $z = (z_1, \dots, z_n)$, of the form

$$q(x,z) = \sum_{j,l} p^+_{jl}(x,z) h_j(x) p^-_{jl}(x,z). \tag{51}$$

The algebraic version of constraints (49a-49c) defines the non-commutative linear independence constraint qualification (ncLICQ).

Definition 9. *An NPO Problem (2) satisfies non-commutative LICQ if there exist n-tuples of symmetric polynomials in $2n$ variables $P_0(x,z), P_1(x,z), \dots, P_{m+m'}(x,z)$, linear in the z variables, such that*

$$\begin{aligned}
&\sum_{j=0}^{m+m'} P_j(x,z) - z \in (\mathcal{J}^Z)^{\times n},\\
&\forall i \neq k\ \exists s_i(x,z) \text{ such that } \nabla_x g_i(P_k(x,z)) + s_i(x,z) g_i(x) + g_i(x) s_i(x,z)^* \in \mathcal{J}^Z,\\
&\nabla_x h_j(P_k(x,z)) \in \mathcal{J}^Z, \quad \forall k \neq j+m,\\
&\exists \alpha^+, \alpha^- \text{ such that } (P_i(x,z))_k - \sum_l \alpha^+_{ikl}(x) \nabla_x g_i(z) \alpha^-_{ikl}(x) \in \mathcal{J}^Z, \quad i = 1, \dots, m, \quad k = 1, \dots, n,\\
&\exists \beta^+, \beta^- \text{ such that } (P_{j+m}(x,z))_k - \sum_l \beta^+_{jkl}(x) \nabla_x h_j(z) \beta^-_{jkl}(x) \in \mathcal{J}^Z, \quad j = 1, \dots, m', \quad k = 1, \dots, n.
\end{aligned} \tag{52}$$

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

Remark 1. *Conditions (49a-49c) are required to hold only when z is an n-tuple of polynomials on x. In this regard, for some NPO problems, one can further relax the definition of ncLICQ provided above. For instance, if there exists a polynomial $r(x)$ such that $[r(x), x_i] \in \mathcal{J}$, for $i = 1, \dots, n$, then it will hold that $[r(x), z_i] = 0$, for $i = 1, \dots, n$. Consequently, one can redefine the set $\mathcal{J}^Z$ in eq. (51) to also include polynomials of the form $\{s^+(x)[r(x), z_i]s^-(x)\}_i$. The proofs of Theorems 10, 15 below follow through with such a modified definition.*

Example 1. *Consider an NPO with no inequality constraints ($m = 0$) and the following equality constraints:*

$$0 = h_j(x) := x_j^2 - 1, \quad j = 1, \dots, n. \tag{53}$$

This system satisfies the LICQ. Indeed, define

$$\begin{aligned}
(P_j(x,z))_k &:= \delta_{jk} \frac{1}{2}(z_j + x_j z_j x_j),\\
(P_0(x,z))_k &:= \frac{1}{2}(z_k - x_k z_k x_k).
\end{aligned} \tag{54}$$

It can be easily verified that the n-tuples of polynomials $P_0, \dots, P_n$ satisfy conditions (52).

Example 2. *Consider an NPO with no equality constraints ($m'=0$) and the following inequality constraints:*

$$0\leq g_i^+(x):=1+x_i,\ 0\leq g_i^-(x):=1-x_i,\quad i=1,\ldots,n. \tag{55}$$

This system also satisfies the LICQ, with 'projective' polynomials:

$$\begin{aligned}
&(P_i^+(x,z))_k:=\delta_{ik}\frac{1}{2}\{z_i,(1-x_i)\},\\
&(P_i^-(x,z))_k:=\delta_{ik}\frac{1}{2}\{z_i,(1+x_i)\},\\
&(P_0(x,z))_k:=0.
\end{aligned} \tag{56}$$

2. *The effect of ncLICQ*

In this section, we will prove that, if an Archimedean NPO problem satisfies ncLICQ, then it *essentially* satisfies the operator opttimality conditions. More specifically:

Theorem 10. *Consider a NPO (2) that satisfies both ncLICQ and the Archimedean condition, and let $\sigma^\star$ be one of its minimizers. Then, 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*

$$\begin{aligned}
&\mu_i^k(pp^*)+\epsilon_i\|p\|_2^2\geq 0,\quad \forall p\in\mathcal{P},\ \deg(p)\leq k,\\
&\mu_i^k(pg_lp^*)+\epsilon_i\|p\|_2^2\geq 0,\quad \forall p\in\mathcal{P},\ \deg(p)\leq k-\left\lceil\frac{\deg(g_l)}{2}\right\rceil,\ i=1,\ldots,m,\ l=1,\ldots,m,\\
&\mu_i^k(s^+h_js^-)=0,\quad \forall s^+,s^-\in\mathcal{P},\ \deg(s^+)+\deg(s^-)\leq 2k-\deg(h_j),\ j=1,\ldots,m,\\
&\mu_i^k(g_i)=0,\quad i=1,\ldots,m,\\
&\lambda_j^k(s^+h_js^-)=0,\quad \forall s^+,s^-\in\mathcal{P},\ \deg(s^+)+\deg(s^-)\leq 2k-\deg(h_j),\ j=1,\ldots,m',\\
\sigma^\star\left(\nabla_xf(t(x))\right)-\sum_i\mu_i^k&\left(\nabla_xg_i(t(x))\right)-\sum_j\lambda_j^k\left(\nabla_xh_j(t(x))\right)=0,\quad \forall t\in\mathcal{P}^{\times n},\\
&\deg\left(\nabla_xf\right),\deg\left(\nabla_xg_i\right),\deg\left(\nabla_xh_j\right)\leq 2k-\deg(t),
\end{aligned} \tag{57}$$

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

Moreover, for any $i\in\{1,\ldots,m\}$ for which there exist $p_i,s_i\in\mathcal{P}$, $r_i\in\mathbb{R}^+$ such that the polynomial

$$\nabla_xg_i(p_i)+s_ig_i+g_is_i^*-r_i \tag{58}$$

is an SOS, one can take $\epsilon_i=0$.

Remark 2. *Taking $\epsilon_i=0$ for all i, the SDP constraints (57) reduce to the conditions (16) that operator optimality implies on the SDP relaxation (4). The reason why the theorem demands non-zero $\{\epsilon_i\}$ is to prevent that $\mu_i^k(p)=\infty$ for some polynomnials $p\in\mathcal{P}^{2k}$. As we will see, condition (58) implies the existence of a bounded functional $\mu_i:\mathcal{P}\to\mathbb{C}$; thus one can take $\epsilon_i=0$.*

Remark 3. *For non-zero ϵ_i, enforcing the first two lines of (57) amounts to demanding the k^{th}-order moment and localizing matrices (see (17) for a definition) $M^k(\mu_i),M_l^k(\mu_i)$ to satisfy*

$$M_i^k(\mu_i)+\epsilon_i\mathbb{I}\geq 0,\ M_l^k(\mu_i)+\epsilon_i\mathbb{I}\geq 0,\quad l=1,\ldots,m. \tag{59}$$

To prove Theorem 10 (and, later, Theorem 15), we will make use of the following technical result.

Lemma 11. *Let the n-tuple of symmetric polynomials $q(x)$ satisfy*

$$\nabla_xh_j(q(x))\in\mathcal{J},\quad j=1,\ldots,m', \tag{60}$$

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

$$h_j(X^\star)=0,\quad j=1,\ldots,m'. \tag{61}$$

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

$$\begin{aligned} & h_j(X(t)) = 0, \quad j = 1, \ldots, m', \ t \in [-\epsilon, \epsilon], \\ & X(0) = X^\star, \\ & \left.\frac{dX(t)}{dt}\right|_{t=0} = q(X^\star). \end{aligned} \tag{62}$$

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

$$s_i(x)g_i(x) + g_i(x)s_i(x)^* + \nabla_x g_i(q(x)), \quad i = 1, \ldots, m, \tag{63}$$

are sums of squares, then the trajectory $X(t)$ also satisfies:

$$g_i(X(t)) \geq 0, \quad i = 1, \ldots, m, \tag{64}$$

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

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

$$\begin{aligned} & \frac{dX(t)}{dt} = q(X), \\ & X(0) = X^\star. \end{aligned} \tag{65}$$

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 (62).

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 (60), we have that

$$\begin{aligned} \frac{dh_j(X(t))}{dt} &= \nabla_x h_j \left(\frac{dX(t)}{dt}\right) = \nabla_x h_j(q(X(t))) \\ &= \sum_{j',l} r^+_{jj'l}(X(t))h_{j'}(X(t))r^-_{jj'l}(X(t)). \end{aligned} \tag{66}$$

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

$$\begin{aligned} & \frac{dH_j(t)}{dt} = \sum_{j',l} r^+_{jj'l}(X(t))H_{j'}(t)r^-_{jj'l}(X(t)), \quad j = 1, \ldots, m', \\ & H_j(0) = h_j(X(0)) = h_j(X^\star) = 0, \quad j = 1, \ldots, m'. \end{aligned} \tag{67}$$

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

$$H_j(k+1; \Delta) = H_j(k; \Delta) + \Delta \sum_{j',l} r^+_{jj'l}(X(k\Delta))H_{j'}(k; \Delta)r^-_{jj'l}(X(k\Delta)), \quad j = 1, \ldots, m'. \tag{68}$$

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

$$0 = H_j(t) = h_j(X(t)), \tag{69}$$

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. (62).

Similarly, let us assume that eq. (64) 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:

$$\frac{d\omega_i(t)}{dt} = \omega_i(t)s_i(X(t)), \quad \omega_i(0) = \mathbb{I}. \tag{70}$$

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

$$\begin{aligned}
\frac{dG_i(t)}{dt} &= \omega(t)(s_i(X(t))g_i(X(t)) + g_i(X(t))s_i^*(X(t)) + \nabla_x g_i(q(X(t))\omega_i(t)^* \\
&\overset{(63)}{=} \omega_i(t)\left(\sum_l s_{il}^*(X(t))s_{il}(X(t)) + \sum_{i'l} s_{ii'l}^*(X(t))g_{i'}(X(t))s_{ii'l}(X(t)) + \sum_{jl} s_{ijl}^+(X(t))h_j(X(t))s_{ijl}^-(X(t))\right)\omega_i(t)^* \\
&\overset{H(t)=0}{=} \omega_i(t)\left(\sum_l s_{il}^*(X(t))s_{il}(X(t)) + \sum_{i'l} s_{ii'l}^*(X(t))\omega_{i'}(t)^{-1}G_{i'}(t)(\omega_{i'}(t)^{-1})^* s_{ii'l}(X(t))\right)\omega_i(t)^*.
\end{aligned} \tag{71}$$

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

$$\begin{aligned}
&G_i(k+1;\Delta) = G_i(k;\Delta) \\
&+\Delta\omega_i(\Delta k)\left(\sum_l s_{il}^*(X(\Delta k))s_{il}(X(\Delta k)) + \sum_{i'l} s_{ii'l}^*(X(\Delta k))\omega_{i'}(\Delta k)^{-1}G_{i'}(k;\Delta))(\omega_{i'}(\Delta k)^{-1})^* s_{ii'l}(X(\Delta k))\right)\omega_i(\Delta k)^*.
\end{aligned} \tag{72}$$

Clearly, starting from positive semidefinite operators

$$G_i(0;\Delta) = g_i(X^\star) \geq 0, \quad i = 1, \ldots, m, \tag{73}$$

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. (64) holds. □

Proof of Theorem 10. Let $\sigma^\star$ be the minimizer of Problem (2), and suppose that ncLICQ holds. First, we define the space of polynomials generated by the gradients of the constraints $\{h_j\}_j$,

$$\mathbb{S}_j := \text{span}\{\nabla_x h_j(p) : p \in \mathcal{P}\}. \tag{74}$$

Next, we define the linear functionals $\lambda_j : \mathbb{S}_j + \mathcal{J} \to \mathbb{C}$ by the relation

$$\lambda_j(\nabla_x h_j(p) + \mathcal{J}) := \sigma^\star(\nabla_x f(P_{j+m}(x,p))). \tag{75}$$

Note that, by (52), $\nabla_x h_j(p) \equiv \nabla_x h_j(P_j(x, p(x))) \mod \mathcal{J}^Z$. The definition (75) is consistent. Indeed, suppose that there exists $p \in \mathcal{P}^{\times n}$, $q \in \mathcal{J}$ such that $\nabla_x h_j(p) + q = 0$. Then, by the last line of eq. (52), there exists $q' \in \mathcal{J}$ such that

$$\begin{aligned}
P_{j+m}(x,p) &= \sum_l \beta_{jl}^+ \nabla_x h_j(p)\beta_{jl}^- + q' \\
&= -\sum_l \beta_{jl}^+ q \beta_{jl}^- + q' =: q'' \in \mathcal{J}.
\end{aligned} \tag{76}$$

Thus,

$$\nabla_x f\left(P_{j+m}(x,p)\right) \in \mathcal{J}, \tag{77}$$

and so

$$\sigma^\star(\nabla_x f(P_{j+m}(x, p(x)))) = 0, \tag{78}$$

as desired.

Now, given a polynomial $p \in \mathcal{P}$, let us give it a seminorm through the relation

$$\|p\|_{\text{SOS}} := \inf\{K \in \mathbb{R}^+ : K^2 - p^*p \text{ admits an SOS decomposition}\}. \tag{79}$$

It turns out that the functional λ_j just defined is continuous with respect to the $\|\cdot\|_{\text{SOS}}$ seminorm. Indeed, due to the last line of (52) and relation (75), there exists $\Lambda \in \mathbb{R}^+$ such that, for all $p \in \mathbb{S}_j + \mathcal{J}$,

$$|\lambda_j(p)| \leq \Lambda \|p\|_{\text{SOS}}. \tag{80}$$

We next invoke the Hahn-Banach theorem (see, e.g., [49, Theorem III.6]) to extend each of these functionals λ_j to $\lambda_j : \mathcal{P} \to \mathbb{C}$ satisfying (80) for all $p \in \mathcal{P}$.

Let us now construct the 'state' functionals $\{\mu_i\}_i$. For every i, define the space of polynomials

$$\mathbb{T}_i' := \mathbb{T}_i + \mathcal{J} + \{sg_i + g_i s^* : s \in \mathcal{P}\}, \tag{81}$$

with

$$\mathbb{T}_i := \text{span}\{\nabla_x g_i(p) : p \in \mathcal{P}\}. \tag{82}$$

We define the linear functional $\mu_i : \mathbb{T}_i' \to \mathbb{C}$ through the relation

$$\mu_i \left(\nabla_x g_i(p) + q + sg_i + g_i s^*\right) := \sigma^\star(\nabla_x f(P_i(x,p))), \tag{83}$$

for any $q \in \mathcal{J}$. Again, this definition is consistent. Let $p \in \mathcal{P}^{\times n}$, $q \in \mathcal{J}$, $s \in \mathcal{P}$ such that

$$sg_i + g_i s^* + q + \nabla_x g_i(p) = 0. \tag{84}$$

This implies that the two polynomials

$$\pm(sg_i + g_i s^* + \nabla_x g_i(p)) \tag{85}$$

are sums of squares.

Let $(\mathcal{H}^\star, \psi^\star, X^\star)$ be the result of applying the GNS construction to $\sigma^\star$. Taking $q = P_i(x, \pm p(x))$ in Lemma 11, this means that there exist $\epsilon > 0$ and two feasible trajectories $\{X^\pm(t) : t \in [0, \epsilon]\}$ such that

$$X^\pm(0) = X^\star, \quad \left.\frac{dX^\pm(t)}{dt}\right|_{t=0} = \pm P_i(X^\star, p(X^\star)). \tag{86}$$

Now,

$$\begin{aligned} \left.\frac{d\psi^\star\left(f(X^\pm(t))\right)}{dt}\right|_{t=0} &= \psi^\star\left(\nabla_x f\left(\left.\frac{dX^\pm(t)}{dt}\right|_{t=0}\right)\right) \\ &= \psi^\star\left(\nabla_x f\left(\pm P_i(x,p)\right)\Big|_{x=X^\star}\right) = \pm\sigma^\star(\nabla_x f(P_i(x,p))). \end{aligned} \tag{87}$$

However, the left-hand side must be non-negative, because $X^\pm(0) = X^\star$ is minimal. It follows that

$$\sigma^\star(\nabla_x f(P_i(x,p))) = 0. \tag{88}$$

One can similarly show that the functionals μ_i are, in a sense, positive. Namely, for $p \in \mathbb{T}_i'$, p admitting an SOS decomposition, it is the case that $\mu_i(p) \geq 0$. Indeed, if

$$sg_i + g_i s^* + q + \nabla_x g_i(p) \tag{89}$$

is a sum of squares, then, by the first and second lines of eq. (52), there exist $s_i' \in \mathcal{P}$ such that

$$s_i' g_i + g_i (s_i')^* + \nabla_x g_i(P_i(x,p)) \tag{90}$$

is also SOS. Therefore, by Lemma 11, there exists a feasible trajectory $\{X(t) : t \in [0, \epsilon]\}$, satisfying

$$X(0) = X^\star, \quad \left.\frac{dX(t)}{dt}\right|_{t=0} = P_i(X^\star, p(X^\star)). \tag{91}$$

Again, by differentiating $\psi^\star(f(X(t))$ at $t = 0$, we find that the right-hand side of eq. (83) is non-negative.

Now, we claim that

$$\sigma^\star(\nabla_x f(p)) = \sum_i \mu_j(\nabla_x g_i(p)) + \sum_j \lambda_j(\nabla_x h_j(p)), \tag{92}$$

thus proving that the last line of eq. (57) holds.

Invoking the second and third lines of (52), this relation clearly holds for polynomials of the form $p = P_k(x, q)$, $k = 1, \ldots, m' + m$. By the first line of (52), it suffices to prove that the relation holds holds for $p = P_0(x, q)$ to show that it holds generally. In turn, this implies that we need to prove that

$$\sigma^\star(\nabla_x f(P_0(x, q))) = 0. \tag{93}$$

This again follows from Lemma 11. In this case, by eq. (52),

$$\nabla_x g_i(\pm P_0(x, q)) + s_i(x, \pm q(x))g_i + g_i s_i(x, \pm q(x))^* \in \mathcal{J}, \tag{94}$$

for all i. In particular, both polynomials are SOS. In addition, also by eq. (52), we have that

$$\nabla_x h_j(P_0(x, \pm q)) \in \mathcal{J}, \tag{95}$$

for all j. Thus, there exists trajectories $X^\pm(t) : t \in [0, \epsilon]$, with $X^\pm(0) = X^\star$ and $\frac{dX}{dt}\Big|_{t=0} = \pm P_0(X^\star, q(X^\star))$ such that

$$g_i(X^\pm(t)) \geq 0,\ h_j(X^\pm(t)) = 0, \quad \forall i, j. \tag{96}$$

Since both trajectories are feasible, then $\psi^\star(f(X^\pm(t)))$ must be minimal at $t = 0$. It thus follows that

$$0 \leq \frac{d\psi^\star(f(X^\pm(t)))}{dt}\Big|_{t=0} = \pm\sigma^\star\left(\nabla_x f(P_0(x, q(x)))\right), \tag{97}$$

and therefore eq. (93) holds.

It remains to extend the action of $\{\mu_i\}_i$ to $\mathcal{P}^{2k}$ in such a way that the constraints in (57) are satisfied. Given $k \in \mathbb{N}$, consider the semidefinite program:

$$\begin{aligned}
&\min \epsilon \\
\text{s.t. } & M^k(\mu_i^k) + \epsilon\mathbb{I} \geq 0, \\
& M_l^k(\mu_i^k) + \epsilon\mathbb{I} \geq 0, \quad l = 1, \ldots, m, \\
& \mu_i(sg_i + g_i s^* + \nabla_x g_i(p) + q) = \sigma^\star(\nabla_x f(P_i(x, p))), \quad \forall p, s \in \mathcal{P}, q \in \mathcal{J}, \\
& \deg(p) + \deg(\nabla_x g_i),\ \deg(s) + \deg(g_i),\ \deg(P_i(x, p)) + \deg(\nabla_x f) \leq 2k, \\
& \epsilon \geq 0.
\end{aligned} \tag{98}$$

where M^k ($\{M_l^k\}_l$) denotes (denote) the k^{th} order moment matrix (localizing matrices) of μ_i, see (17). The dual of the problem can be shown to be

$$\begin{aligned}
&\max -\mu_i(p) \\
\text{s.t. } & p \in \mathbb{T}_i', \\
& p - \mathrm{tr}\big((\vec{o}^T)^* Z\vec{o}\big) - \sum_l \mathrm{tr}\left((\vec{o^l}^T)^* Z_l \vec{o^l} g_l\right) \in \mathcal{J}, \\
& Z \geq 0,\ Z_l \geq 0, \quad l = 1, \ldots, m, \\
& \phi + \mathrm{tr}(Z) + \sum_{l=1}^m \mathrm{tr}(Z_l) = 1, \\
& \phi \geq 0,
\end{aligned} \tag{99}$$

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

That is, in this problem one needs to maximize $-\mu_i(p)$ over SOS polynomials $p \in \mathbb{T}_i'$ satisfying certain normalization constraints. Since $\mu_i(p) \geq 0$, for $p \in \mathbb{T}_i'$, p, SOS, the solution of the dual is zero, achieved by taking $Z = 0$, $Z_l = 0$ for all l and $\phi = 1$.

Now, the primal problem (98) 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 solution of (99) is also zero. This implies that one can find feasible points of problem (98) for any $\epsilon > 0$. Conditions (57) thus do not change the solution of problem (2).

We next prove that one can take $\epsilon_i = 0$ if eq. (58) holds. First, let us clarify what values the expression

$$\kappa := \mu_i \left(\nabla_x g_i(p_i) + s_i g_i + g s_i^*\right) \tag{100}$$

can take. By eq. (58), the polynomial on the right-hand side is an SOS that belongs to $\mathbb{T}_i'$, and thus $\kappa \geq 0$.

Suppose that $\kappa = 0$, and let $q \in \mathbb{T}_i'$ be symmetric. Due to the Archimedean condition, there exists $r \in \mathbb{R}^+$ such that $r - q$ is SOS. Thus, by (58), the polynomial

$$\frac{r}{r_i} \left(\nabla_x g_i(p_i) + s_i g_i + g s_i^*\right) - q \tag{101}$$

admits an SOS decomposition. The polynomial above also belongs to $\mathbb{T}_i'$, so we can evaluate it with μ_i, from which we infer that

$$-\mu_i(q) \geq 0. \tag{102}$$

Exploiting the fact that there exists $r' \in \mathbb{R}^+$ with $r' + q$ being an SOS, we correspondingly deduce that $\mu_i(q) \geq 0$. It thus follows that, if $\kappa = 0$, then $\mu_i = 0$, and so we can trivially extend the state functional μ_i to $\mathcal{P}$. In particular, we can take $\epsilon_i = 0$ in eq. (57).

Suppose, on the contrary, that $\kappa > 0$, and consider the convex cone of polynomials

$$T_i := \{p \in \mathbb{T}_i' : \mu_i(p) \geq 0\} + \{\sum_j q_j^* q_j + \sum_{j,l} q_{j,l}^* g_l q_{j,l} : l = 1, \ldots, m,\, q_j, q_{j,l} \in \mathcal{P}\}. \tag{103}$$

On one hand, this cone can be proven algebraically solid [48], i.e., there exists $t \in T_i$ such that, for all symmetric q, there exists $\delta > 0$, with $t + \delta' q \in T_i$, for all $0 < \delta' \leq \delta$. Indeed, due to the Archimedean condition, we can choose $t = 1$.

On the other hand, we know that $\mu_i(\nabla_x g_i(-p_i)) = -\kappa < 0$, and thus $\nabla_x g_i(-p_i) \notin T_i$. By the Eidelheit-Kakutani Separation theorem [50, Theorem III.3.2], we therefore have that there exists a non-zero functional $\tilde{\mu}_i$ that separates $\nabla_x g_i(-p_i)$ from the cone T_i. That is,

$$\tilde{\mu}_i(\nabla_x g_i(-p_i)) \leq 0,\ \tilde{\mu}_i(t) \geq 0, \quad \forall t \in T_i. \tag{104}$$

From the definition of T_i, the functional $\tilde{\mu}_i$ must be non-negative on SOS polynomials. In addition, for $p \in \mathbb{T}_i'$, $\mu_i(p) \geq 0$ implies that $\tilde{\mu}_i(p) \geq 0$.

Now, consider the functional

$$\hat{\mu}_i := \frac{\kappa}{\kappa'} \tilde{\mu}_i, \tag{105}$$

with $\kappa' := \tilde{\mu}_i(\nabla_x g_i(p_i)) > 0$ (if κ' were zero, then $\tilde{\mu}_i$ could be shown to vanish everywhere, as we saw it happened with μ_i). We claim that $\hat{\mu}_i$ is an extension of μ_i. Indeed, let $p \in \mathbb{T}_i'$. Then, the two polynomials in $\mathbb{T}_i'$

$$q^{\pm} := \pm \left(\frac{1}{\kappa} \mu_i(p) \nabla_x g_i(p_i) - p\right) \tag{106}$$

satisfy $\mu_i(q^{\pm}) = 0 \geq 0$. Thus,

$$0 \leq \hat{\mu}_i(q^{\pm}) = \pm \left(\mu_i(p) - \hat{\mu}_i(p)\right). \tag{107}$$

It follows that $\hat{\mu}_i(p) = \mu_i(p)$ for all $p \in \mathbb{T}_i'$, and, again, we can take $\epsilon_i = 0$ in eq. (57). $\square$

C. Mangasarian-Fromovitz constraint qualification

1. *Classical MFCQ and its non-commutative extension*

For classical optimization problems (8), Mangasarian-Fromovitz constraint qualification (MFCQ) reads [21]:

$$\begin{aligned} &\{\partial_x h_j\}_j, \text{ linearly independent}, \\ &\exists z \in \mathbb{R}^n \text{ with } \partial_x g_i \cdot z > 0, \quad i \in \mathbb{A}(x). \end{aligned} \tag{108}$$

MFCQ is known to be sufficient for the KKT conditions to hold. Contrary to LICQ, MFCQ allows one to bound the values of all Lagrange multipliers [21].

The non-classical analog of MFCQ is non-commutative Mangasarian-Fromovitz constraint qualification (ncMFCQ):

Definition 12. *Consider an NPO Problem (2). We say that the problem satisfies non-commutative Mangasarian-Fromovitz (ncMFCQ) constraint qualification if, on one hand, the equality constraints are linearly independent, i.e., if there exist symmetric n-tuples $P_0(x,z), P_1(x,z), \ldots, P_{m'}(x,z)$ with*

$$\sum_k P_k(x,z) - z \in (\mathcal{J}^Z)^{\times n}, \tag{109a}$$

$$\nabla_x h_j(P_k(x,z)) \in \mathcal{J}^Z, \quad \forall k \neq j,\ j = 1, \ldots, m', \tag{109b}$$

$$\exists \beta^+, \beta^- \text{ such that } (P_j(x,z))_k - \sum_l \beta^+_{jkl}(x) \nabla_x h_j(z) \beta^-_{jkl}(x) \in \mathcal{J}^Z. \tag{109c}$$

On the other hand, there exist $R, L \in \mathbb{R}^+$ and an n-tuple of polynomials $\hat{q}(x)$ such that

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

and, for $i = 1, \ldots, m$, the polynomial

$$L g_i(x) + \nabla_x g_i(\hat{q}(x)) - R \tag{111}$$

is SOS.

Remark 4. *If the goal is to generalize the condition $\partial_x g_i \cdot z > 0$, the reader might wonder why we demand the seemingly strong condition (111), instead of simply requiring that, for some $R' \in \mathbb{R}^+$, $s_i \in \mathcal{P}$, the polynomials*

$$\omega_i := s_i g_i(x) + g_i s_i + \nabla_x g_i(\hat{q}(x)) - R', \tag{112}$$

for $i = 1, \ldots, m$, are SOS. In fact, conditions (111), (112) are equivalent. The implication (111) $\Rightarrow$ (112) is obvious, so let us work out the opposite one. Suppose that ω_i is SOS. Let $K, \Lambda \in \mathbb{R}^+$ be such that the polynomials $K - s_i s_i^*$, $\Lambda - g_i$ are SOS for all i (the existence of K, Λ is a consequence of the Archimedean condition). Then, for any $\nu \in \mathbb{R}^+$, the polynomial

$$\begin{aligned}
&\omega_i + \left(\nu s_i - \frac{g_i}{\nu}\right)\left(\nu s_i - \frac{g_i}{\nu}\right)^* + \nu^2(K - s_i s_i^*) + \frac{1}{\Lambda \nu^2}\left(g_i(\Lambda - g_i) g_i + (g_i + \Lambda) g_i (g_i + \Lambda)\right) \\
&= \frac{\Lambda g_i}{\nu^2} + \nabla_x g_i(\hat{q}) - (R' - \nu^2 K)
\end{aligned} \tag{113}$$

is also SOS. Choosing ν small enough, we can make the last term be negative, thus recovering condition (111).

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 (16) is a bounded optimization problem.

Lemma 13. *Assume that Problem* (2) *satisfies the Archimedean condition and ncMFCQ. Then we can, without loss of generality, bound the SOS seminorm of the state multipliers μ_i and λ_j in the KKT conditions. In particular, we can find $K \in \mathbb{R}^+$ such that*

$$|\mu_i(p)|, |\lambda_j(p)| \leq K \|p\|_{\text{SOS}}, \quad \forall i, j, \tag{114}$$

where the $\|\cdot\|_{\text{SOS}}$ seminorm is defined in (79).

Proof. Take $p = \hat{q}$ in eq. (11c). Then we have that

$$\begin{aligned}
\sigma^\star\left(\nabla_x f(\hat{q}(x))\right) &\overset{(110)}{=} \sum_i \mu_i\left(\nabla_x g_i(\hat{q}(x))\right) \\
&\overset{(111)}{\geq} R \sum_i \mu_i(1) - L \sum_i \mu_i(g_i) \\
&\overset{(11b)}{=} R \sum_i \mu_i(1).
\end{aligned} \tag{115}$$

In turn, provided that the original NPO satisfies the Archimedean condition, there exists $K \in \mathbb{R}^+$ such that

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

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 $\frac{K}{R}$.

For any $s \in \mathcal{P}$, taking $p = P_j(x, s(x))$ in eq. (11c) implies that

$$\sigma^\star\left(\nabla_x f(P_j(x,s(x)))\right) - \sum_i \mu_i\left(\nabla_x g_i(P_j(x,s(x)))\right) = \lambda_j\left(\nabla_x h_j(P_j(x,s(x)))\right) \overset{(109a),(109b)}{=} \lambda_j\left(\nabla_x h_j(s(x))\right). \tag{117}$$

Using the Archimedean condition, the bound $\mu_i(1) \leq \frac{K}{R}$ and (109c), it is easy to find K such that eq. (114) holds for $p = \nabla_x h_j(s(x))$. Finally, since $\lambda_j = 0$ on $\mathcal{J}$ by the KKT conditions, applying the Hahn-Banach extension theorem allows us to extend λ_j from $\mathbb{S}_j + \mathcal{J}$, where $\mathbb{S}_j$ is defined as in (74), to $\mathcal{P}$ while preserving the norm. □

The non-commutative MFCQ condition also implies that the SOS polynomial g_i somehow 'dominates' an infinitesimal multiple of $\nabla_x g_i(\hat{q})$.

Lemma 14. *The non-commutative MFCQ condition implies that the polynomial*

$$g_i + t\nabla_x g_i(\hat{q}) - tR \tag{118}$$

admits an SOS decomposition for all $0 \leq t \leq \frac{1}{L}$.

Proof. For any $\gamma \in \mathbb{R}^+ \cup \{0\}$, if p is an SOS polynomial, then so is $p + \gamma g_i$. To arrive at eq. (118), given $0 < t < \frac{1}{L}$, add γg_i to equation (111), with $\gamma = \frac{1}{t} - \frac{1}{L}$ and then multiply everything by t. □

2. The effect of ncMFCQ

Theorem 15. *Let Problem (2) satisfy the Archimedean property and ncMFCQ. Then Problem (2) satisfies the operator optimality constraints (11a), (11b), (11c).*

Proof. Consider the set

$$\mathcal{N} := \{a \in \mathcal{P} \mid \|a\|_{\text{SOS}} = 0\}$$

of nullvectors w.r.t. the seminorm (79). It induces a proper norm on the quotient algebra $\mathcal{A} := \mathcal{P}/\mathcal{N}$. Observe that an element $a \in \mathcal{P}$ is in $\mathcal{N}$ iff for each $K \in \mathbb{R}^+$, $K - a^*a$ admits an SOS decomposition. In particular, each positive linear functional on $\mathcal{P}$ must kill $\mathcal{N}$. Hence each ρ_j induces a positive linear functional on $\mathcal{A}$, still called ρ_j. Similarly, each linear functional $\tau : \mathcal{P} \to \mathbb{C}$ bounded w.r.t. the seminorm (79) in the sense that for some $\Lambda \in \mathbb{R}$,

$$|\tau(p)| \leq \Lambda \|p\|_{\text{SOS}} \quad \forall p \in \mathcal{P}, \tag{119}$$

induces a (bounded) linear functional on $\mathcal{A}$. We shall abuse notation and sometimes identify elements of $\mathcal{P}$ with their images in $\mathcal{A}$.

Consider the set of bounded linear functionals $\tau : \mathcal{A} \to \mathbb{C}$ endowed with the (dual space) norm

$$\|\tau\|_{\text{SOS}} := \sup\{|\tau(p)| : 1-p,\, 1+p \text{ admit an SOS decomposition}\}. \tag{120}$$

Next, we define the set S of linear functionals $s : \mathcal{A}^{\times n} \to \mathbb{C}$ of the form

$$s(q) := \sum_i \mu_i(\nabla_x g_i(q)) + \sum_j \lambda_j(\nabla_x h_j(q)), \tag{121}$$

where $\mu_i : \mathcal{A} \to \mathbb{C}$ is an unnormalized state satisfying

$$\mu_i(g_i) = 0,\ \mu_i(pp^*) \geq 0,\ \mu_i(pg_kp^*) \geq 0, \quad \forall p \in \mathcal{A},\ k = 1, \ldots, m, \tag{122}$$

and $\lambda_j : \mathcal{A} \to \mathbb{C}$ is a bounded Hermitian linear functional, with

$$\lambda_j(s^+h_ks^-) = 0, \quad \forall s^+, s^- \in \mathcal{P},\ k = 1, \ldots, m'. \tag{123}$$

(Observe that (121) is well-defined since the lifts of bounded functionals μ_i, λ_i to maps on $\mathcal{P}$ vanish identically on $\mathcal{N}$ and $\nabla_x h(q) \in \mathcal{N}$ for $q \in \mathcal{N}^{\times n}$.) Let $\overline{S}^{\text{w}*}$ be the closure of S under the weak-$*$ topology [49, Sec. IV.5]. By construction, both S and $\overline{S}^{\text{w}*}$ are convex.

Assuming that the NPO Problem (2) is Archimedean, let $\sigma^\star$ be a minimizer. We then define the linear map $F : \mathcal{A}^{\times n} \to \mathbb{C}$

$$F(q) := \sigma^\star(\nabla_x f(q)). \tag{124}$$

There are two possibilities: (1) $F \in \overline{S}^{\mathrm{w}*}$; (2) $F \notin \overline{S}^{\mathrm{w}*}$.

In the first case, there exists a net of state and bounded linear functional tuples $(\mu_i^k, \lambda_j^k)_k$ such that, for all $q \in \mathcal{P}$,

$$\sigma^\star(\nabla_x f(q)) = \lim_{k\to\infty} \sum_i \mu_i^k(\nabla_x g_i(q)) + \sum_j \lambda_j^k(\nabla_x h_j(q)). \tag{125}$$

Following the proof of Lemma 13, this can be seen to imply the existence of functionals $\{\mu_i\}_i$ $\{\lambda_j\}_j$ such that eqs. (11a) (11b) and (11c) hold.

More precisely: take $q = \hat{q}$, the polynomial in (125). Then from (110) and (111) we have that

$$\sigma^\star(\nabla_x f(\hat{q})) = \lim_{k\to\infty} \sum_i \mu_i^k(\nabla_x g_i(\hat{q})) \overset{(118)}{\geq} R \lim_{k\to\infty} \sum_i \mu_i^k(1). \tag{126}$$

From this relation, it follows that the nets $\{(\mu_i^k)_k\}_i$ are bounded. Thus the Banach-Alaoglu theorem [49, Theorem IV.21] applies: the closed unit ball in the dual space $\mathcal{A}^*$ is compact. Therefore, there exists a subnet $(k_l)_l$ such that the limits

$$\mu_i := \lim_{l\to\infty} \mu_i^{k_l}, \quad i = 1, \dots, m, \tag{127}$$

exist. For arbitrary $p \in \mathcal{P}$, taking $q = P_0(x,p)$ in eq. (125), we thus have

$$\sigma^\star(\nabla_x f(P_0(x,p))) = \sum_i \mu_i(\nabla_x g_i(P_0(x,p))), \tag{128}$$

where we invoked the relation (109b) to annihilate the terms with the λ's. Next, we define the linear functionals $\lambda_j : \mathbb{S}_j + \mathcal{J} \to \mathbb{C}$ (see eq. (74) for the definition of $\mathbb{S}_j$) through the relation

$$\lambda_j(\nabla_x h_j(p) + q) := \sigma^\star(\nabla_x f(P_j(x,p))) - \sum_i \mu_i(\nabla_x g_i(P_j(x,p))), \tag{129}$$

for $q \in \mathcal{J}$, $p \in \mathcal{P}$. This definition is consistent by virtue of eq. (109c), which implies that, if $\nabla_x h_j(p) + q = 0$, the right-hand side vanishes. Moreover, also by eq. (109c), we have that there exists $K \in \mathbb{R}^+$ such that

$$\lambda_j(p) \leq K \|p\|_{SOS}, \tag{130}$$

for $p \in \mathbb{S}_j + \mathcal{J}$. The functionals $\{\lambda_j\}_j$ can thus be extended to $\mathcal{A}$, by virtue of the Hahn-Banach theorem. From eqs. (129), (109b) and (128), it follows that the relation

$$\sigma^\star(\nabla_x f(q)) = \sum_i \mu_i(\nabla_x g_i(q)) + \sum_j \lambda_j(\nabla_x h_j(q)). \tag{131}$$

holds for $q = P_j(x,p)$, for $j = 0, \dots, m'$ and all $p \in \mathcal{P}$. By eq. (109a), the equation above hence holds for all q. This is the stationarity condition (11c), and, by construction, the linear functionals $\{\mu_i\}_i$, $\{\lambda_j\}$ satisfy (11b) and (11a).

To prove the theorem, it therefore suffices to show that $F \notin \overline{S}^{\mathrm{w}*}$ leads to a contradiction. So let us assume that $F \notin \overline{S}^{\mathrm{w}*}$. Since $\overline{S}^{\mathrm{w}*}$ is closed in the weak-$*$ topology, by the Hahn-Banach separation theorem [49, Theorem V.4(c)], there exists a weak-$*$ continuous linear witness that separates F from $\overline{S}^{\mathrm{w}*}$. That is, by [49, Theorem IV.20], such a witness comes from evaluation at a point $\alpha \in \mathcal{A}^n$. Thus the n-tuple of polynomials α satisfies

$$F(\alpha) = \kappa < 0,\ s(\alpha) \geq 0, \quad \forall s \in \overline{S}^{\mathrm{w}*}. \tag{132}$$

The right-hand side implies that, for all i, and all states μ_i satisfying (122),

$$\mu_i(\nabla_x g_i(\alpha)) \geq 0. \tag{133}$$

We now claim that for each $\epsilon > 0$ there exists $L^i_\epsilon > 0$ such that

$$L^i_\epsilon g_i + \nabla_x g_i(\alpha) + \epsilon \tag{134}$$

is positive semidefinite. Suppose otherwise. Then there exists $\epsilon > 0$ such that for all $N > 0$ there is a (norm one) state μ_N with

$$\mu_N(N g_i + \nabla_x g_i(\alpha)) < -\epsilon. \tag{135}$$

Letting $C := \|\nabla_x g_i(\alpha)\|$, (135) yields

$$0 \leq \mu_N(g_i) < \frac{1}{N}(-\epsilon + C). \tag{136}$$

Again by Banach-Alaoglu, the sequence (μ_N) has a weak-$*$ convergent subsequence. Call the limit state μ. By (136), $\mu(g_i) = 0$. From (135), since $\mu_N(g_i) \geq 0$, $\mu_N(\nabla_x g_i(\alpha)) < -\epsilon$. Thus by weak-$*$ convergence, $\mu(\nabla_x g_i(\alpha)) \leq -\epsilon$. But this contradicts (133).

Now, call $(\mathcal{H}^\star, \psi^\star, X^\star)$ the result of applying the GNS construction on $\sigma^\star$. By the same argument that led to eq. (118), eq. (134) implies that

$$g_i(X^\star) + t\nabla_x g_i(\alpha(X^\star))) + t\epsilon\mathbb{I} \geq 0, \tag{137}$$

for $t \leq \frac{1}{L^i_\epsilon}$. For convenience, in the following we assume that $\epsilon < R$, where $R \in \mathbb{R}^+$ is the constant appearing in (111).

Eq. (132) also implies that $\lambda_j(\nabla_X h_j(\alpha(X))) = 0$ for all normalized linear functionals λ_j. In particular, we have that

$$\nabla_X h_j(\alpha(X^\star))\Big|_{X=X^\star} = 0, \quad j = 1, \ldots, m'. \tag{138}$$

By eq. (109c), it hence follows that

$$P_j(X^\star, \alpha(X^\star)) = 0, \quad j = 1, \ldots, m'. \tag{139}$$

Thus, by the completeness relation (109a) we arrive that

$$\alpha(X^\star) = \sum_{j=0}^{m'} P_j(X^\star, \alpha(X^\star)) = P_0(X^\star, \alpha(X^\star)). \tag{140}$$

By a similar argument, the polynomial $\hat{q}$ appearing in the ncMFCQ conditions (110), (111) also satisfies $\hat{q}(X^\star) = P_0(X, \hat{q}(X^\star))$.

We now invoke Lemma 11, with $q = (1 - \frac{\epsilon}{R})P_0(x, \alpha) + \frac{\epsilon}{R}P_0(x, \hat{q})$. We then have that there exists a trajectory $\{X(t) : t \in [-\epsilon', \epsilon']\}$ of operators in $B(\mathcal{H}^\star)$ satisfying the equality constraints

$$h_j(X(t)) = 0, \quad j = 1, \ldots, m', \tag{141}$$

and such that

$$X(0) = X^\star, \quad \frac{dX(t)}{dt}\Big|_{t=0} = \left(1 - \frac{\epsilon}{R}\right) P_0(X^\star, \alpha(X^\star)) + \frac{\epsilon}{R} P_0(X^\star, \hat{q}(X^\star)) = \left(1 - \frac{\epsilon}{R}\right) \alpha(X^\star) + \frac{\epsilon}{R}\hat{q}(X^\star). \tag{142}$$

Now, suppose that t satisfies $0 < t \leq \min(\epsilon', \frac{1}{L^i_\epsilon}, \frac{1}{L})$. Then we have that

$$\begin{aligned} g_i(X(t)) &= g_i(X^\star) + t\nabla_x g_i\left(\frac{dX(t)}{dt}\Big|_{t=0}\right)\Big|_{x=X^\star} + o(t) \\ &= \left(1 - \frac{\epsilon}{R}\right)(g_i(X^\star) + t\nabla_x g_i(\alpha(X^\star))) + \frac{\epsilon}{R}(g_i(X^\star) + t\nabla_x g_i(\hat{q}(X^\star))) + o(t) \\ &\overset{(137),(118)}{\geq} t\frac{\epsilon^2}{R} + o(t). \end{aligned} \tag{143}$$

It follows that, for some $0 < \epsilon'' < \epsilon'$, $\{g_i(X(t))\}_i$ are positive semidefinite for $t \in [0, \epsilon'']$. Thus, $\{X(t) : t \in [0, \epsilon'']\}$ is a trajectory of feasible operators.

Finally, consider the expression

$$\frac{d\sigma^\star(f(X(t)))}{dt}\Big|_{t=0} = F\left(\frac{dX(t)}{dt}\Big|_{t=0}\right) = \left(1 - \frac{\epsilon}{R}\right) F(\alpha) + \frac{\epsilon}{R}F(\hat{q}) = -\kappa\left(1 - \frac{\epsilon}{R}\right) + \frac{\epsilon}{R}F(\hat{q}). \tag{144}$$

The left-hand side is non-negative, since $\sigma^\star(f(X(0)))$ is minimal by assumption. The right-hand side tends to $-\kappa$ as $\epsilon \to 0$. We therefore end up in a contradiction, and so $F \in \bar{S}$.

We have just proven that there exist unnormalized states $\mu_i : \mathcal{P} \to \mathbb{C}$ and functionals $\lambda_i : \mathcal{P} \to \mathbb{C}$ satisfying eq. (11b), (11a) and (11c). It thus follows that the operator optimality conditions apply to Problem (1). $\square$

VI. 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,

$$[z_k, y_l] = 0, \quad \forall k, l. \tag{145}$$

If the remaining constraints on z are convex, then one can prove that the operator optimality conditions hold in some sense for polynomials of z. The following lemma is the key to arrive at this result.

Lemma 16. *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 functional* [1]*. Given some* concave *non-commutative polynomials $\{\hat{g}_i\}$ and* affine *polynomials $\{\hat{h}_i\}$, consider the optimization problem*

$$\begin{aligned} &\min_{Z \in \mathcal{A}^{\times q}} \hat{f}(Z) \\ \text{s.t. } &\hat{g}_i(Z) \geq 0, \quad i = 1, \ldots, m_Z, \\ &\hat{h}_j(Z) = 0, \quad j = 1, \ldots, m'_Z. \end{aligned} \tag{146}$$

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

Further assume that Problem (146) admits a strictly feasible point, i.e., there exists a feasible tuple $\hat{Z} \in \mathcal{A}^{\times q}$ such that

$$\begin{aligned} &\hat{g}_i(\hat{Z}) > 0, \quad i = 1, \ldots, m_Z, \\ &\hat{h}_j(\hat{Z}) = 0, \quad j = 1, \ldots, m'_Z. \end{aligned} \tag{147}$$

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

$$\mu_i(\hat{g}_i(Z^\star)) = 0, \quad i = 1, \ldots, m_Z, \tag{148}$$

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

$$\min_{Z \in \mathcal{A}_h^{\times q}} \mathcal{L}(Z; \mu, \lambda), \tag{149}$$

with

$$\mathcal{L}(Z; \mu, \lambda) := \hat{f}(Z) - \sum_i \mu_i(\hat{g}_i(Z)) - \sum_j \lambda_j(\hat{h}_j(Z)). \tag{150}$$

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:

$$\begin{aligned} A :=&\{(r, S, T) : r \in \mathbb{R},\, S \in \mathcal{A}_h^{\times m},\, T \in \mathcal{A}_h^{\times m'}, \\ &\quad \exists Z_1, \ldots, Z_n \in \mathcal{A}_h, \hat{f}(Z) \leq r,\, -\hat{g}_i(Z) \leq S_i,\, \hat{h}_j(Z) = T_j,\, \forall i, j\}, \\ B :=&\{(\nu, 0, 0) : \nu < \hat{f}(Z^\star)\}. \end{aligned} \tag{151}$$

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 [49, Theorem V.4(a)] implies that there exists a separating linear functional (ϕ, μ, λ) and $\alpha \in \mathbb{R}$ such that

$$\phi r + \sum_i \mu_i(S_i) + \sum_j \lambda_j(T_j) \geq \alpha, \quad \forall (r, S, T) \in A, \tag{152a}$$

---

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

$$\phi\nu \le \alpha, \quad \forall(\nu, 0, 0) \in B. \tag{152b}$$

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. (152a) arbitrarily small, just by replacing S_i with $S_i + uyy^*$, with $u \in \mathbb{R}^+$ sufficiently large. It follows that, for all i, $\mu_i(yy^*) \ge 0$, i.e., $\{\mu_i\}_i$ are non-normalized states of $\mathcal{A}$.

Notice as well that we can choose ν to be arbitrarily small in eq. (152b). It follows that $\phi \ge 0$. We next prove that $\phi > 0$. Suppose, on the contrary, that $\phi = 0$. From eqs. (152a), (152b) we have that

$$\phi\hat{f}(Z) - \sum_i \mu_i(g_i(Z)) - \sum_j \lambda_j(h_i(Z)) \ge \alpha \ge \phi\hat{f}(Z^\star), \quad \forall Z \in \mathcal{A}^{\times p}. \tag{153}$$

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

$$\phi(\hat{f}(\hat{Z}) - f(Z^\star)) \ge \sum_i \mu_i(g_i(\hat{Z})). \tag{154}$$

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. (152a) implies that

$$\sum_j \lambda_j(h_j(Z)) \ge \alpha, \quad \forall Z. \tag{155}$$

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. (153) by ϕ, we have that

$$\hat{f}(Z) - \sum_i \tilde{\mu}_i(\hat{g}_i(Z)) - \sum_j \tilde{\lambda}_j(\hat{h}_j(Z)) \ge \hat{f}(Z^\star), \quad \forall Z, \tag{156}$$

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

Finally, take $Z = Z^\star$ in eq. (156). We arrive at:

$$f(Z^\star) - \sum_i \tilde{\mu}_i(\hat{g}_i(Z^\star)) \ge f(Z^\star). \tag{157}$$

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 (148). In that case,

$$\mathcal{L}(Z^\star, \tilde{\mu}, \tilde{\lambda}) = \hat{f}(Z^\star), \tag{158}$$

and so, by the above equation and (156), $Z^\star$ is a global solution of the unconstrained problem (149). □

Theorem 17. *Consider an NPO (2) with variables $x = (y, z)$, with $z = (z_1, \ldots, z_q)$ such that*

1. *For fixed Y, $f(Y, z)$ is a convex polynomial on the variables z.*
2. *The only constraints involving both types of variables y and z are the following:*

$$[y_r, z_s] = 0, \quad \forall r, s. \tag{159}$$

3. *The remaining constraints involving variables of type z are*

$$\begin{aligned} \hat{g}_i(z) &\ge 0, \quad i = 1, \ldots, m_Z, \\ \hat{h}_j(z) &= 0, \quad j = 1, \ldots, m'_Z, \end{aligned} \tag{160}$$

where $\{\hat{g}_i\}_i$ are concave non-commutative polynomials and $\{\hat{h}_j\}_j$ are linearly independent affine polynomials (linear independence in the usual sense).

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

$$\hat{g}_i(Q(z)) - r \tag{161}$$

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

$$\hat{h}_j(Q(z)) = \sum_{l,j'} s_{jj'l}(z)\hat{h}_{j'}(z)s'_{jj'l}(z), \tag{162}$$

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

Denote by $\mathcal{P}_Z$ the set of polynomials on $z_1, \ldots, z_q$. Then, besides the state optimality conditions (11d), it is legitimate to add to Problem (2) the constraints:

$$\begin{aligned}
\exists \mu_i : \mathcal{P}_Z \to \mathbb{C},\ & \mu_i(pp^*) \geq 0, \quad \forall p \in \mathcal{P}_Z,\ i = 1, \ldots, m_Z, \\
& \mu_i(p\hat{g}_l p^*) \geq 0, \quad \forall p \in \mathcal{P}_Z,\ i = 1, \ldots, m_Z,\ l = 1, \ldots, m_Z, \\
& \mu_i(s\hat{h}_j s') = 0, \quad \forall s, s' \in \mathcal{P}_Z,\ i = 1, \ldots, m_Z,\ j = 1, \ldots, m'_Z, \\
& \mu_i(\hat{g}_i) = 0, \quad i = 1, \ldots, m_Z, && \text{(163a)} \\
\exists \lambda_j : \mathcal{P}_Z \to \mathbb{C},\ & \lambda_j(p + p^*) \in \mathbb{R}, \quad \forall p \in \mathcal{P}_Z,\ j = 1, \ldots, m'_Z, \\
& \lambda_j(s\hat{h}_j s') = 0, \quad \forall s, s' \in \mathcal{P}_Z,\ j = 1, \ldots, m'_Z, && \text{(163b)} \\
\sigma\left(\nabla_z f(p)\right) - \sum_i \mu_i \left(\nabla_z \hat{g}_i(p)\right) & - \sum_j \lambda_j \left(\nabla_z \hat{h}_j(p)\right) = 0, \quad \forall p \in \mathcal{P}_Z^{\times q}. && \text{(163c)}
\end{aligned}$$

Call $(\mathcal{H}^\star, \psi^\star, X^\star)$ the solution of the original NPO (1), with $X^\star = (Y^\star, Z^\star)$. Then, both states $\{\mu_i\}_i$ and functionals $\{\lambda_i\}$ can be further restricted to act on the algebra generated by $Z_1^\star, \ldots, Z_q^\star$.

Proof. Let $(\mathcal{H}^\star, \sigma^\star, X^\star)$ be the solution of Problem (2), with $X^\star = (Y^\star, Z^\star)$. Call $\mathcal{A}$ the algebra generated by $Z^\star$. Since the only relations connecting X with Z are the commutation relations (159), it follows that the solution $p^\star$ of Problem (2) satisfies

$$\begin{aligned}
p^\star = \min_{Z \in \mathcal{A}^{\times q}} & \hat{f}(Z) \\
\text{such that } & \hat{g}_i(Z) \geq 0, i = 1, \ldots, m_Z, \\
& \hat{h}_j(Z) = 0, j = 1, \ldots, m'_Z,
\end{aligned} \tag{164}$$

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

$$\hat{f}(Z) = \sigma^\star(f(Y^\star, Z)). \tag{165}$$

Moreover, one of the minimizers of (164) is $Z = Z^\star$.

In addition, by eqs. (161), (162), 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 16 and conclude that $Z^\star$ is the solution of the unconstrained Problem (149), for some states $\{\mu_i\}_i$ and symmetric linear functionals $\{\lambda_j\}_j$. Next, for any q-tuple of symmetric polynomials p on z, consider the following trajectory in $\mathcal{A}^{\times q}$

$$Z(t) := Z^\star + tp(Z^\star). \tag{166}$$

Since $Z^\star$ is a minimizer of Problem (149), it follows that

$$0 = \frac{d\mathcal{L}(Z(t); \mu, \lambda)}{dt}\Big|_{t=0} = \sigma^\star\left(\nabla_z f(Y^\star, z)(p(Z^\star))\right) - \sum_i \mu_i^k \left(\nabla_z \hat{g}_i(p)\Big|_{Z=Z^\star}\right) - \sum_j \lambda_j \left(\nabla_z \hat{h}_j(p)\Big|_{Z=Z^\star}\right) = 0. \tag{167}$$

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

VII. 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

$$C(P) := \sum_{a,b,x,y} C(a, b, x, y)P(a, b|x, y) \tag{168}$$

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

$$\begin{aligned} c^\star = \min \sigma & \left(\frac{1}{2} \sum_{a,b,x,y} C(a, b, x, y)\{E_{a|x}, F_{b|y}\} \right) \\ \text{s.t. } & E_{a|x} \geq 0, \quad \forall a, x, \\ & \sum_a E_{a|x} - \mathbb{I} = 0, \quad \forall x, \\ & F_{b|y} \geq 0, \quad \forall b, y, \\ & \sum_b F_{b|y} - \mathbb{I} = 0, \quad \forall y, \\ & [E_{a|x}, F_{b|y}] = 0, \quad \forall a, b, x, y. \end{aligned} \tag{169}$$

As we can appreciate, taking the partition $X = (E, F)$, the NPO satisfies the conditions of Theorem (17), 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 (169) can be chosen such that the non-commuting variables are all projectors [51]. That is,

$$\begin{aligned} (E^\star_{a|x})^2 = E^\star_{a|x}, \quad \forall a, x, \\ (F^\star_{b|y})^2 = F^\star_{b|y}, \quad \forall b, y. \end{aligned} \tag{170}$$

Next, we apply Theorem 17 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). By the last paragraph of the statement of the theorem, we can, not only demand the state σ fulfill the 'projector relations'

$$\sigma(s((E_{a|x})^2 - E_{a|x})s') = \sigma(s((F_{b|y})^2 - E_{b|y})s'), \quad \forall s, s', \tag{171}$$

but also the Lagrangian multipliers of Alice's $\mu^A_{a|x}, \lambda^A_x$ and Bob's $\mu^B_{b|y}, \lambda^B_y$ can be constrained to satisfy

$$\mu^A_{a|x}(s((E_{a'|x'})^2 - E_{a'|x'})s') = 0, \tag{172a}$$

$$\lambda^A_x(s((E_{a|x'})^2 - E_{a|x'})s') = 0, \tag{172b}$$

$$\mu^B_{b|y}(s((F_{b'|y'})^2 - F_{b'|y'})s') = 0, \tag{172c}$$

$$\lambda^B_y(s((F_{b|y'})^2 - F_{b|y'})s') = 0. \tag{172d}$$

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:

$$\mu^A_{a|x}(ss^*) \geq 0, \quad \forall a, x, \forall s \in \mathcal{P}_E, \tag{173a}$$

$$\mu^A_{a|x}\Big(s\big((E_{a'|x'})^2 - E_{a'|x'}\big)s'\Big) = 0, \quad \forall a, a', x, x', \forall s, s' \in \mathcal{P}_E \tag{173b}$$

$$\mu^A_{a|x}\Big(s\Big(\sum_{a'} E_{a'|x'} - 1\Big)s'\Big) = 0, \quad \forall a, x, x', \forall s, s' \in \mathcal{P}_E, \tag{173c}$$

$$\lambda^A_x\Big(s\big((E_{a|x'})^2 - E_{a|x'}\big)s'\Big) = 0, \quad \forall a, x, x', \forall s, s' \in \mathcal{P}_E \tag{173d}$$

$$\lambda_x^A\Big(s\Big(\sum_a E_{a|x}-1\Big)s'\Big)=0,\quad \forall x,\forall s,s'\in\mathcal{P}_E, \tag{173e}$$

$$\mu^A_{a|x}(E_{a|x})=0,\quad \forall a,x, \tag{173f}$$

$$\sigma\left(\frac{1}{2}\sum_{b,y}C(a,b,x,y)\{p,F_{b|y}\}\right)=\mu^A_{a|x}(p)+\lambda^A_x(p),\quad \forall a,x,\forall p\in\mathcal{P}_E. \tag{173g}$$

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.

A. Only two outcomes

When a,b can only take two values, it is customary to rewrite Problem (169) in terms of 'dichotomic operators' A_x, B_y. The problem to solve is thus

$$\begin{aligned}
&\min\ \sigma(H)\\
&\text{s.t. } \frac{1-A_x}{2}\ge 0,\ \frac{1+A_x}{2}\ge 0,\quad \forall x,\\
&\qquad \frac{1-B_y}{2}\ge 0,\ \frac{1+B_y}{2}\ge 0,\quad \forall y,\\
&\qquad [A_x,B_y]=0,\quad \forall x,y
\end{aligned} \tag{174}$$

where H is the Bell polynomial

$$\frac{1}{2}\sum_{x,y}c_{xy}\{A_x,B_y\}+\sum_x d_xA_x+\sum_y e_yB_y. \tag{175}$$

To simplify notation, we define the polynomials

$$\mathcal{F}_x:=\sum_y\frac{1}{2}c_{xy}B_y+\frac{1}{2}d_x\mathbb{I}, \tag{176a}$$

$$\mathcal{G}_y:=\sum_x\frac{1}{2}c_{xy}A_x+\frac{1}{2}e_y\mathbb{I}, \tag{176b}$$

which allow us to express H as

$$H=\sum_x\{\mathcal{F}_x,A_x\}+\sum_y e_yB_y=\sum_y\{\mathcal{G}_y,B_y\}+\sum_x d_xA_x. \tag{177}$$

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

$$(A^\star_x)^2=(B^\star_y)^2=1. \tag{178}$$

Thus, once more we can apply Theorem 17 to the algebra $\mathcal{A}$ generated by $A^\star_1,\dots,A^\star_n$ and conclude that one can add new state multipliers μ^+_x,μ^-_x to the problem, with the properties:

$$\mu^\pm_x(s(A^2_{x'}-1)s')=0,\quad \forall x,x',\ \forall s,s'\in\mathcal{A}, \tag{179a}$$

$$\mu^\pm_x\left(\frac{1\pm A_x}{2}\right)=0,\quad \forall x, \tag{179b}$$

$$\sigma(\{p,\mathcal{F}_x\})=\mu^+_x(p)-\mu^-_x(p),\quad \forall x,\forall p\in\mathcal{A}. \tag{179c}$$

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

Let $E^\pm_x:=\frac{1\pm A_x}{2}$. From eq. (179b) and the positivity of $\mu^\pm_x$, an analogous argument to the one used to derive eqs. (38) shows that

$$\mu^\pm_x\left(E^\pm_x p\right)=\mu^\pm_x\left(pE^\pm_x\right)=0\quad \forall p\in\mathcal{A}. \tag{180}$$

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

$$\mu_x^{\pm}\left(\{A_x, q\}\right) = \mp\mu_x^{\pm}\left(q + A_x q A_x\right) = 0. \tag{181}$$

Thus, if we set $p = -\{A_x, q\}$ in eq. (179c), we arrive that

$$-\sigma\left(\{\{A_x, q\}, \mathcal{F}_x\}\right) = \mu^+(q + A_x q A_x) + \mu^-(q + A_x q A_x). \tag{182}$$

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

$$-\sigma\left(\{\{A_x, ss^*\}, \mathcal{F}_x\}\right) = \mu^+(ss^* + A_x ss^* A_x) + \mu^-(ss^* + A_x ss^* A_x) \geq 0, \forall s \in \mathcal{P}_E. \tag{183}$$

Setting $p = [A_x, q]$ in eq. (179c) and using eq. (180), we obtain another useful constraint:

$$\sigma\left(\{[A_x, q], \mathcal{F}_x\}\right) = 0, \quad \forall q \in \mathcal{A}. \tag{184}$$

Constraints (183), (184) are, respectively, extra positivity and linear conditions that one can apply to the already existing variables of the 'quantum NPO' (169).

B. Numerical implementation

In order to implement numerically the constraints (183) and (184), together with the analogous constraints for Bob and the state optimality conditions (11d), 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

$$\sigma\left(\mathcal{F}_x[A_x, m_i^A]\right) = 0, \tag{185a}$$

$$\sigma\left(\mathcal{G}_y[B_y, m_i^B]\right) = 0, \tag{185b}$$

$$\sigma\left([H, m_i]\right) = 0, \tag{185c}$$

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

$$\alpha_{ij}^x := -\sigma\left(\mathcal{F}_x\{A_x, m_i^{A^*} m_j^A\}\right), \tag{186a}$$

$$\beta_{ij}^y := -\sigma\left(\mathcal{G}_y\{B_y, m_i^{B^*} m_j^B\}\right), \tag{186b}$$

$$\gamma_{ij} := \sigma\left(m_i^* H m_j - \frac{1}{2}\{H, m_i^* m_j\}\right). \tag{186c}$$

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 (186).

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 [52]. Although the vanilla NPA hierarchy is always strictly feasible [53], this is in general not true when additional constraints are enforced [54]. 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. (185c) to rewrite eq. (186c) as

$$\gamma_{ij} = \sigma(m_i^*[H, m_j]). \tag{187}$$

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 (188), 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 calculation were done using the software Moment [46]. We are particularly interested in checking whether we have achieved convergence at some level. To certify that, we test for the existence of a rank loop [2] or, as it is known in the mathematics literature, the flatness condition [14, 55–58]. We remark under the corresponding table whether it holds.

We start with a tilted version of the CHSH inequality [59], with an additional $\tau(A_0 + B_0)$ term [60, 61]. In full correlation notation the coefficients table is given by

$$\left(\begin{array}{c|cc} 0 & \tau & 0 \\ \hline \tau & 1 & 1 \\ 0 & 1 & -1 \end{array}\right). \tag{188}$$

The results are shown in Tables V and VI. As τ goes to 1 the level at which the NPA hierarchy converges 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 V. Results for the tilted CHSH inequality with $\tau = 0.95$. For comparison, the analytical answer is 3.9001 6389 9372. With the KKT constraints we get a rank loop at level 3, and without at level 4.

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

Table VI. Results for the tilted CHSH inequality with $\tau = 0.99$. For comparison, the analytical answer is 3.9800 0132 8893. With the KKT constraints we get a rank loop at level 4, and without at level 7.

Our next example is the well-studied I3322 inequality [62–64]. In full correlation notation the coefficients table is given by

$$\frac{1}{4}\left(\begin{array}{c|ccc} 0 & -1 & -1 & 0 \\ \hline -1 & -1 & -1 & -1 \\ -1 & -1 & -1 & 1 \\ 0 & -1 & 1 & 0 \end{array}\right). \tag{189}$$

The results are shown in Table VII. Its maximal violation is conjectured to occur only for an infinite-dimensional system [63], 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 (186), 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 VII. 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_{4422}^{20} inequality [65], that had a gap between the best known lower bound and the best known upper bound (see Table IV of Ref. [66]). In full correlation notation the coefficients table is given by

$$\frac{1}{4}\left(\begin{array}{r|rrrr} -12 & -1 & -1 & -2 & 4 \\ \hline -1 & -1 & 1 & 1 & 2 \\ -1 & 1 & -1 & 1 & 2 \\ -2 & 1 & 1 & -2 & 2 \\ 4 & 2 & 2 & 2 & -2 \end{array}\right). \tag{190}$$

The results are shown in Table VIII. For increased efficiency the calculations here were done without the positivity conditions (186), 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 VIII. Results for the inequality I_{4422}^{20}. 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.

VIII. 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.

While the state optimality conditions (11d) hold for all problems, the operator optimality conditions (11a) – (11c) 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 full KKT conditions hold. However, this property is difficult to verify for most NPO problems.

Thus, we generalized two known ‘classical’ qualification constraints: Linear Independence Constraint Qualification (LICQ) and Mangasarian-Fromovitz Constraint Qualification (MFCQ), both of which legitimate the use of the KKT conditions in NPO. We also presented very mild conditions that guarantee that at least some relaxed form of the KKT 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 KKT conditions improved a lot the speed of convergence, 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 [40]. It is intriguing whether the state optimality conditions can be integrated within renormalization flow techniques, like those proposed in [39]. 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. was 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).

[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).

[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] 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.

[38] X. Han, Quantum many-body bootstrap (2020), arXiv:2006.06002 [cond-mat.str-el].

[39] 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].

[40] 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].

[41] 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].

[42] 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].

[43] F. Verstraete and J. I. Cirac, Matrix product states represent ground states faithfully, Physical Review B **73**, 094423 (2006), arXiv:cond-mat/0505140.

[44] 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.

[45] 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).

[46] A. Garner and M. Araújo, Moment, `https://github.com/ajpgarner/moment` (2023).

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

[48] N. Ozawa, About the connes embedding conjecture: Algebraic approaches, Japanese Journal of Mathematics **8**, 147–183 (2013).

[49] M. Reed and B. Simon, *Methods of modern mathematical physics. I: Functional analysis.* (1980).

[50] A. Barvinok, *A Course in Convexity*, Graduate studies in mathematics (American Mathematical Society, 2002).

[51] 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**, 10.1063/1.3514538 (2011).

[52] 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].

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

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

[55] 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.

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

[57] 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.

[58] 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].

[59] 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).

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

[61] 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].

[62] 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.

[63] 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].

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

[65] 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].

[66] 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 VII, together with the state optimality conditions (11d).

$$
\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}
\tag{A1}
$$