Approximation Limits of Linear Programs (Beyond Hierarchies)

Gábor Braun[1], Samuel Fiorini[2], Sebastian Pokutta[3], and David Steurer[4]

[1]Universität Leipzig, Institut für Informatik, PF 100920, 04009 Leipzig, Germany.
Email: gabor.braun@informatik.uni-leipzig.de
[2]Department of Mathematics, Université libre de Bruxelles CP 216, Bd. du Triomphe, 1050 Brussels, Belgium. *Email:* sfiorini@ulb.ac.be
[3]ISyE, Georgia Institute of Technology, Atlanta, GA, USA. *Email:* sebastian.pokutta@isye.gatech.edu
[4]Department of Computer Science, Cornell University, Ithaca, NY 14853, United States.
Email: dsteurer@cs.princeton.edu

July 2, 2013

Abstract

We develop a framework for proving approximation limits of polynomial-size linear programs from lower bounds on the nonnegative ranks of suitably defined matrices. This framework yields unconditional impossibility results that are applicable to *any* linear program as opposed to only programs generated by hierarchies. Using our framework, we prove that $O(n^{1/2-\epsilon})$-approximations for CLIQUE require linear programs of size $2^{n^{\Omega(\epsilon)}}$. This lower bound applies to linear programs using a certain encoding of CLIQUE as a linear optimization problem. Moreover, we establish a similar result for approximations of semidefinite programs by linear programs.

Our main technical ingredient is a quantitative improvement of Razborov's rectangle corruption lemma (1992) for the high error regime, which gives strong lower bounds on the nonnegative rank of shifts of the unique disjointness matrix.

1 Introduction

1.1 Context

Linear programs (LPs) play a central role in the design of approximation algorithms, see, e.g., [Vazirani, 2001, Williamson and Shmoys, 2011, Lau et al., 2011]. Therefore, understanding the limitations of LPs as tools for designing approximation algorithms is an important question.

The first generation of results studied the limitations of *specific* LPs by seeking to determine their integrality gaps. The second generation of results, pioneered by Arora et al. [2002], studied the limitations of *structured* LPs such as those generated by *lift-and-project procedures* or *hierarchies* (e.g., Sherali and Adams [1990] and Lovász and Schrijver [1991]). See the previous work section below for a more detailed account of the relevant literature.

In this work, we start a third generation of results that apply to *any* LP for a given problem. For example, our lower bounds address the following question: Is there a polynomial-size linear programming relaxation LP_n for CLIQUE that achieves a $n^{\Theta(1)}$-approximation for all graphs with at most n vertices? We develop a framework for reducing these kind of questions to lower bounds on the nonnegative rank of certain matrices obtained from a linear encoding of the problem considered. Then we prove lower bounds on the nonnegative rank of the matrices for CLIQUE. Although we mainly focus on LPs, our framework readily generalizes to semidefinite programs (SDPs).

Linear Encodings We consider combinatorial optimization problems[1] that can be encoded in a linear fashion by specifying a set of feasible solutions represented as binary vectors and a set of admissible (linear) objective functions represented by their coefficient vectors. An instance of a given linear encoding is specified by a dimension d and admissible objective function $w \in \mathbb{R}^d$. Solving the instance means finding a feasible solution $x \in \{0,1\}^d$ such that $w^\intercal x = \sum_{i=1}^d w_i x_i$ is minimum (or maximum). The optimum value of the instance is thus the minimum (or maximum) value of $w^\intercal x$ for a feasible $x \in \{0,1\}^d$.

We require that every instance of the problem can be mapped to an instance of the linear encoding in such a way that feasible solutions to an instance of the problem can be converted in polynomial time to feasible solutions to the corresponding instance of the linear encoding without deteriorating their objective function values, and vice-versa. In this case, we say that a linear encoding *faithfully* encodes the problem. For graph problems such as the maximum clique problem (CLIQUE), such a linear encoding does not allow the set of feasible solutions to depend on the input graph, which is encoded solely in the objective function. Constraints are only allowed to depend on the size n of the ground set.

Example 1 (Linear encoding of metric TSP)**.** With the natural linear encoding of the metric traveling salesman problem (metric TSP) the feasible solutions are the characteristic vectors (or incidence vectors) of tours of the complete graph over $[n] := \{1, 2, \ldots, n\}$ for some $n \geqslant 3$, and the admissible objective functions are all nonnegative vectors $w = (w_{ij})$ such that $w_{ik} \leqslant w_{ij} + w_{jk}$ for all distinct i, j and k in $[n]$. All vectors are encoded in $\mathbb{R}^d$, where $d = \binom{n}{2}$.

In general, a linear encoding determines two nested convex sets $P \subseteq Q$ in $\mathbb{R}^d$ for each d. The set P is the convex hull of the feasible solutions of dimension d, thus P is a 0/1-polytope[2]. In case of a minimization problem, the set Q is defined by all inequalities of the form $w^\intercal x \geqslant \delta$ satisfied by P where w is an admissible objective function of dimension d, and δ is chosen as large as possible. For a maximization problem, the inequalities are of the form $w^\intercal x \leqslant \delta$ and δ is chosen as small as possible. In other words, Q is the tightest relaxation of P with the given facet coefficients.

(Approximate) Extended Formulations We begin by illustrating the concept on our former example.

Example 2 (Approximate extended formulation of metric TSP)**.** We return to Example 1. It is known that the Held-Karp relaxation K of the metric TSP has integrality gap at most $3/2$ (see Held and Karp [1970], Wolsey [1980]). In geometric terms, this means that $P \subseteq K \subseteq 2/3 \cdot Q$. Although K is defined by an exponential number of inequalities, it is known that it can be reformulated with a polynomial number of constraints by adding a polynomial number of variables, see, e.g., Carr et al. [2009]. That is, the Held-Karp relaxation K has a polynomial-size extended formulation.

Formally, an *extended formulation* (EF) of a polyhedron $K \subseteq \mathbb{R}^d$ is a linear system in variables $(x, y) \in \mathbb{R}^{d+k}$ such that, for every $x \in \mathbb{R}^d$, we have $x \in K$ if and only if there exists $y \in \mathbb{R}^k$ such that (x, y) is a solution to the system. The *size* of an EF is the number of *inequalities* in the system, thus the variables and equalities are not counted. This turns out to be the right definition of size. It can be shown that an EF can always be brought into *slack form* $Ex + Fy = g$, $y \geqslant \mathbf{0}$ without increasing

[1]We assume some familiarity with combinatorial optimization. See, e.g., Schrijver [2003].

[2]We also assume some familiarity with (convex) polytopes and polyhedra, see the standard reference Ziegler [1995].

its size. We will mainly consider EFs in slack form. For these, the size equals the number of extra variables.

The *extension complexity* $\mathrm{xc}(K)$ of the polyhedron K is defined as the minimum size of an EF of K. Most of the LP relaxations that appear in the context of approximation algorithms actually have polynomial extension complexity. This is in particular the case of the relaxations obtained from an initial polynomial size relaxation at a bounded level of any of the common linear programming hierarchies.

Let $\rho \geqslant 1$. Then we say that $Ex + Fy = g$, $y \geqslant \mathbf{0}$ is a ρ-approximate EF of a given maximization problem, w.r.t. a given linear encoding of this problem, if the maximum value of $w^\intercal x$ on $K := \{x \in \mathbb{R}^d \mid \exists y : Ex + Fy = g,\ y \geqslant \mathbf{0}\}$ is at least the optimum value for every $w \in \mathbb{R}^d$ and at most ρ times the optimum value for every *admissible* $w \in \mathbb{R}^d$. Geometrically, this is equivalent to $P \subseteq K \subseteq \rho Q$. For minimization problems, the definitions are similar. In this case, we have $P \subseteq K \subseteq \rho^{-1} Q$.

Nonnegative Factorizations A *rank-r nonnegative factorization* of an $m \times n$ matrix M is a decomposition of M as a product $M = TU$ of nonnegative matrices T and U of sizes $m \times r$ and $r \times n$, respectively. The *nonnegative rank* $\mathrm{rank}_+(M)$ of M is the minimum rank r of nonnegative factorizations of M. In case M is zero, we let $\mathrm{rank}_+(M) = 0$. It is quite useful to notice that the nonnegative rank of M is also the minimum number of nonnegative rank-1 matrices whose sum is M. From this, we see immediately that the nonnegative rank of M is at least the nonnegative rank of any of its submatrices.

The factorization theorem of Yannakakis [1991] (see [Yannakakis, 1988] for the conference version) states that extension complexity of a polytope K is precisely the nonnegative rank of any of its slack matrices. If K is the convex hull of $\{v_1, \dots, v_n\} \subseteq \mathbb{R}^d$ and the set of solutions to $A_1 x \leqslant b_1$, $\dots$, $A_m x \leqslant b_m$ then the *slack matrix* of K with respect to these outer and inner descriptions is the $m \times n$ nonnegative matrix S with entries $S_{ij} := b_i - A_i v_j$. Yannakakis' theorem states that $\mathrm{xc}(K) = \mathrm{rank}_+(S)$ for every polytope K and every slack matrix S of K. This theorem can be straightforwardly generalized to the case where K is an unbounded polyhedron, see, e.g., Conforti et al. [2010], or Theorem 1 below.

The Link to Communication Complexity Yannakakis' factorization theorem initiated an interplay between the extension complexity of polytopes and (classical) communication complexity.[3] The relevant concept here is randomized communication protocol with private randomness and *nonnegative outputs* computing a (nonnegative) function $M : X \times Y \to \mathbb{R}_+$ *in expectation*. For the sake of simplicity, we call this a *protocol computing M in expectation*.

Faenza et al. [2011] show that, considering M as a matrix, the minimum complexity of a protocol computing M in expectation equals $\log(\mathrm{rank}_+(M)) + \Theta(1)$. (This was proved independently by Zhang [2012].) Thus proving bounds on the nonnegative rank of M amounts to proving bounds on the required amount of communication for computing M in expectation.

It is not hard to see that this last quantity is bounded from below by the nondeterministic communication complexity of the support of M because every protocol computing M in expectation can be turned into a nondeterministic protocol for the support of M. Equivalently, the nonnegative rank of the matrix M is bounded from below by the minimum number of 1-monochromatic rectangles covering the support of M.

(Unique) Disjointness In the *disjointness problem* (DISJ), both Alice and Bob receive a subset of $[n]$. They have to determine whether the two subsets are disjoint. The disjointness problem is central to communication complexity, see Chattopadhyay and Pitassi [2010] for a survey.

[3] We assume some familiarity with communication complexity. See Kushilevitz and Nisan [1997].

A related problem that captures the hardness of the disjointness problem is the *unique disjointness problem* (UDISJ), that is, the promise version of the disjointness problem where the two subsets are guaranteed to have at most one element in common. Denoting the binary encoding of the sets of Alice and Bob by $a, b \in \{0,1\}^n$, respectively, this amounts to computing the Boolean function $\text{UDISJ}(a, b) := 1 - a^\intercal b$ on the set of pairs $(a, b) \in \{0,1\}^n \times \{0,1\}^n$ with $a^\intercal b \in \{0,1\}$. Viewing it as a partial $2^n \times 2^n$ matrix, we call UDISJ the *unique disjointness matrix*.

It is known that the communication complexity of UDISJ is $\Omega(n)$ bits for deterministic, nondeterministic and even randomized communication protocols [Kalyanasundaram and Schnitger, 1992, Razborov, 1992, Bar-Yossef et al., 2004]. One consequence of this is that the nonnegative rank of *any* matrix obtained from UDISJ by filling arbitrarily the blank entries (for pairs (a, b) with $a^\intercal b > 1$) and perhaps adding rows and/or columns is still $2^{\Omega(n)}$. Indeed, the support of the resulting matrix has $\Omega(n)$ nondeterministic communication complexity because it contains UDISJ.

1.2 Previous Work

In a recent paper Fiorini et al. [2012] proved strong lower bounds on the size of LPs expressing the traveling salesman problem (TSP), or more precisely on the size of EFs of the TSP polytope. Their proof works by embedding the UDISJ in a slack matrix of the TSP polytope of the complete graph on $\Theta(n^4)$ vertices (a more recent version of this paper uses $\Theta(n^2)$ vertices). This solved a question left open in Yannakakis [1991]. We use a similar approach for approximate EFs, which requires lower bounds on the nonnegative rank of partial matrices obtained from the UDISJ matrix by adding a positive offset to all the entries.

Our results are closely related to previous work in communication complexity for the (unique) disjointness problem and related problems. Lower bounds of $\Omega(n)$ on the randomized, bounded error communication complexity of disjointness were established in Kalyanasundaram and Schnitger [1992]. In Razborov [1992] the distributional complexity of unique disjointness problem was analyzed, which in particular implies the result of Kalyanasundaram and Schnitger [1992]. The main tool here is Razborov's rectangle corruption lemma showing that in every large rectangle, the number of 0-entries is proportional to the number of 1-entries. This ensures that monochromatic 1-rectangles have to be small and therefore a large number is needed to cover all 1-entries; a lower bound for the nondeterministic communication complexity. It is precisely this lemma that was used in Fiorini et al. [2012] to establish lower bounds on the extension complexity of the cut polytope, the stable set polytope, and the TSP polytope. The most recent proof that the randomized, bounded error communication complexity of DISJ is $\Omega(n)$ is due to Bar-Yossef et al. [2004] and is based on information theoretic arguments. This leads to a lower bound for randomized communication within a high-error regime, that is, when the error probability is close to $1/2$. Here we derive a strong generalization dealing with shifts for approximate EFs and we recover the high-error regime bound.

There has been extensive work on LP and SDP hierarchies/relaxations and their limitations; we will be only able to list a few here. In Charikar et al. [2009], strong lower bounds (of $2 - \epsilon$) on the integrality gap for n^ϵ rounds of the Sherali-Adams hierarchy when applied to (natural relaxations of) VERTEX COVER, Max CUT, SPARSEST CUT have been been established via embeddings into ℓ_2; see also Charikar et al. [2010] for limits and tradeoffs in metric embeddings. For integrality gaps of linear (and also SDP) relaxations for the KNAPSACK problem see Karlin et al. [2011]. A nice overview of the differences and similarities of the Sherali-Adams, the Lovász-Schrijver and the Lasserre hierarchies/relaxations can be found in Laurent [2003]. Similar to the level of a hierarchy, we have the notion of *rank* for the Lovász-Schrijver relaxation and rank correspond to a similar complexity measure as the level. The rank is the minimum number of application of the Lovász-Schrijver operator N until we obtain the integral hull of the polytope under consideration.

Rank lower bounds of n for Lovász-Schrijver relaxations of CLIQUE have been obtained in Cook and Dash [2001]; a similar result for Sherali-Adams hierarchy can be found in Laurent [2003]. In Singh and Talwar [2010] integrality gaps, after adding few rounds of Chvátal-Gomory cuts, have been studied for problems including k-CSP, Max CUT, VERTEX COVER, and UNIQUE LABEL COVER showing that in some cases (e.g., k-CSP) the gap can be significantly reduced whereas in most other cases the gap remains high. In the context of SDP relaxations, in particular formulations derived from the Lovász-Schrijver N_+ hierarchies (see Lovász and Schrijver [1991]) and the Lasserre hierarchies (see Lasserre [2002]). For example, in Arora et al. [2009] an $O(\sqrt{\log n})$ upper bound on a suitable SDP relaxation of the SPARSEST CUT problem was obtained. For lower bounds in terms of rank, see e.g., Schoenebeck [2008] for the k-CSP in the Lasserre hierarchy or Schoenebeck et al. [2007] for VERTEX COVER in the semidefinite Lovász-Schrijver hierarchy. Motivated by the Unique Games Conjecture, several works studied upper and lower bounds for SDP hierarchy relaxations of Unique Games (see for example, Guruswami and Sinop [2011], Barak et al. [2011, 2012b,a]). In Fiorini et al. [2012] a characterization of semidefinite EFs via one-way quantum communication complexity is established.

Approximate EFs have been studied before, for specific problems, e.g., KNAPSACK in Bienstock [2008], or as a general tool, see Vyve and Wolsey [2006]. The idea of considering a pair P, Q as we do here first appeared in Pashkovich [2012] and similar ideas appeared earlier in Gillis and Glineur [2010]. For recent results on computation of nonnegative rank see Arora et al. [2012].

1.3 Contribution

The contribution of the present paper is threefold.

(i) We develop a framework for proving lower bounds on the sizes of approximate EFs. Through a generalization of Yannakakis' factorization theorem, we characterize the minimum size of a ρ-approximate EF as the nonnegative rank of any slack matrix of a *pair* of nested polyhedra. Thus we reduce the task of proving approximation limits for LPs to the task of obtaining lower bounds on the nonnegative ranks of associated matrices. Typically, these matrices have no zeros, which renders it impossible to use nondeterministic communication complexity. We emphasize the fact that the results obtained within our framework are unconditional. In particular, they do not rely on $\mathrm{P} \neq \mathrm{NP}$.

(ii) We extend Razborov's rectangle corruption lemma to deal with shifts of the UDISJ matrix. As a consequence, we prove that the nonnegative rank of any matrix obtained from the UDISJ matrix by adding a constant offset to every entry is still $2^{\Omega(n)}$. Moreover, the nonnegative rank is still $2^{\Omega(n^{2\epsilon})}$ when the offset is at most $n^{1/2-\epsilon}$. To our knowledge, these are the first strong lower bounds on the nonnegative rank of matrices that contain no zeros. (Furthermore, the relative difference between any two entries of some of our shifted UDISJ matrices is tiny.) Our extension of Razborov's lemma allow us to recover known lower bounds for DISJ in the high-error regime of Bar-Yossef et al. [2004].

(iii) We obtain a strong hardness result for CLIQUE w.r.t. a natural linear encoding of CLIQUE. From the results described above, we prove that the size of every $O(n^{1/2-\epsilon})$-approximate EF for CLIQUE is $2^{\Omega(n^{2\epsilon})}$. Finally, we observe that the same bounds hold for approximations of SDPs by LPs. This suggests that SDP-based approximation algorithms can be significantly stronger than LP-based approximation algorithms. The inapproximability of SDPs by LPs has some interesting consequences. In particular we cannot expect to convert SDP-based approximation algorithms into LP-based ones by approximating the PSD-cone via linear programming.

We point out that our framework readily generalizes to SDPs by replacing nonnegative rank with PSD rank (see Gouveia et al. [2011] or Fiorini et al. [2012] for a definition of the PSD rank).

Finally, we report that the results of this paper have inspired further research. In particular, Braverman and Moitra [2013] improved our lower bound on the nonnegative rank of shifted UDISJ matrices and obtain super-polynomial lower bounds for shifts up to $O(n^{1-\epsilon})$, hence matching the algorithmic hardness of approximation for CLIQUE. To achieve this, they pioneered information-theoretic methods for proving lower bounds on the nonnegative rank. An alternative information theoretic approach for lower bounding the nonnegative rank which simplifies and slightly improves the results in Braverman and Moitra [2013] has been presented in Braun and Pokutta [2013]. The latter also establishes that the hard pair that we present here has high average case and adversarial approximate extension complexity.

1.4 Outline

We begin in Section 2 by setting up our framework for studying approximate extended formulations of combinatorial optimization problems. Then we extend Razborov's rectangle corruption lemma in Section 3 and use this to prove strong lower bounds on the nonnegative rank of shifts of the UDISJ matrix. Finally, we draw consequences for CLIQUE and approximations of SDPs by LPs in Section 4.

2 Framework for Approximation Limits of LPs

In this section we establish our framework for studying approximation limits of LPs. First, we define in details the concepts of linear encodings and approximate extended formulations. Second, we prove a factorization theorem for pairs of nested polyhedra reducing existential questions on approximate EFs to the computation of nonnegative ranks of certain slack matrices.

2.1 Preliminaries

A (convex) *polyhedron* is a set $P \subseteq \mathbb{R}^d$ that is the intersection of a finite collection of closed halfspaces. In other words, P is a polyhedron if and only if P is the set of solutions of a finite system of linear inequalities and possibly equalities. (Note that every equality can be represented by a pair of inequalities.) Equivalently, a set $P \subseteq \mathbb{R}^d$ is a polyhedron if and only if P is the Minkowski sum of the convex hull $\mathrm{conv}\,(V)$ of a finite set V of points and the conical hull $\mathrm{cone}\,(R)$ of a finite set R of vectors, that is, $P = \mathrm{conv}\,(V) + \mathrm{cone}\,(R)$.

Let $P \subseteq \mathbb{R}^d$ be a polyhedron. The *dimension* of P is the dimension of its affine hull $\mathrm{aff}(P)$. A *face* of P is a subset $F := \{x \in P \mid w^\intercal x = \delta\}$ of P such that P satisfies the inequality $w^\intercal x \leqslant \delta$. Note that face F is again a polyhedron. A *vertex* is a face of dimension 0, i.e., a point. A *facet* is a face of dimension one less than P. The inequality $w^\intercal x \leqslant \delta$ is called *facet-defining* if the face F it defines is a facet. The *recession cone* $\mathrm{rec}\,(P)$ of P is the set of directions $v \in \mathbb{R}^d$ such that, for a point p in P, all points $p + \lambda v$ where $\lambda \geqslant 0$ belong to P. The recession cone of P does not depend on the base point p, and is again a polyhedron (even more, it is a polyhedral cone). The elements of the recession cone are sometimes called *rays*.

A *(convex) polytope* $P \subseteq \mathbb{R}^d$ is a bounded polyhedron. Equivalently, P is a polytope if and only if P is the convex hull $\mathrm{conv}\,(V)$ of a finite set V of points. Let $P \subseteq \mathbb{R}^d$ be a polytope. Every (finite or infinite) set V such that $P = \mathrm{conv}\,(V)$ contains all the vertices of P. Letting $\mathrm{vert}(P)$ denote the vertex set of P, then we have $P = \mathrm{conv}\,(\mathrm{vert}(P))$. Every (finite) system describing P contains all the facet-defining inequalities of P, up to scaling by positive numbers and adding equalities satisfied by *all* points of P. Conversely, a linear description of P can be obtained by picking one defining inequality per facet and adding a system of equalities describing $\mathrm{aff}(P)$. A 0/1-*polytope* in $\mathbb{R}^d$ is simply the convex hull of a subset of $\{0,1\}^d$.

2.2 Linear Encodings of Problems and Approximate EFs

A *linear encoding* of a (combinatorial optimization) problem is a pair $(\mathcal{L}, \mathcal{O})$ where $\mathcal{L} \subseteq \{0,1\}^*$ is the set of *feasible solutions* to the problem and $\mathcal{O} \subseteq \mathbb{R}^*$ is the set of *admissible objective functions*. An *instance* of the linear encoding is a pair (d, w) where d is a positive integer and $w \in \mathcal{O} \cap \mathbb{R}^d$. Solving the instance (d, w) means finding $x \in \mathcal{L} \cap \{0,1\}^d$ such that $w^\intercal x$ is either maximum or minimum, according to the type of problem under consideration.

For every fixed dimension d, a linear encoding $(\mathcal{L}, \mathcal{O})$ naturally defines a pair of nested convex sets $P \subseteq Q$ where

$$P := \operatorname{conv}\left(\{x \in \{0,1\}^d \mid x \in \mathcal{L}\}\right), \quad \text{and}$$
$$Q := \{x \in \mathbb{R}^d \mid \forall w \in \mathcal{O} \cap \mathbb{R}^d : w^\intercal x \leqslant \max\{w^\intercal x \mid x \in P\}\}$$

if the goal is to maximize and $Q := \{x \in \mathbb{R}^d \mid \forall w \in \mathcal{O} \cap \mathbb{R}^d : w^\intercal x \geqslant \min\{w^\intercal x \mid x \in P\}\}$ if the goal is to minimize. Intuitively, the vertices of P encode the feasible solutions of the problem under consideration and the defining inequalities of Q encode the admissible linear objective functions. Notice that P is always a $0/1$-polytope but Q might be unbounded and, in some pathological cases, nonpolyhedral. Below, we will mostly consider the case where Q is polyhedral, that is, defined by a finite number of "interesting" inequalities.

Given a linear encoding $(\mathcal{L}, \mathcal{O})$ of a maximization problem, and $\rho \geqslant 1$, a *ρ-approximate extended formulation* (EF) is an extended formulation $Ex + Fy = g$, $y \geqslant \mathbf{0}$ with $(x, y) \in \mathbb{R}^{d+r}$ such that

$$\max\{w^\intercal x \mid Ex + Fy = g,\ y \geqslant \mathbf{0}\} \geqslant \max\{w^\intercal x \mid x \in P\} \quad \text{for all} \quad w \in \mathbb{R}^d \quad \text{and}$$
$$\max\{w^\intercal x \mid Ex + Fy = g,\ y \geqslant \mathbf{0}\} \leqslant \rho \max\{w^\intercal x \mid x \in P\} \quad \text{for all} \quad w \in \mathcal{O} \cap \mathbb{R}^d.$$

Letting $K := \{x \in \mathbb{R}^d \mid \exists y \in \mathbb{R}^r : Ex + Fy = g,\ y \geqslant \mathbf{0}\}$, we see that this is equivalent to $P \subseteq K \subseteq \rho Q$. For a minimization problem, we require

$$\min\{w^\intercal x \mid Ex + Fy = g,\ y \geqslant \mathbf{0}\} \leqslant \min\{w^\intercal x \mid x \in P\} \quad \text{for all} \quad w \in \mathbb{R}^d \quad \text{and}$$
$$\min\{w^\intercal x \mid Ex + Fy = g,\ y \geqslant \mathbf{0}\} \geqslant \rho^{-1} \min\{w^\intercal x \mid x \in P\} \quad \text{for all} \quad w \in \mathcal{O} \cap \mathbb{R}^d.$$

This is equivalent to $P \subseteq K \subseteq \rho^{-1} Q$.

We require the following *faithfulness condition*: every instance of the problem can be mapped to an instance of the linear encoding in such a way that feasible solutions to an instance of the problem can be converted in polynomial time to feasible solutions to the corresponding instance of the linear encoding without deteriorating their objective function values, and vice-versa. Roughly speaking, we ask that each instance of the problem can be encoded as an instance of the linear encoding.

Example 3 (Max k-SAT)**.** Consider the maximum k-SAT problem (Max k-SAT), where k is constant. Letting $u_1, \ldots, u_n$ denote the variables of a Max k-SAT instance, we encode the problem in dimension $d = \Theta(n^k)$. For each nonempty clause C of size at most k, we introduce a variable x_C. Collectively, these variables define a point $x \in \mathbb{R}^d$. Given a truth assignment, we set x_C to 1 if C is satisfied and otherwise we set x_C to 0. Letting n vary, this defines a language $\mathcal{L} \subseteq \{0,1\}^*$. We let $\mathcal{O} := \{0,1\}^*$.

The pair $(\mathcal{L}, \mathcal{O})$ defines a linear encoding of Max k-SAT because each instance of Max k-SAT can be encoded as an instance of $(\mathcal{L}, \mathcal{O})$. More precisely, to any given set of clauses over n variables, we can associate a dimension $d = \Theta(n^k)$ and weight vector $w \in \{0,1\}^d$ such that maximizing $\sum w_C x_C$ for $x \in \mathcal{L} \cap \{0,1\}^d$ corresponds to finding a truth assignment that maximizes the number of satisfied clauses.

Finally, we remark that the EF defined by the inequalities $0 \leqslant x_C \leqslant 1$ and $x_C \leqslant \sum_{u_i \in C} x_{\{u_i\}} + \sum_{\bar{u}_i \in C}(1 - x_{\{u_i\}})$ for all clauses C is a polynomial-size $4/3$-approximate EF for Max k-SAT, as follows from Goemans and Williamson [1994].

2.3 Factorization Theorem for Pairs of Nested Polyhedra

Let P and Q be polyhedra with $P \subseteq Q \subseteq \mathbb{R}^d$. An *extended formulation* (EF) *of the pair* P, Q is a system $Ex + Fy = g$, $y \geqslant \mathbf{0}$ defining a polyhedron $K := \{x \in \mathbb{R}^d \mid Ex + Fy = g,\ y \geqslant \mathbf{0}\}$ such that $P \subseteq K \subseteq Q$. We denote by $\mathrm{xc}(P, Q)$ the minimum size of an EF of the pair P, Q.

Now consider an inner description $P := \mathrm{conv}(\{v_1, \ldots, v_n\}) + \mathrm{cone}(\{r_1, \ldots, r_k\})$ of P and an outer description $Q := \{x \in \mathbb{R}^d \mid Ax \leqslant b\}$ of Q, where the system $Ax \leqslant b$ consists of m inequalities: $A_1 x \leqslant b_1, \ldots, A_m x \leqslant b_m$. The *slack matrix of the pair* P, Q w.r.t. these inner and outer descriptions is the $m \times (n+k)$ matrix $S^{P,Q} = [\, S^{P,Q}_{\text{vertex}}\ S^{P,Q}_{\text{ray}} \,]$ given by block decomposition into a vertex and ray part:

$$
\begin{aligned}
S^{P,Q}_{\text{vertex}}(i,j) &:= b_i - A_i v_j, && i \in [m],\ j \in [n],\\
S^{P,Q}_{\text{ray}}(i,j) &:= -A_i r_j, && i \in [k],\ j \in [n].
\end{aligned}
$$

Our first result gives an essentially exact characterization of $\mathrm{xc}(P, Q)$ in terms of the nonnegative rank of the slack matrix of the pair P, Q. It states that the minimum extension complexity $\mathrm{xc}(P, Q)$ of a polyhedron sandwiched between P and Q equals the nonnegative rank of $S^{P,Q}$ (minus 1, in some cases). The result readily generalizes Yannakakis's factorization theorem [Yannakakis, 1991], which concerns the case $P = Q$. A result similar to ours appeared in Pashkovich [2012].

Theorem 1. *With the above notations, we have* $\mathrm{rank}_+(S^{P,Q}) - 1 \leqslant \mathrm{xc}(P, Q) \leqslant \mathrm{rank}_+(S^{P,Q})$ *for every slack matrix of the pair* P, Q*. If the affine hull of* P *is not contained in* Q *and* $\mathrm{rec}(Q)$ *is not full-dimensional, we have* $\mathrm{xc}(P, Q) = \mathrm{rank}_+(S^{P,Q})$*. In particular, this holds when* P *and* Q *are polytopes of dimension at least* 1.

Proof. First, we deal with degenerate cases. Observe that $\mathrm{xc}(P, Q) = 0$ if and only if there exists an affine subspace containing P and contained in Q, that is, if and only if the affine hull of P is contained in Q. In this case, we have $\mathrm{rank}_+(S^{P,Q}) \in \{0, 1\}$, so the theorem holds.

Now assume that the affine hull of P is not contained in Q. Then, $\mathrm{rank}_+(S^{P,Q}) \geqslant 1$ because having $\mathrm{rank}_+(S^{P,Q}) = 0$ means either that $S^{P,Q}$ is empty, that is, $m = 0$ or $n + k = 0$, or that $S^{P,Q}$ is the zero matrix. In all cases, this contradicts our assumption that the affine hull of P is not contained in Q.

Next, let $S^{P,Q} = TU$ be any rank-r nonnegative factorization of $S^{P,Q}$ with $r = \mathrm{rank}_+(S^{P,Q}) \geqslant 1$. This factorization decomposes into blocks: $S^{P,Q}_{\text{vertex}} = TU_{\text{vertex}}$ and $S^{P,Q}_{\text{ray}} = TU_{\text{ray}}$. Consider the system

$$Ax + Ty = b,\ y \geqslant \mathbf{0} \tag{1}$$

and the corresponding polyhedron $K := \{x \in \mathbb{R}^d \mid Ax + Ty = b,\ y \geqslant \mathbf{0}\}$.

We verify now that $P \subseteq K \subseteq Q$. The inclusion $K \subseteq Q$ simply follows from $Ty \geqslant \mathbf{0}$. For the inclusion $P \subseteq K$, pick a vertex v_j and observe that $(x, y) = (v_j, U^j_{\text{vertex}})$ satisfies (1), where U^j_{vertex} denotes the jth column of U, because $Av_j + TU^j_{\text{vertex}} = Av_j + b - Av_j = b$ and $U^j \geqslant \mathbf{0}$. Similarly, for every ray r_j we obtain a ray (r_j, U^j_{ray}) of K as $Ar_j + TU^j_{\text{ray}} = 0$ and $U^j_{\text{ray}} \geqslant \mathbf{0}$.

Thus we obtain that (1) is a size-r EF of the pair P, Q. Therefore, $\mathrm{xc}(P, Q) \leqslant \mathrm{rank}_+(S^{P,Q})$.

Finally, suppose that the system

$$Ex + Fy = g,\ y \geqslant \mathbf{0} \tag{2}$$

defines a size-r EF of the pair P, Q. Let $L \subseteq \mathbb{R}^{d+r}$ denote the polyhedron defined by (2), and let $K \subseteq \mathbb{R}^d$ denote the orthogonal projection of L into x-space.

Since $P \subseteq K$, for each point v_j, there exists $w_j \in \mathbb{R}^r_+$ such that $(v_j, w_j) \in L$. Similarly, for each ray r_j there exists a $z_j \in \mathbb{R}^r_+$ with (r_j, z_j) a ray of L. Let W be the matrix with columns the w_j, and Z be the matrix with columns the z_j

Since $K \subseteq Q$, by Farkas's lemma, $Ax \leqslant b$ can be derived from (2), i.e., $A = TE$, $b = Tg + c$ and $TF \geqslant 0$ for some $c \geqslant 0$ and T. This gives the factorizations $S^{P,Q}_{\text{vertex}} = (TF)W + c \cdot \mathbf{1}$ and $S^{P,Q}_{\text{ray}} = (TF)Z$, resulting in the rank-$(r+1)$ nonnegative factorization $S^{P,Q} = \left[\, TF \;\, c \,\right] \cdot \left[\begin{smallmatrix} W & Z \\ \mathbf{1} & 0 \end{smallmatrix}\right]$. Taking $r = \text{xc}(P, Q)$, we find $\text{rank}_+(S^{P,Q}) \leqslant \text{xc}(P, Q) + 1$.

Finally, when $\text{rec}\,(Q)$ is not full-dimensional, then c above can be chosen to be $\mathbf{0}$. This simplifies the factorization, and yields the sharper inequality $\text{rank}_+(S^{P,Q}) \leqslant \text{xc}(P, Q)$. □

Let P, Q be as above and $\rho \geqslant 1$. Then $\rho Q = \{x \in \mathbb{R}^d \mid Ax \leqslant \rho b\}$ and the slack matrix of the pair $P, \rho Q$ is related to the slack matrix of the pair P, Q in the following way:

$$S^{P,\rho Q}_{\text{vertex}}(i,j) = \rho b_i - A_i v_j = (\rho - 1)b_i + b_i - A_i v_j = S^{P,Q}_{ij} + (\rho - 1)b_i,$$
$$S^{P,\rho Q}_{\text{ray}}(i,j) = S^{P,Q}_{ij}.$$

Theorem 1 directly yields the following result.

Theorem 2. *Consider a maximization problem and linear encoding for this problem. Let $P, Q \subseteq \mathbb{R}^d$ be the pair of polyhedra associated with the linear encoding, and let $\rho \geqslant 1$. Consider any slack matrix $S^{P,Q}$ for the pair P, Q and the corresponding slack matrix $S^{P,\rho Q}$ for the pair $P, \rho Q$. Then the minimum size of a ρ-approximate EF of the problem, w.r.t. the considered linear encoding, is* $\text{rank}_+(S^{P,\rho Q}) + \Theta(1)$, *where the constant is* 0 *or* 1. *For a minimization problem, the minimum size of a ρ-approximate EF is* $\text{rank}_+(S^{P,\rho^{-1}Q}) + \Theta(1)$.

Fixing $\rho \geqslant 1$, Theorem 2 characterizes the minimum number of inequalities in any LP providing a ρ-approximation for the problem under consideration. We point out that the theorem directly generalizes to SDPs, by replacing nonnegative rank by PSD rank [Gouveia et al., 2011, Fiorini et al., 2012]. Here, we focus on LPs and nonnegative rank. As a matter of fact, strong lower bounds on the PSD rank seem to be currently lacking.

2.4 A Problem with no Polynomial-Size Approximate EF

We conclude this section with an example showing the necessity to restrict the set of admissible objective functions rather than allowing every $w \in \mathbb{R}^*$ (that is $P = Q$).

Let $K_n = (V_n, E_n)$ denote the n-vertex complete graph. For a set X of vertices of K_n, we let $\delta(X)$ denote the set of edges of K_n with one endpoint in X and the other in its complement $\bar{X}$. This set $\delta(X)$ is known as the *cut* defined by X. For a subset F of edges of K_n, we let $\chi^F \in \mathbb{R}^{E_n}$ denote the *characteristic vector* (or *incidence vector*) of F, with $\chi^F_e = 1$ if $e \in F$ and $\chi^F_e = 0$ otherwise. The *cut polytope* $\text{CUT}(n)$ is defined as the convex hull of the characteristic vectors of all cuts in the complete graph $K_n = (V_n, E_n)$. That is,

$$\text{CUT}(n) := \text{conv}\left(\{\chi^{\delta(X)} \in \mathbb{R}^{E_n} \mid X \subseteq V_n\}\right).$$

A related object is the *cut cone*, defined as the cone generated by the *cut-vectors* $\chi^{\delta(X)}$:

$$\text{CUT-CONE}(n) := \text{cone}\left(\{\chi^{\delta(X)} \in \mathbb{R}^{E_n} \mid X \subseteq V_n\}\right).$$

Consider the maximum cut problem (Max CUT) with *arbitrary* weights, and its usual linear encoding. With this encoding we have $P = Q = \mathrm{CUT}(n)$. Our next result states that this problem has no ρ-approximate EF, whatever $\rho \geqslant 1$ is. Intuitively, this phenomenon stems from the fact that, because $\mathbf{0}$ is a vertex of the cut polytope, every approximate EF necessarily 'captures' all facets of the cut polytope incident to $\mathbf{0}$ (see Figure 1). These facets define the cut cone, which turns out to have high extension complexity. Although this follows rather easily from ideas of Fiorini et al. [2012], we include a proof here for completeness.

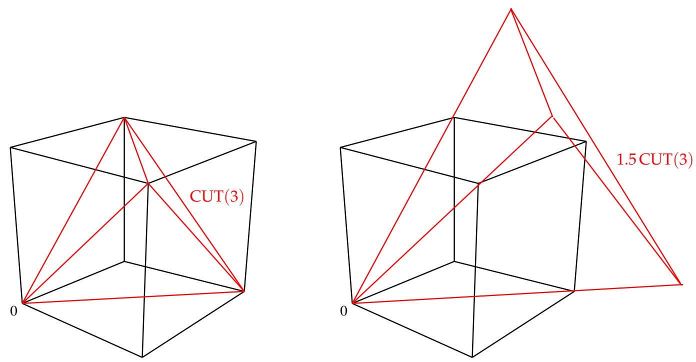

Figure 1: $\mathrm{CUT}(3)$ and a dilate $\rho\,\mathrm{CUT}(3)$ for $\rho = 1.5$.

Proposition 3. *For every $\rho \geqslant 1$, every ρ-approximate EF of the Max CUT problem with arbitrary weights has $2^{\Omega(n)}$ size. More precisely, disregarding the value of $\rho \geqslant 1$, we have $\mathrm{xc}(\mathrm{CUT}(n), \rho\,\mathrm{CUT}(n)) = 2^{\Omega(n)}$.*

Proof. Let $Ex + Fy = g$, $y \geqslant \mathbf{0}$ denote a minimum size ρ-approximate EF of $\mathrm{CUT}(n)$. We claim that

$$Ex + Fy - \lambda g = 0,\ y \geqslant \mathbf{0},\ \lambda \geqslant \mathbf{0} \tag{3}$$

is an EF of the cut cone. Let K be the polyhedron obtained by projecting the set of solutions of (3) into x-space. Clearly, K is a cone containing all the cut-vectors $\chi^{\delta(X)}$, from which we get that $\mathrm{CUT\text{-}CONE}(n) \subseteq K$. Now take any point (x, y, λ) satisfying (3). If $\lambda = 0$ then necessarily $x = 0$ because $Ex + Fy = 0, y \geqslant 0$ defines the recession cone of a polyhedron that projects into $\rho\,\mathrm{CUT}(n)$, which is bounded. In this case we have $x = 0 \in \mathrm{CUT\text{-}CONE}(n)$. Assume that $\lambda > 0$. Then $E\lambda^{-1}x + F\lambda^{-1}y = g$ and $\lambda^{-1}y \geqslant 0$ which implies that $\lambda^{-1}x$ is in $\rho\,\mathrm{CUT}(n)$. Thus $\rho^{-1}\lambda^{-1}x$ is in $\mathrm{CUT}(n)$ and x is thus a positive combination of cut-vectors, hence $x \in \mathrm{CUT\text{-}CONE}(n)$. This yields $K \subseteq \mathrm{CUT\text{-}CONE}(n)$. In conclusion, $K = \mathrm{CUT\text{-}CONE}(n)$ and (3) is an EF of the cut cone. The size of this EF is at most $2r + \binom{n}{2}$, where r denotes the size of the given ρ-approximate EF of $\mathrm{CUT}(n)$. Thus $\mathrm{xc}(\mathrm{CUT\text{-}CONE}(n)) \leqslant 2r + \binom{n}{2}$.

By using the correlation mapping (see [Laurent and Deza, 1997, p. 55]), the cut cone has the same extension complexity as its corresponding *correlation cone*, defined as

$$\mathrm{COR\text{-}CONE}(n-1) := \mathrm{cone}\left(\left\{ \begin{pmatrix} b_0 \\ b \end{pmatrix} \begin{pmatrix} b_0 \\ b \end{pmatrix}^{\intercal} \,\middle|\, b_0 \in \{0,1\}, b \in \{0,1\}^{n-2} \right\}\right).$$

We claim that the unique disjointness matrix on $[n-2]$ can be embedded in a slack matrix of $\mathrm{COR\text{-}CONE}(n-1)$. To prove this, consider the $(n-1) \times (n-1)$ rank-1 positive semidefinite matrices

$$T_a := \begin{pmatrix} -1 \\ a \end{pmatrix} \begin{pmatrix} -1 \\ a \end{pmatrix}^{\intercal} \qquad \text{and} \qquad U^b := \begin{pmatrix} 1 \\ b \end{pmatrix} \begin{pmatrix} 1 \\ b \end{pmatrix}^{\intercal} \tag{4}$$

where $a, b \in \{0,1\}^{n-2}$. The Frobenius inner product $\langle T_a, z\rangle \geqslant 0$ of T_a with any correlation matrix $z = \binom{b_0}{b}\binom{b_0}{b}^\intercal$ is nonnegative because both matrices are positive semidefinite. Thus $\langle T_a, z\rangle \geqslant 0$ is valid for all points $z \in$ COR-CONE$(n-1)$, for all $a \in \{0,1\}^{n-2}$. Moreover, $\langle T_a, U^b\rangle = (1 - a^\intercal b)^2$ for all $a, b \in \{0,1\}^{n-2}$ and thus $\langle T_a, U^b\rangle = \text{UDISJ}(a, b)$ provided $a^\intercal b \in \{0,1\}$.

From what precedes, the slack of correlation matrix U^b with respect to the valid inequality $\langle T_a, z\rangle \geqslant 0$ is UDISJ(a, b) provided $a^\intercal b \in \{0,1\}$. Therefore, COR-CONE$(n-1)$ has a slack matrix that contains UDISJ on $[n-2]$. Because the nonnegative rank of any matrix containing UDISJ is $2^{\Omega(n)}$ (this follows from [Razborov, 1992], see [Fiorini et al., 2012, Theorem 1]), we conclude that the nonnegative rank of some slack matrix of COR-CONE$(n-1)$ is $2^{\Omega(n)}$. From the Theorem 1 applied to $P = Q =$ COR-CONE$(n-1)$, it follows that xc(COR-CONE$(n-1)) = 2^{\Omega(n)}$. Thus we get

$$2r + \binom{n}{2} \geqslant \text{xc}(\text{CUT-CONE}(n)) = \text{xc}(\text{COR-CONE}(n-1)) = 2^{\Omega(n)}$$

from which we obtain $r \geqslant \frac{1}{2}2^{\Omega(n)} - \binom{n}{2} = 2^{\Omega(n)}$. The result then follows immediately. □

3 Extension of Razborov's Lemma and Shifts of Unique Disjointness

In the first subsection we generalize Razborov's famous lemma on the disjointness problem (see Razborov [1992] or Kushilevitz and Nisan [1997, Lemma 4.49] for the original version). In the next subsection we apply it to shift the UDISJ matrix without significantly decreasing its nonnegative rank, which will be used in later sections to obtain lower bounds on approximate extended formulations.

The main improvements to Razborov's lemma are threefold: (i) the dependence in the error parameter ϵ is made explicit; (ii) better analytical estimations are employed to improve overall strength of the statement; (iii) probabilities are generalized to expected values to homogenize the proof and allow arbitrary nonnegative rank-1 matrices instead of rectangles, which is more natural for nonnegative rank lower bounds.

3.1 Extension of Razborov's Rectangle Corruption Lemma

Suppose that $n \equiv 3 \pmod 4$ and let

$$\ell := \frac{n+1}{4}.$$

We define the following distribution μ on pairs (a, b) of subsets of $[n]$. We flip a biased coin and with probability $1/4$ and choose (a, b) uniformly among the pairs of ℓ-subsets intersecting in exactly one element; with probability $3/4$, we choose (a, b) uniformly among the pairs of disjoint ℓ-subsets.

Let A denote the event that a and b are disjoint ℓ-subsets, and B denote the event that a and b are ℓ-subsets intersecting in exactly one element (B for "Barely intersecting"). We have

$$A = \{(a, b) \mid |a \cap b| = 0,\ |a| = |b| = \ell\}, \quad \text{and} \quad B = \{(a, b) \mid |a \cap b| = 1,\ |a| = |b| = \ell\}.$$

We see that

$$\mathbb{P}[A] = \frac{3}{4}, \qquad \mathbb{P}[B] = \frac{1}{4}$$

and the conditional distribution of μ given either A or B is uniform.

Lemma 4. *Let n, ℓ, μ, A and B be as above. For every nonnegative functions f and g defined on $2^{[n]} \times 2^{[n]}$ we introduce a random variable X such that $X(a, b) := f(a)g(b)$. Then for every $0 < \epsilon < 1$:*

$$(1-\epsilon)\,\mathbb{E}[X \mid A] - \mathbb{E}[X \mid B] \leqslant \|X \restriction (A \cup B)\|_\infty 2^{-\frac{\epsilon^2}{16\ln 2}\ell + O(\log \ell)}, \tag{5}$$

where the constant in the $O(\log \ell)$ is absolute, and $X \restriction (A \cup B)$ denotes the restriction of X to $A \cup B$.

We note that the lemma holds for all nonnegative functions f and g if and only if it holds for all *binary* functions f and g. In this case X is the indicator[4] of a rectangle R, that is $X = I_R$, and (5) becomes

$$\frac{4}{3}(1-\epsilon)\,\mathbb{P}\left[R \cap A\right] - 4\,\mathbb{P}\left[R \cap B\right] \leqslant 2^{-\frac{\epsilon^2}{16\ln 2}\ell + O(\log \ell)},$$

which is a strengthened version of Razborov's original lemma. In the proof below, whenever this helps the intuition, the reader can assume that X is the indicator of a rectangle.

Our proof is inspired by the version in Kushilevitz and Nisan [1997, Lemma 4.49] and we adopt similar notations.

Proof of Lemma 4. The proof is in four main steps.

Step 1: Redefinition of distribution μ and rewriting of $\mathbb{E}\left[X \mid A\right]$ and $\mathbb{E}\left[X \mid B\right]$. First, we redefine the distribution μ in an alternative fashion. Let $T = (T_1, T_2, \{i\})$ be a uniformly chosen partition of $[n]$ into two subsets T_1, T_2 with $2\ell - 1$ elements and one singleton $\{i\}$. Given T we choose ℓ-subsets a and b independently in $T_1 \cup \{i\}$ and $T_2 \cup \{i\}$, respectively. We flip a fair coin to decide whether i is an element of a. With probability $1/2$, we select a as a uniform random ℓ-subset of $T_1 \cup \{i\}$ containing $\{i\}$. With probability $1/2$, we choose a as a uniform random ℓ-subset of T_1. We choose b similarly by using T_2 instead of T_1. Note that the probability of B under this distribution is $\mathbb{P}\left[B\right] = \mathbb{P}\left[i \in a, i \in b\right] = (1/2)^2 = 1/4$. For A we get $\mathbb{P}\left[A\right] = 1 - \mathbb{P}\left[A\right] = 1 - 1/4 = 3/4$. Hence, by symmetry reasons, the conditional distribution is uniform given either A or B, hence the marginal distribution on pairs (a, b) is exactly μ.

We begin by rewriting $\mathbb{E}\left[X \mid B\right]$ and then $\mathbb{E}\left[X \mid A\right]$ in terms of the following functions of T:

$$\begin{aligned} \mathrm{Row}_0(T) &:= \mathbb{E}\left[f(a) \mid T, i \notin a\right], & \mathrm{Row}_1(T) &:= \mathbb{E}\left[f(a) \mid T, i \in a\right], \\ \mathrm{Col}_0(T) &:= \mathbb{E}\left[g(b) \mid T, i \notin b\right], & \mathrm{Col}_1(T) &:= \mathbb{E}\left[g(b) \mid T, i \in b\right]. \end{aligned}$$

We note the following nice interpretation of $\mathrm{Row}_0(T) + \mathrm{Row}_1(T)$ and $\mathrm{Col}_0(T) + \mathrm{Col}_1(T)$, that we will use at the end of the proof:

$$\begin{aligned} \mathbb{E}\left[f(a) \mid T\right] &= \underbrace{\mathbb{E}\left[f(a) \mid T, i \in a\right]}_{\mathrm{Row}_1(T)} \cdot \underbrace{\mathbb{P}\left[i \in a \mid T\right]}_{1/2} + \underbrace{\mathbb{E}\left[f(a) \mid T, i \notin a\right]}_{\mathrm{Row}_0(T)} \cdot \underbrace{\mathbb{P}\left[i \notin a \mid T\right]}_{1/2} \\ &= \frac{\mathrm{Row}_0(T) + \mathrm{Row}_1(T)}{2}, \\ \mathbb{E}\left[g(b) \mid T\right] &= \frac{\mathrm{Col}_0(T) + \mathrm{Col}_1(T)}{2}. \end{aligned} \tag{6}$$

Note that: (i) the distribution of (a, b) conditioned on a given T is a product distribution (this local independence property is the main reason why we reinterpret the distribution μ); (ii) the marginal distributions of a conditioned on $(T, i \in a, i \in b)$ and $(T, i \in a)$ are the same (and similarly for b, we can remove the condition $i \in a$). From these facts, we get

$$\begin{aligned} \mathbb{E}\left[X \mid B\right] &= \mathbb{E}\left[f(a)g(b) \mid i \in a, i \in b\right] \\ &= \mathbb{E}\left[\mathbb{E}\left[f(a)g(b) \mid T, i \in a, i \in b\right]\right] \\ &= \mathbb{E}\left[\mathbb{E}\left[f(a) \mid T, i \in a, i \in b\right] \mathbb{E}\left[g(b) \mid T, i \in a, i \in b\right]\right] \\ &= \mathbb{E}\left[\mathbb{E}\left[f(a) \mid T, i \in a\right] \mathbb{E}\left[g(b) \mid T, i \in b\right]\right] \\ &= \mathbb{E}\left[\mathrm{Row}_1(T)\,\mathrm{Col}_1(T)\right]. \end{aligned} \tag{7}$$

[4]We write I_C for the indicator of any event C.

By similar arguments, we find

$$\begin{aligned}\mathbb{E}\left[X \mid A\right] &= \frac{1}{3}\,\mathbb{E}\left[f(a)g(b) \mid i \notin a, i \notin b\right] + \frac{1}{3}\,\mathbb{E}\left[f(a)g(b) \mid i \in a, i \notin b\right] + \frac{1}{3}\,\mathbb{E}\left[f(a)g(b) \mid i \notin a, i \in b\right] \\ &= \frac{1}{3}\,\mathbb{E}\left[\mathrm{Row}_0(T)\,\mathrm{Col}_0(T)\right] + \frac{1}{3}\,\mathbb{E}\left[\mathrm{Row}_1(T)\,\mathrm{Col}_0(T)\right] + \frac{1}{3}\,\mathbb{E}\left[\mathrm{Row}_0(T)\,\mathrm{Col}_1(T)\right].\end{aligned}$$

Pick an $(2\ell - 1)$-subset T_2 of $[n]$, that we consider fixed for the time being. The marginal distributions of a conditioned on the events T_2, $(T_2, i \in a)$ and $(T_2, i \notin a)$ are the same, namely, the uniform distribution on the ℓ-subsets of $[n] \setminus T_2$. (Note that fixing T_2 does not fix i, which could be any element of $[n] \setminus T_2$.) Thus, we get

$$\mathbb{E}\left[f(a) \mid T_2, i \notin a\right] = \mathbb{E}\left[f(a) \mid T_2, i \in a\right] = \mathbb{E}\left[f(a) \mid T_2\right]. \tag{8}$$

On the other hand, we have

$$\begin{aligned}\mathbb{E}\left[\mathrm{Row}_0(T) \mid T_2\right] &= \mathbb{E}\left[\mathbb{E}\left[f(a) \mid T, i \notin a\right] \mid T_2\right] \\ &= \mathbb{E}\left[\frac{\mathbb{E}\left[f(a) I_{i \notin a} \mid T\right]}{\mathbb{P}\left[i \notin a \mid T\right]} \,\middle|\, T_2\right] \\ &= 2\,\mathbb{E}\left[f(a) I_{i \notin a} \mid T_2\right] \\ &= \mathbb{E}\left[f(a) \mid T_2, i \notin a\right]\end{aligned} \tag{9}$$

and similarly

$$\mathbb{E}\left[\mathrm{Row}_1(T) \mid T_2\right] = \mathbb{E}\left[f(a) \mid T_2, i \in a\right].$$

From (8), we conclude

$$\mathbb{E}\left[\mathrm{Row}_0(T) \mid T_2\right] = \mathbb{E}\left[\mathrm{Row}_1(T) \mid T_2\right].$$

Therefore (letting T_2 vary),

$$\begin{aligned}\mathbb{E}\left[\mathrm{Row}_1(T)\,\mathrm{Col}_0(T)\right] &= \mathbb{E}\left[\mathbb{E}\left[\mathrm{Row}_1(T)\,\mathrm{Col}_0(T) \mid T_2\right]\right] \\ &= \mathbb{E}\left[\mathbb{E}\left[\mathrm{Row}_1(T) \mid T_2\right]\mathrm{Col}_0(T)\right] \\ &= \mathbb{E}\left[\mathbb{E}\left[\mathrm{Row}_0(T) \mid T_2\right]\mathrm{Col}_0(T)\right] \\ &= \mathbb{E}\left[\mathbb{E}\left[\mathrm{Row}_0(T)\,\mathrm{Col}_0(T) \mid T_2\right]\right] \\ &= \mathbb{E}\left[\mathrm{Row}_0(T)\,\mathrm{Col}_0(T)\right].\end{aligned}$$

The second and fourth equalities above are due to the fact that $\mathrm{Col}_0(T)$ is constant when T_2 is fixed. This is because $\mathrm{Col}_0(T) = \mathbb{E}\left[g(b) \mid T, i \notin b\right]$ depends only on T_2, as the marginal distribution of b given $(T, i \notin b)$ is uniform on the ℓ-subsets of T_2.

Exchanging the roles of rows and columns, we have

$$\mathbb{E}\left[\mathrm{Row}_1(T)\,\mathrm{Col}_0(T)\right] = \mathbb{E}\left[\mathrm{Row}_0(T)\,\mathrm{Col}_0(T)\right].$$

In conclusion, we find the following simple expression for $\mathbb{E}\left[X \mid A\right]$:

$$\mathbb{E}\left[X \mid A\right] = \mathbb{E}\left[\mathrm{Row}_0(T)\,\mathrm{Col}_0(T)\right]. \tag{10}$$

Step 2: Estimation of $\mathbb{E}[X \mid A] - \mathbb{E}[X \mid B]$. Using the inequality $|x| \geqslant x$, which is valid for all $x \in \mathbb{R}$, we get

$$\begin{aligned}\mathrm{Row}_1(T) &\geqslant \mathrm{Row}_0(T) - |\,\mathrm{Row}_0(T) - \mathrm{Row}_1(T)|, \quad \text{and}\\ \mathrm{Col}_1(T) &\geqslant \mathrm{Col}_0(T) - |\,\mathrm{Col}_0(T) - \mathrm{Col}_1(T)|.\end{aligned}$$

Thus

$$\begin{aligned}&\mathrm{Row}_0(T)\,\mathrm{Col}_0(T) - \mathrm{Row}_1(T)\,\mathrm{Col}_1(T)\\ &\leqslant \mathrm{Row}_0(T)\,\mathrm{Col}_0(T) - (\mathrm{Row}_0(T) - |\,\mathrm{Row}_0(T) - \mathrm{Row}_1(T)|) \cdot (\mathrm{Col}_0(T) - |\,\mathrm{Col}_0(T) - \mathrm{Col}_1(T)|)\\ &\leqslant \mathrm{Row}_0(T)|\,\mathrm{Col}_0(T) - \mathrm{Col}_1(T)| + |\,\mathrm{Row}_0(T) - \mathrm{Row}_1(T)|\,\mathrm{Col}_0(T).\end{aligned} \tag{11}$$

This argument is depicted on Figure 2.

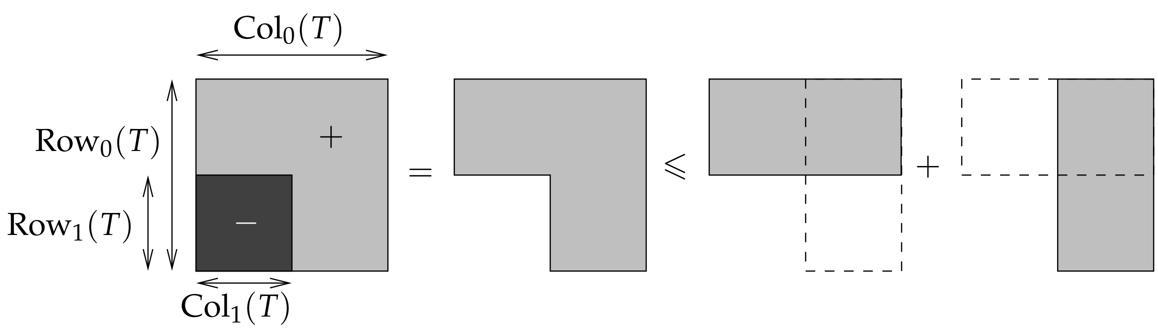

Figure 2: The estimation of $\mathrm{Row}_0(T)\,\mathrm{Col}_0(T) - \mathrm{Row}_1(T)\,\mathrm{Col}_1(T)$.

In Step 3 below, we will define two events, row-big(T) and column-big(T). The event small(T) is such that small(T) holds if and only if *any* of the row-big(T) and column-big(T) does not hold. Thus

$$1 = I_{\text{row-big}(T)\cap\text{column-big}(T)} + I_{\text{small}(T)}. \tag{12}$$

From (11),

$$\begin{aligned}&(\mathrm{Row}_0(T)\,\mathrm{Col}_0(T) - \mathrm{Row}_1(T)\,\mathrm{Col}_1(T)) \cdot I_{\text{row-big}(T)\cap\text{column-big}(T)}\\ &\leqslant (\mathrm{Row}_0(T)|\,\mathrm{Col}_0(T) - \mathrm{Col}_1(T)| + |\,\mathrm{Row}_0(T) - \mathrm{Row}_1(T)|\,\mathrm{Col}_0(T)) \cdot I_{\text{row-big}(T)\cap\text{column-big}(T)}\\ &\leqslant \mathrm{Row}_0(T)|\,\mathrm{Col}_0(T) - \mathrm{Col}_1(T)| \cdot I_{\text{column-big}(T)} + |\,\mathrm{Row}_0(T) - \mathrm{Row}_1(T)|\,\mathrm{Col}_0(T) \cdot I_{\text{row-big}(T)}.\end{aligned}$$

Moreover, we obviously have

$$(\mathrm{Row}_0(T)\,\mathrm{Col}_0(T) - \mathrm{Row}_1(T)\,\mathrm{Col}_1(T)) \cdot I_{\text{small}(T)} \leqslant \mathrm{Row}_0(T)\,\mathrm{Col}_0(T) \cdot I_{\text{small}(T)}.$$

Below, we will prove

$$\mathbb{E}\left[\mathrm{Row}_0(T)|\,\mathrm{Col}_0(T) - \mathrm{Col}_1(T)| \cdot I_{\text{column-big}(T)}\right] \leqslant \frac{\epsilon}{2}\,\mathbb{E}\left[\mathrm{Row}_0(T)\,\mathrm{Col}_0(T)\right], \tag{13}$$

$$\mathbb{E}\left[|\,\mathrm{Row}_0(T) - \mathrm{Row}_1(T)|\,\mathrm{Col}_0(T) \cdot I_{\text{row-big}(T)}\right] \leqslant \frac{\epsilon}{2}\,\mathbb{E}\left[\mathrm{Row}_0(T)\,\mathrm{Col}_0(T)\right], \quad \text{and} \tag{14}$$

$$\mathbb{E}\left[\mathrm{Row}_0(T)\,\mathrm{Col}_0(T) \cdot I_{\text{small}(T)}\right] \leqslant \|X \restriction (A \cup B)\|_\infty\, 2^{-\frac{\epsilon^2}{16\ln 2} - O(\log \ell)} \tag{15}$$

By (7), (10) and (12), these upper bounds imply

$$\begin{aligned}
&\mathbb{E}[X \mid A] - \mathbb{E}[X \mid B] \\
&= \mathbb{E}[\mathrm{Row}_0(T)\,\mathrm{Col}_0(T) - \mathrm{Row}_1(T)\,\mathrm{Col}_1(T)] \\
&= \mathbb{E}\left[(\mathrm{Row}_0(T)\,\mathrm{Col}_0(T) - \mathrm{Row}_1(T)\,\mathrm{Col}_1(T)) \cdot (I_{\text{row-big}(T)\cap\text{column-big}(T)} + I_{\text{small}(T)})\right] \\
&\leqslant 2\frac{\epsilon}{2}\,\mathbb{E}[\mathrm{Row}_0(T)\,\mathrm{Col}_0(T)] + \|X \upharpoonright (A \cup B)\|_\infty 2^{-\frac{\epsilon^2}{16\ln 2}\ell - O(\log \ell)} \\
&= \epsilon\,\mathbb{E}[X \mid A] + \|X \upharpoonright (A \cup B)\|_\infty 2^{-\frac{\epsilon^2}{16\ln 2}\ell - O(\log \ell)}
\end{aligned}$$

from which the result clearly follows, by rearranging.

Step 3. One-sided error estimation via entropy argument in the "big" case. Let $\delta > 0$ be a constant to be chosen later. Essentially, δ will be the coefficient of ℓ in the exponent. Let row-big(T) denote the event $\mathbb{E}[f(a) \mid T_2] > 2^{-\delta\ell-1}\left\|f \upharpoonright \binom{[n]\setminus T_2}{\ell}\right\|_\infty$ where $f \upharpoonright \binom{[n]\setminus T_2}{\ell}$ denotes the restriction of f to ℓ-subsets of $[n] \setminus T_2$. The event column-big(T) is defined in a similar way. These events depend only on T_2 and T_1, respectively.

Let T_2 be fixed and assume that row-big(T) holds. In particular $\mathbb{E}[f(a) \mid T_2]$ is positive. Because $\binom{2\ell-1}{\ell-1} = \binom{2\ell-1}{\ell}$, the probability of a given T_2 is the same as the probability of a given T, for every fixed choice of i. Thus, we have

$$\mathbb{E}[f(a) \mid T_2] = \sum_{\substack{x\subseteq[n]\setminus T_2 \\ |x|=\ell}} \frac{1}{\binom{2\ell}{\ell}} f(x) = \mathbb{E}[f(a) \mid T].$$

(This holds when $f(a)$ is replaced by any function of a.)

We can define s as a random ℓ-subset of $[n] \setminus T_2$ with distribution

$$\mathbb{P}[s = x \mid T_2] = \frac{f(x)}{\binom{2\ell}{\ell}\mathbb{E}[f(a) \mid T_2]} = \frac{f(x)}{\sum_{\substack{y\subseteq[n]\setminus T_2 \\ |y|=\ell}} f(y)} \leqslant \frac{2^{\delta\ell+1}}{\binom{2\ell}{\ell}}.$$

Let us introduce the shorthand notation $\lambda := \mathbb{P}[i \in s \mid T_2]$. Then

$$\lambda = \frac{\sum_{\substack{x\subseteq[n]\setminus T_2 \\ |x|=\ell,\ x\ni i}} f(x)}{\sum_{\substack{y\subseteq[n]\setminus T_2 \\ |y|=\ell}} f(y)} = \frac{\frac{1}{\binom{2\ell}{\ell}}\sum_{\substack{x\subseteq[n]\setminus T_2 \\ |x|=\ell,\ x\ni i}} f(x)}{\frac{1}{\binom{2\ell}{\ell}}\sum_{\substack{y\subseteq[n]\setminus T_2 \\ |y|=\ell}} f(y)} = \frac{\mathbb{E}[f(a) I_{i\in a} \mid T]}{\mathbb{E}[f(a) \mid T_2]}.$$

Hence,

$$\mathrm{Row}_1(T) = 2\,\mathbb{E}[f(a) I_{i\in a} \mid T] = 2\,\mathbb{E}[f(a) \mid T_2] \cdot \mathbb{P}[i \in s \mid T_2] = 2\lambda\,\mathbb{E}[f(a) \mid T_2], \tag{16}$$

$$\mathrm{Row}_0(T) = 2\,\mathbb{E}[f(a) I_{i\notin a} \mid T] = 2\,\mathbb{E}[f(a) \mid T_2] \cdot \mathbb{P}[i \notin s \mid T_2] = 2(1-\lambda)\,\mathbb{E}[f(a) \mid T_2]. \tag{17}$$

We now estimate the entropy of s. On the one hand, by subadditivity of the entropy, we get the following upperbound on $H(s \mid T_2)$:

$$H(s \mid T_2) \leqslant \sum_{j\in[n]\setminus T_2} H\left(I_{j\in s} \,\middle|\, T_2\right) = 2\ell\,\mathbb{E}[H(\lambda) \mid T_2].$$

In this last equation, $H(\lambda)$ denotes the binary entropy of λ. On the other hand, we get a lower bound on $H(s \mid T_2)$ from our upper bound on the distribution of s (which induces "flatness" of the distribution):

$$\begin{aligned} H(s \mid T_2) &= \sum_x \mathbb{P}\left[s = x \mid T_2\right] \log \frac{1}{\mathbb{P}\left[s = x \mid T_2\right]} \\ &\geqslant \sum_x \mathbb{P}\left[s = x \mid T_2\right] \log \frac{\binom{2\ell}{\ell}}{2^{\delta\ell+1}} = \log \frac{\binom{2\ell}{\ell}}{2^{\delta\ell+1}} = 2\ell\left(1 - \frac{\delta}{2} - O\left(\frac{\log \ell}{\ell}\right)\right). \end{aligned}$$

This implies

$$\frac{\delta}{2} + O\left(\frac{\log \ell}{\ell}\right) \geqslant \mathbb{E}\left[1 - H(\lambda) \mid T_2\right]. \tag{18}$$

To estimate this expression, we use the Taylor expansion of the binary entropy function at $1/2$:

$$1 - H(x) \geqslant \frac{(1-2x)^2}{2\ln 2}.$$

Hence (18) yields

$$\frac{\delta}{2} + O\left(\frac{\log \ell}{\ell}\right) \geqslant \frac{\mathbb{E}\left[(1-2\lambda)^2 \,\middle|\, T_2\right]}{2\ln 2} \geqslant \frac{\left(\mathbb{E}\left[|1-2\lambda| \mid T_2\right]\right)^2}{2\ln 2}.$$

From (8), (9) we have $\mathbb{E}\left[f(a) \mid T_2\right] = \mathbb{E}\left[\mathrm{Row}_0(T) \mid T_2\right]$. Using (17) and (16), we derive

$$\begin{aligned} \mathbb{E}\left[|\mathrm{Row}_0(T) - \mathrm{Row}_1(T)| \mid T_2\right] &= \mathbb{E}\left[|2(1-\lambda)\,\mathbb{E}\left[f(a) \mid T_2\right] - 2\lambda\,\mathbb{E}\left[f(a) \mid T_2\right]| \mid T_2\right] \\ &= 2\,\mathbb{E}\left[|1-2\lambda| \mid T_2\right]\mathbb{E}\left[f(a) \mid T_2\right] \\ &\leqslant 2\sqrt{\delta'}\,\mathbb{E}\left[\mathrm{Row}_0(T) \mid T_2\right]. \end{aligned}$$

with

$$\delta' := \left(\delta + O\left(\frac{\log \ell}{\ell}\right)\right)\ln 2. \tag{19}$$

We now globalize to prove (14):

$$\begin{aligned} &\mathbb{E}\left[|\,\mathrm{Row}_0(T) - \mathrm{Row}_1(T)|\,\mathrm{Col}_0(T) I_{\text{row-big}(T)}\right] \\ &= \mathbb{E}\left[\mathbb{E}\left[|\,\mathrm{Row}_0(T) - \mathrm{Row}_1(T)|\,\mathrm{Col}_0(T) I_{\text{row-big}(T)} \,\middle|\, T_2\right]\right] \\ &= \mathbb{E}\left[\mathbb{E}\left[|\,\mathrm{Row}_0(T) - \mathrm{Row}_1(T)| I_{\text{row-big}(T)} \,\middle|\, T_2\right]\mathrm{Col}_0(T)\right] \\ &\leqslant \mathbb{E}\left[2\sqrt{\delta'}\,\mathbb{E}\left[\mathrm{Row}_0(T) \mid T_2\right]\mathrm{Col}_0(T)\right] \\ &= 2\sqrt{\delta'}\,\mathbb{E}\left[\mathrm{Row}_0(T)\,\mathrm{Col}_0(T)\right] \end{aligned}$$

We require $\frac{\epsilon}{2} = 2\sqrt{\delta'}$, from which we express δ in terms of ϵ using (19):

$$\delta = \frac{\delta'}{\ln 2} - O\left(\frac{\log \ell}{\ell}\right) = \frac{\epsilon^2}{16\ln 2} - O\left(\frac{\log \ell}{\ell}\right)$$

This concludes the proof of (14). Equation (13) follows by exchanging rows and columns.

Step 4: Error estimation in the "small" case. Suppose that for some given T, $\mathrm{small}(T)$ holds because row-big(T) does not hold (the argument is similar in case column-big(T) does not hold). Then, using (6),

$$\mathrm{Row}_0(T) \leqslant \mathrm{Row}_0(T) + \mathrm{Row}_1(T) = 2\,\mathbb{E}\left[f(a) \mid T\right].$$

Thus

$$\begin{aligned}\mathrm{Row}_0(T)\,\mathrm{Col}_0(T) &\leqslant 2\,\mathbb{E}\left[f(a) \mid T\right] \cdot \mathbb{E}\left[g(b) \mid T, i \notin b\right] \\ &\leqslant 2^{-\delta\ell} \left\| f(a) \restriction \binom{[n] \setminus T_2}{\ell} \right\|_\infty \cdot \left\| g(b) \restriction \binom{T_2}{\ell} \right\|_\infty \\ &\leqslant 2^{-\delta\ell} || f(a) g(b) \restriction (A \cup B) ||_\infty\end{aligned}$$

This is easily seen to imply (15). $\square$

3.2 Lower Bounds for Shifts of Unique Disjointness

Now we apply Lemma 4 to show that the nonnegative rank (and hence the complexity of computation in expectation) of any shifted version of the unique disjointness matrix remains high. More precisely, let $M \in \mathbb{R}_+^{2^n \times 2^n}$; for convenience we index the rows and columns with elements in $\{0,1\}^n$. We say that M is a *ρ-extension of UDISJ*, if $M_{ab} = \rho$ whenever $|a \cap b| = 0$ and $M_{ab} = \rho - 1$ whenever $|a \cap b| = 1$ with $a, b \in \{0,1\}^n$. Note that for these pairs M has exclusively positive entries whenever $\rho > 1$. For $\rho = 1$ a nonnegative rank of $2^{\Omega(n)}$ was already shown in Fiorini et al. [2012] via nondeterministic communication complexity. We now extend this result for a wide range of ρ using Lemma 4.

Theorem 5 (Nonnegative rank of UDISJ shifts)**.** *Let $M \in \mathbb{R}_+^{2^n \times 2^n}$ be a ρ-extension of UDISJ as above.*

(i) If ρ is a fixed constant, then $\mathrm{rank}_+(M) = 2^{\Omega(n)}$.

(ii) If $\rho = O(n^\beta)$ for some constant $\beta < 1/2$ then $\mathrm{rank}_+(M) = 2^{\Omega(n^{1-2\beta})}$.

Proof. Without loss of generality, assume $n \equiv 3 \pmod 4$. Let $r = \mathrm{rank}_+(M)$. Regarding the $2^n \times 2^n$ matrix M as a function from $2^{[n]} \times 2^{[n]}$ to $\mathbb{R}$, we can write $M = \sum_{i=1}^r X_i$ where $X_i(a,b) = f_i(a) g_i(b)$ for some nonnegative functions f_i and g_i defined over $2^{[n]}$. Then

$$\mathbb{E}\left[M \mid A\right] = \rho \quad \text{and} \quad \mathbb{E}\left[M \mid B\right] = \rho - 1.$$

On the other hand, by applying Lemma 4 to each $i \in [r]$ and summing up all equations we find

$$\begin{aligned}(1-\epsilon)\,\mathbb{E}\left[M \mid A\right] - \mathbb{E}\left[M \mid B\right] &\leqslant \sum_{i=1}^r \|X_i \restriction (A \cup B)\|_\infty 2^{-\frac{\epsilon^2}{16\ln 2}\ell + O(\log \ell)} \\ &\leqslant r\,\|M \restriction (A \cup B)\|_\infty 2^{-\frac{\epsilon^2}{16\ln 2}\ell + O(\log \ell)}\end{aligned}$$

where $\ell = \frac{n+1}{4}$ as before. By plugging in the values of $\mathbb{E}\left[M \mid A\right]$, $\mathbb{E}\left[M \mid B\right]$ and $\|M \restriction (A \cup B)\|_\infty$, we get

$$(1-\epsilon)\rho - \rho + 1 \leqslant r \cdot \rho \cdot 2^{-\frac{\epsilon^2}{16\ln 2}\ell + O(\log \ell)} \iff r \geqslant \left(\frac{1}{\rho} - \epsilon\right) 2^{\frac{\epsilon^2}{16\ln 2}\ell - O(\log \ell)}.$$

If ρ is constant, this last expression is $2^{\Omega(n)}$ provided ϵ is chosen sufficiently close to 0. This proves part (i) of the theorem.

If $\rho \leqslant Cn^\beta$ for some positive constant C, then we can take $\epsilon = \frac{1}{2Cn^\beta}$. Thus $\frac{1}{\rho} - \epsilon \geqslant \frac{1}{2Cn^\beta} = \Omega(n^{-\beta})$. This leads to the lower bound $r \geqslant 2^{\Omega(n^{1-2\beta})}$ as claimed in part (ii). $\square$

4 Polyhedral Inapproximability of CLIQUE and SDPs

We will now use Theorem 5 in combination with Theorem 2 to lower bound the sizes of certain approximate EFs. First, we pinpoint a pair P, Q of nested polyhedra that will be the source of our polyhedral inapproximability results. Second, we give a faithful linear encoding of CLIQUE and prove strong lower bounds on the sizes of approximate EFs for CLIQUE w.r.t. this encoding. Third, we focus on approximations of SDPs by LPs.

4.1 A Hard Pair

Let n be a positive integer. The *correlation polytope* $\mathrm{COR}(n)$ is defined as the convex hull of all the $n \times n$ rank-1 binary matrices of the form $bb^\intercal$ where $b \in \{0,1\}^n$. In other words,

$$\mathrm{COR}(n) = \mathrm{conv}\left(\{bb^\intercal \mid b \in \{0,1\}^n\}\right).$$

This will be our inner polytope P. Next, let

$$Q = Q(n) := \{x \in \mathbb{R}^{n \times n} \mid \langle 2\,\mathrm{diag}(a) - aa^\intercal, x\rangle \leqslant 1,\ a \in \{0,1\}^n\},$$

where $\langle \cdot, \cdot \rangle$ denotes the Frobenius inner product. This will be our outer polyhedron Q.

Then the following is known, see [Fiorini et al., 2012]. First, $P \subseteq Q$. Second, denoting by $S^{P,Q}$ the slack matrix of the pair P, Q, we have $S^{P,Q}_{ab} = (1 - a^\intercal b)^2$. Thus, for $\rho \geqslant 1$, we have $S^{P,\rho Q}_{ab} = (1 - a^\intercal b)^2 + \rho - 1$. Observe that the matrix $S^{P,\rho Q}$ is a ρ-extension of UDISJ and therefore has high nonnegative rank via Theorem 5; moreover it has positive entries everywhere for $\rho > 1$. Together with Theorem 1 this implies that every polyhedron sandwiched between $P = \mathrm{COR}(n)$ and ρQ has large extension complexity. We obtain the following theorem.

Theorem 6 (Lower bounds for approximate EFs of the hard pair)**.** *Let $\rho \geqslant 1$, let n be a positive integer and let $P = \mathrm{COR}(n)$, $Q = Q(n)$ be as above. Then the following hold:*

(i) If ρ is a fixed constant, then $\mathrm{xc}(P, \rho Q) = 2^{\Omega(n)}$.

(ii) If $\rho = O(n^\beta)$ for some constant $\beta < 1/2$, then $\mathrm{xc}(P, \rho Q) = 2^{\Omega(n^{1-2\beta})}$.

4.2 Polyhedral Inapproximability of CLIQUE

We define a natural linear encoding for the maximum clique problem (CLIQUE) as follows. Let n denote the number of vertices of the input graph. We define a $d = n^2$ dimensional encoding. The variables are denoted by x_{ij} for $i, j \in [n]$. Thus $x \in \mathbb{R}^{n \times n}$. The interpretation is that a set of vertices X is encoded by $x_{ij} = 1$ if $i, j \in X$ and $x_{ij} = 0$ otherwise. Note that $X = \{i : x_{ii} = 1\}$ can be recovered from only the diagonal variables. This defines the set $\mathcal{L} \subseteq \{0,1\}^*$ of feasible solutions. Notice that $x \in \{0,1\}^{n \times n}$ is feasible if and only if it is of the form $x = bb^\intercal$ for some $b \in \{0,1\}^n$, the characteristic vector of X. Thus we have $P = \mathrm{COR}(n)$ for the inner polytope.

An objective function $w \in \mathbb{R}^{n \times n}$ is admissible if $w_{ii} \in \{0,1\}$ for the diagonal coefficients and $w_{ij} = w_{ji} \in \{-1, 0\}$ for the off-diagonal coefficients. This defines the set $\mathcal{O} \subseteq \{-1,0,1\}^*$ of admissible objective functions.

Given a graph G such that $V(G) \subseteq [n]$, we let $w_{ii} := 1$ for $i \in V(G)$, $w_{ii} := 0$ for $i \in [n] \setminus V(G)$, $w_{ij} = w_{ji} := -1$ when ij is a non-edge of G, and $w_{ij} = w_{ji} := 0$ otherwise. We denote the resulting weight vector by w^G. Notice that for a graph G with $V(G) = [n]$, we have $w^G = I - A(\overline{G})$ where I is the $n \times n$ identity matrix, $A(\overline{G})$ is the adjacency matrix of the complement of G. A feasible solution $x = bb^\intercal \in \{0,1\}^{n \times n}$ maximizes $\langle w^G, x\rangle$ only if b is the characteristic vector (or incidence vector) of a clique of G. Indeed, if $b = \chi^X$ and ij is a non-edge of G with $i, j \in X$ then removing i or j from X increases $\langle w, x\rangle$. Moreover, the maximum of $\langle w^G, x\rangle$ over $x \in \{0,1\}^{n\times n}$ feasible is the clique number $\omega(G)$. Therefore, $(\mathcal{L}, \mathcal{O})$ defines a valid linear encoding of CLIQUE. We denote the outer convex set of this linear encoding by Q^{all}. It is actually the polyhedron defined as $Q^{\text{all}} = \{x \in \mathbb{R}^{n\times n} \mid \forall \text{ graphs } G \text{ s.t. } V(G) \subseteq [n] : \langle w^G, x\rangle \leqslant \omega(G),\ \forall i \neq j \in [n] : x_{ij} \geqslant 0\}$. We will now show that $Q^{\text{all}} \subseteq Q$.

Lemma 7. *Let Q^{all}, Q be as above, then $Q^{\text{all}} \subseteq Q$.*

Proof. We will show that Q is a relaxation of Q^{all} by restricting to stable sets $a \subseteq [n]$ as graphs. Let $G = (a, \emptyset)$ be the stable set supported on a with $a \subseteq [n]$. Slightly abusing notation, we will also identify a with its characteristic vector in $\{0,1\}^n$. With the definition from above we have $\omega^a = I_a - A(K_a)$. Now let $x \in \mathbb{R}^{n\times n}$ be arbitrary. Clearly, $\langle I_a, x\rangle = \langle \operatorname{diag}(a), x\rangle$ and we have $\langle A(K_a), x\rangle = \langle aa^T, x\rangle - \langle \operatorname{diag}(a), x\rangle$, where the latter summand comes from the fact that in $\langle aa^T, x\rangle$ we count the main diagonal, as purported edges (i,i) although they are not counted in $\langle A(K_a), x\rangle$. All in all we obtain

$$\langle \omega^a, x\rangle = \langle I_a - A(K_a), x\rangle = \langle \operatorname{diag}(a), x\rangle - (\langle aa^T, x\rangle - \langle \operatorname{diag}(a), x\rangle) = \langle 2\operatorname{diag}(a), x\rangle - \langle aa^T, x\rangle \leq 1,$$

where the last inequality follows from $\omega(G) = 1$ as G is a stable set. We conclude $Q^{\text{all}} \subseteq Q$. ☐

Because Q^{all} is contained in the polyhedron Q defined above, every K satisfying $P \subseteq K \subseteq \rho Q^{\text{all}}$ also satisfies $P \subseteq K \subseteq \rho Q$. Hence, Theorem 6 yields the following result.

Theorem 8 (Polyhedral inapproximability of CLIQUE)**.** *W.r.t. the linear encoding defined above, CLIQUE has an $O(n^2)$-size n-approximate EF. Moreover, every $n^{1/2-\epsilon}$-approximate EF of CLIQUE has size $2^{\Omega(n^{2\epsilon})}$, for all $0 < \epsilon < 1/2$.*

Proof. The n-approximate EF of CLIQUE is trivial: it is defined by the system $\mathbf{0} \leqslant x \leqslant \mathbf{1}$, or in slack form $x - y = \mathbf{0}$, $x + z = \mathbf{1}$, $y \geqslant \mathbf{0}$, $z \geqslant \mathbf{0}$. We claim that this defines a n-approximate EF of CLIQUE of size $2n^2$. Indeed, letting $K = [0,1]^{n\times n}$ denote the polytope defined by this EF, we have $P \subseteq K$. Moreover, $\max\{\langle w, x\rangle \mid x \in K\} \leqslant n \leqslant n \cdot \max\{\langle w, x\rangle \mid x \in P\}$ for all admissible objective functions w of dimension $n \times n$ with a nonzero diagonal. In case an admissible w has $w_{ii} = 0$ for all $i \in [n]$, we have $\max\{\langle w, x\rangle \mid x \in K\} = 0 = \max\{\langle w, x\rangle \mid x \in P\}$. Our claim and the first part of the theorem follows.

The second part of the theorem follows directly from Theorem 6 and the fact that $Q^{\text{all}} \subseteq Q$. ☐

4.3 Polyhedral Inapproximability of SDPs

In this section we show that there exists a spectrahedron with small semidefinite extension complexity but high approximate extension complexity; i.e., any sufficiently fine polyhedral approximation is large. This indicates that in general it is not possible to approximate SDPs arbitrarily well using LPs, so that SDPs are indeed a much stronger class of optimization problems. (The situation looks quite different for SOCPs, see Ben-Tal and Nemirovski [2001].) The result follows from Theorem 6 and Fiorini et al. [2012].

We denote the cone of all $r \times r$ symmetric positive semidefinite matrices (shortly, the PSD cone) by $\mathbb{S}^r_+$. A *semidefinite EF* of a convex set $S \subseteq \mathbb{R}^d$ is a system $Ex + Fy = g$, $y \in \mathbb{S}^r_+$ such that $x \in S$ if and only if $\exists y \in \mathbb{R}^{r(r+1)/2}$ with $Ex + Fy = g$, $y \in \mathbb{S}^r_+$. Thus a convex set admits a semidefinite EF if and only if it is a spectrahedron. The *size* of the semidefinite EF $Ex + Fy = g$, $y \in \mathbb{S}^r_+$ is simply r. The *semidefinite extension complexity* of a spectrahedron $S \subseteq \mathbb{R}^d$ is the minimum size of a semidefinite EF of S. This is denoted by $\operatorname{xc}_{SDP}(S)$. A *rank-$r$ PSD-factorization* of a nonnegative matrix $M \in \mathbb{R}^{m\times n}$ is given by two vectors $U \in (\mathbb{S}^r_+)^m$ and $V \in (\mathbb{S}^r_+)^n$, so that $M_{ij} = U_i V^j$ where the scalar product is the Frobenius product; the *PSD-rank of M* is the smallest r such that there exists such a factorization. Yannakakis' factorization theorem can be generalized to the SDP-case (see Fiorini et al. [2012]), i.e., the semidefinite extension complexity is equal to the PSD-rank of any associated slack matrix.

Let $P = \mathrm{COR}(n)$ be the correlation polytope and $Q = Q(n) \subseteq \mathbb{R}^{n\times n}$ be the polyhedron defined above in Section 4.1. Although every polyhedron K sandwiched between P and Q has superpolynomial extension complexity (by Theorem 6, this even applies to polyhedra sandwiched between P and ρQ for $\rho = O(n^{1/2-\epsilon})$), there exists a spectrahedron S sandwiched between P and Q with small semidefinite extension complexity.

Lemma 9 (Existence of spectrahedron)**.** *Let n be a positive integer and let $P = \mathrm{COR}(n)$, $Q = Q(n)$ be as above. Then there exists a spectrahedron S in $\mathbb{R}^{n\times n}$ with $P \subseteq S \subseteq Q$ and $\mathrm{xc}_{SDP}(S) \leqslant n+1$.*

Proof. For $a, b \in \{0,1\}^n$, the matrices $T^a, U_b \in \mathbb{S}_+^{n+1}$ defined in (4) satisfy $\langle T^a, U_b\rangle = (1 - a^\intercal b)^2$. Let $M = M(n) \in \mathbb{R}^{2^n\times 2^n}$ be the matrix defined as $M_{ab} = (1 - a^\intercal b)^2$. The matrix M is a $O(n)$-rank nonnegative matrix extending the UDISJ matrix, and also the slack matrix of the pair P, Q. Then $M = TU$ is a rank-$(n+1)$ PSD-factorization of M.

For convenience write $Q = \{x \in \mathbb{R}^{n\times n} \mid Ax \leqslant \mathbf{1}\}$ with $Ax \leqslant \mathbf{1}$ being the defining system from Section 4.1. Now consider the system $Ax + Ty = \mathbf{1}$, $y \in \mathbb{S}_+^{n+1}$ and $S := \{x \in \mathbb{R}^{n\times n} \mid \exists y : Ax + Ty = \mathbf{1},\ y \in \mathbb{S}_+^{n+1}\}$. First observe that $S \subseteq Q$: since $T^a \in \mathbb{S}_+^{n+1}$ for all $a \in \{0,1\}^n$ and $y \in \mathbb{S}_+^{n+1}$ we have $Ty \geqslant \mathbf{0}$ and thus $Ax \leqslant \mathbf{1}$ holds for all $x \in S$.

In order to show that $P \subseteq S$ recall that M is the slack matrix of the pair P, Q. Therefore, for each vertex $x := bb^\intercal$ of P, we can pick $y := U_b$ from the factorization such that $Ax + Ty = Ax + \mathbf{1} - Ax = \mathbf{1}$ and $y \in \mathbb{S}_+^{n+1}$. It follows that $P \subseteq S$. □

Our final result is the following inapproximability theorem for spectrahedra.

Theorem 10 (Polyhedral inapproximability of SDPs)**.** *Let $\rho \geqslant 1$, and let n be a positive integer. Then there exists a spectrahedron $S \subseteq \mathbb{R}^{n\times n}$ with $\mathrm{xc}_{SDP}(S) \leqslant n+1$ such that for every polyhedron K with $S \subseteq K \subseteq \rho S$ the following hold:*

(i) If ρ is a fixed constant, then $\mathrm{xc}(K) = 2^{\Omega(n)}$.

(ii) If $\rho = O(n^\beta)$ for some constant $\beta < 1/2$, then $\mathrm{xc}(K) = 2^{\Omega(n^{1-2\beta})}$.

Proof. We define S as in Lemma 9. Hence, we have $P \subseteq S \subseteq Q$ and $\mathrm{xc}_{SDP}(S) \leqslant n+1$. As $\mathbf{0} \in S$ this implies in particular $P \subseteq S \subseteq \rho S \subseteq \rho Q$ for $\rho \geqslant 1$. If now K is a polyhedron such that $S \subseteq K \subseteq \rho S$ then it follows $P \subseteq K \subseteq \rho Q$. The result follows from Theorem 6. □

5 Concluding Remarks

We have introduced a general framework to study approximation limits of small LP relaxations. Given a polyhedron Q encoding admissible objective functions and a polytope P encoding feasible solutions, we have proved that any LP relaxation sandwiched between P and a dilate ρQ has extension complexity at least the nonnegative rank of the slack matrix of the pair P, ρQ.

This yields a lower bound depending *only* on the linear encoding of the problem at hand, and applies *independently* of the structure of the actual relaxation. By doing so, we obtain unconditional lower bounds on integrality gaps for small LP relaxations, which hold even in the unlikely event that $P = NP$.

We have proved that every polynomial-size LP relaxation for (a natural linear encoding of) CLIQUE has essentially an $\Omega(\sqrt{n})$ integrality gap.

Finally, our work sheds more light on the inherent limitations of LPs in the context of combinatorial optimization and approximation algorithms, in particular, in comparison to SDPs. We provide strong evidence that certain approximation guarantees can only be achieved via non-LP-based techniques (e.g., SDP-based or combinatorial).

We are convinced that our framework can be used to obtain strong approximation limits for (LP relaxations of) of other well-known problems such as Max CUT, Max k-SAT and VERTEX COVER. The following important question remains open.

Is it possible to show a constant-factor polyhedral inapproximability for Max CUT with non-negative weights (and similarly for VERTEX COVER and many more) for any polynomial-size LP? We conjecture that it is not possible to approximate Max CUT with LPs of poly-size within a factor better than 2. This would be in stark contrast with the ratio achieved by the SDP-based algorithm of Goemans and Williamson [1995] which is known to be optimal, assuming the Unique Games Conjecture Khot [2002], Khot et al. [2004], Mossel et al. [2005].

So far no strong lower bounding technique for semidefinite EFs are known. It is plausible that in the near future we will see lower bounding techniques on the PSD rank that would be suited for studying approximation limits of SDPs. (We remark however that such bounds should not only argue on the zero/nonzero pattern of a slack matrix.)

References

S. Arora, B. Bollobás, and L. Lovász. Proving integrality gaps without knowing the linear program. In *Proc. FOCS*, pages 313–322, 2002.

S. Arora, S. Rao, and U. Vazirani. Expander flows, geometric embeddings and graph partitioning. *J. ACM*, 56(2):5, 2009.

S. Arora, R. Ge, R. Kannan, and A. Moitra. Computing a nonnegative matrix factorization–Provably. accepted for STOC 2012, 2012.

Z. Bar-Yossef, T. Jayram, R. Kumar, and D. Sivakumar. An information statistics approach to data stream and communication complexity. *J. Comput. System Sci.*, 68(4):702–732, 2004.

B. Barak, P. Raghavendra, and D. Steurer. Rounding semidefinite programming hierarchies via global correlation. In *Proc. FOCS*, pages 472–481. IEEE, 2011.

B. Barak, F. G. Brandão, A. Harrow, J. Kelner, D. Steurer, and Y. Zhou. Hypercontractivity, sum-of-squares proofs, and their applications. In *STOC*, 2012a. To appear.

B. Barak, P. Gopalan, J. Håstad, R. Meka, P. Raghavendra, and D. Steurer. Making the long code shorter, 2012b. Manuscript.

A. Ben-Tal and A. Nemirovski. On polyhedral approximations of the second-order cone. *Math. Oper. Res.*, 26:193–205, 2001.

D. Bienstock. Approximate formulations for 0-1 knapsack sets. *Oper. Res. Lett.*, 36(3):317–320, 2008.

G. Braun and S. Pokutta. Common information and unique disjointness. *submitted*, 2013.

M. Braverman and A. Moitra. An information complexity approach to extended formulations. To appear in Proc. STOC, 2013.

R. D. Carr, G. Konjevod, G. Little, V. Natarajan, and O. Parekh. Compacting cuts: A new linear formulation for minimum cut. *ACM Trans. Algorithms*, 5(3):27:1–27:16, July 2009. ISSN 1549-6325. doi: 10.1145/1541885.1541888. URL `http://doi.acm.org/10.1145/1541885.1541888`.

M. Charikar, K. Makarychev, and Y. Makarychev. Integrality gaps for Sherali-Adams relaxations. In *Proc. STOC*, pages 283–292. ACM, 2009.

M. Charikar, K. Makarychev, and Y. Makarychev. Local global tradeoffs in metric embeddings. *SIAM J. Comput.*, 39(6):2487–2512, 2010.

A. Chattopadhyay and T. Pitassi. The story of set disjointness. *SIGACT News*, 41:59–85, 2010.

M. Conforti, G. Cornuéjols, and G. Zambelli. Extended formulations in combinatorial optimization. *4OR*, 8:1–48, 2010.

W. Cook and S. Dash. On the matrix-cut rank of polyhedra. *Math. Oper. Res.*, 26:19–30, 2001.

Y. Faenza, S. Fiorini, R. Grappe, and H. R. Tiwary. Extended formulations, non-negative factorizations and randomized communication protocols. arXiv:1105.4127, 2011.

S. Fiorini, S. Massar, S. Pokutta, and R. de Wolf. Linear vs. semidefinite extended formulations: Exponential separation and strong lower bounds. accepted for STOC 2012, 2012.

N. Gillis and F. Glineur. On the geometric interpretation of the nonnegative rank. arXiv:1009.0880, 2010.

M. X. Goemans and D. P. Williamson. A new 3/4-approximation algorithm for max sat. *SIAM J. Discrete Math.*, 7:313–321, 1994.

M. X. Goemans and D. P. Williamson. Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *J. Assoc. Comput. Mach.*, 42:1115–1145, 1995.

J. Gouveia, P. A. Parrilo, and R. Thomas. Lifts of convex sets and cone factorizations. arXiv:1111.3164, 2011.

V. Guruswami and A. K. Sinop. Lasserre hierarchy, higher eigenvalues, and approximation schemes for quadratic integer programming with PSD objectives. In *FOCS*, 2011.

M. Held and R. Karp. The traveling salesman problem and minimum spanning trees. *Oper. Res.*, 18:1138–1162, 1970.

B. Kalyanasundaram and G. Schnitger. The probabilistic communication complexity of set intersection. *SIAM J. Discrete Math.*, 5:545–557, 1992.

A. R. Karlin, C. Mathieu, and C. T. Nguyen. Integrality gaps of linear and semi-definite programming relaxations for knapsack. In *Proc. IPCO*, pages 301–314, 2011.

S. Khot. On the power of unique 2-prover 1-round games. In *Proc. STOC*, pages 767–775, 2002.

S. Khot, G. Kindler, E. Mossel, and R. O'Donnell. Optimal inapproximability results for Max-Cut and other 2-variable CSPs? In *Proc. FOCS*, pages 146–154, 2004.

E. Kushilevitz and N. Nisan. *Communication complexity*. Cambridge University Press, 1997.

J. B. Lasserre. An explicit equivalent positive semidefinite program for nonlinear 0-1 programs. *SIAM J. Optim.*, 12:756–769, 2002.

L. Lau, R. Ravi, and M. Singh. *Iterative Methods in Combinatorial Optimization*. Cambridge Texts in Applied Mathematics. Cambridge, 2011.

M. Laurent. A comparison of the Sherali-Adams, Lovász-Schrijver, and Lasserre relaxations for 0-1 programming. *Math. Oper. Res.*, pages 470–496, 2003.

M. Laurent and M. Deza. *Geometry of Cuts and Metrics*. Springer-Verlag, Berlin, 1997.

L. Lovász and A. Schrijver. Cones of matrices and set-functions and 0-1 optimization. *SIAM J. Optim.*, 1:166–190, 1991.

E. Mossel, R. O'Donnell, and K. Oleszkiewicz. Noise stability of functions with low influences invariance and optimality. In *Proc. FOCS*, pages 21–30, 2005.

K. Pashkovich. *Extended Formulations for Combinatorial Polytopes*. PhD thesis, Magdeburg Universität, 2012.

A. A. Razborov. On the distributional complexity of disjointness. *Theoret. Comput. Sci.*, 106(2): 385–390, 1992.

G. Schoenebeck. Linear level Lasserre lower bounds for certain k-CSPs. In *Proc. FOCS*, pages 593–602. IEEE, 2008.

G. Schoenebeck, L. Trevisan, and M. Tulsiani. A linear round lower bound for Lovász-Schrijver SDP relaxations of vertex cover. In *Proc. CCC*, pages 205–216. IEEE, 2007.

A. Schrijver. *Combinatorial optimization. Polyhedra and efficiency*. Springer-Verlag, Berlin, 2003.

H. D. Sherali and W. P. Adams. A hierarchy of relaxations between the continuous and convex hull representations for zero-one programming problems. *SIAM J. Discrete Math.*, 3:411–430, 1990.

M. Singh and K. Talwar. Improving integrality gaps via Chvátal-Gomory rounding. In *Approximation, randomization, and combinatorial optimization*, volume 6302 of *Lecture Notes in Comput. Sci.*, pages 366–379. Springer, 2010.

V. V. Vazirani. *Approximation algorithms*. Springer-Verlag, Berlin, 2001. ISBN 3-540-65367-8.

M. Vyve and L. Wolsey. Approximate extended formulations. *Math. Program.*, 105(2):501–522, 2006.

D. P. Williamson and D. B. Shmoys. *The design of approximation algorithms*. Cambridge University Press, Cambridge, 2011.

L. Wolsey. Heuristic analysis, linear programming and branch and bound. *Math. Programming Stud.*, 13:121–134, 1980.

M. Yannakakis. Expressing combinatorial optimization problems by linear programs (extended abstract). In *Proc. STOC*, pages 223–228, 1988.

M. Yannakakis. Expressing combinatorial optimization problems by linear programs. *J. Comput. System Sci.*, 43(3):441–466, 1991.

S. Zhang. Quantum strategic game theory. In *Proc. ITCS 2012*, pages 39–59, 2012.

G. M. Ziegler. *Lectures on Polytopes*, volume 152 of *Graduate Texts in Mathematics*. Springer-Verlag, Berlin, 1995.