2012 IEEE 53rd Annual Symposium on Foundations of Computer Science

Approximation Limits of Linear Programs (Beyond Hierarchies)

Gábor Braun
Institut für Informatik,
Universität Leipzig,
PF 100920, 04009 Leipzig, Germany.
gabor.braun@informatik.uni-leipzig.de

Samuel Fiorini
Department of Mathematics,
Université libre de Bruxelles CP 216,
Bd. du Triomphe, 1050 Brussels, Belgium.
sfiorini@ulb.ac.be

Sebastian Pokutta
ISyE,
Georgia Institute of Technology,
Atlanta, GA 30332, USA.
sebastian.pokutta@isye.gatech.edu

David Steurer
Department of Computer Science,
Cornell University,
Ithaca, NY 14853, United States.
dsteurer@cs.princeton.edu

***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 quadratic approximations for CLIQUE require linear programs of exponential size. (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 certain perturbations of the unique disjointness matrix.

***Keywords*-extended formulations; communication complexity; nonnegative rank; approximation algorithms; polyhedral combinatorics**

I. Introduction

A. Context

Linear programs (LPs) play a central role in the design of approximation algorithms, see, e.g., [1], [2]. 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 [3], studied the limitations of LPs captured by *lift-and-project procedures* or *hierarchies* (e.g., [4], [5]). See the previous work section below for a more detailed account of the relevant literature.

In this work, we develop a framework for a third generation of results that apply to *any* LP for a given problem. For example, our lower bounds address the following question: Are there linear programming relaxations LP_n for CLIQUE of size $\mathrm{poly}(n)$ that achieve $O(1)$-approximations for all graphs with at most n vertices. (In this sense, we prove lower bounds in a model for *non-uniform* computation, whereas hierarchy lower bounds apply to models for uniform computation.)

Although we mainly focus on LPs, our framework readily generalizes to semidefinite programs (SDPs).

Linear Encodings: We consider combinatorial optimization problems 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

0272-5428/12 $26.00 © 2012 IEEE
DOI 10.1109/FOCS.2012.10

linear encoding is specified by a dimension d and an admissible objective function $w \in \mathbb{R}^d$. Solving the instance means finding a feasible solution $x \in \{0,1\}^d$ such that $w^\intercal x$ 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 the linear encoding is *faithful*, i.e., there is a bijection between the instances of the problem and the instances of the linear encoding such that feasible solutions of the two encodings can be converted in polynomial time to each other without deteriorating their objective function values. 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.

For example, with the natural linear encoding of the metric traveling salesman problem (metric TSP) the feasible solutions are the incidence vectors of tours of the complete graph over $[n] := \{1,2,\dots,n\}$ 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 i, j and k in $[n]$. All vectors are encoded in $\mathbb{R}^d$, where $d = \binom{n}{2}$.

Coming back to the general case, 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) and for minimization problems Q is defined by all linear inequalities of the form $w^\intercal x \geqslant \zeta$ where w is an admissible objective function of dimension d and ζ the minimum value of w. For maximization problems Q is defined analogously by the $w^\intercal x \leqslant \zeta$.

(Approximate) Extended Formulations: Returning to our previous example, it is known that the Held-Karp relaxation K of the metric TSP has integrality gap at most $3/2$ (see [6], [7]). 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. That is, the Held-Karp relaxation K has a polynomial-size extended formulation.

Formally, an *extended formulation* (EF) of a polytope $K \subseteq \mathbb{R}^d$ 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. Notice that an an EF can always be brought into *slack form* $Ex + Fy = g$, $y \geqslant \mathbf{0}$ without increasing 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 polytope 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 hierarchies.

Let $\rho \geqslant 1$. Then $Ex + Fy = g$, $y \geqslant \mathbf{0}$ is a ρ-approximate EF of a given maximization problem, w.r.t. a given linear encoding, 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 $P \subseteq K \subseteq \rho Q$, where P and Q come from the linear encoding as defined above. For minimization problems, the definitions are similar with ρ replaced by ρ^{-1}, i.e., $P \subseteq K \subseteq \rho^{-1} Q$.

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 size $m \times r$ and $r \times n$, respectively. The *nonnegative rank* $\mathrm{rank}_+(M)$ of M is the minimum rank r of nonnegative factorizations of M. 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 [8] (see [9] for the conference version) states that extension complexity of a polytope K is precisely the nonnegative rank of any of its slack matrices. Let L be the convex hull of $\{v_1, \dots, v_n\} \subseteq \mathbb{R}^d$, which is the set of solutions to $A_1 x \leqslant b_1, \dots, A_m x \leqslant b_m$ at the same time. The *slack matrix* of L 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 $\mathrm{xc}(L) = \mathrm{rank}_+(S)$ for every polytope L and every slack matrix S of L.

The Link to Communication Complexity: Yannakakis's factorization theorem initiated an interplay between the extension complexity of polytopes and (classical) communication complexity.[1] 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*. [11] show that, considering M as a matrix, the minimum complexity of a protocol computing M in expectation equals $\log(\mathrm{rank}_+(M)) + \Theta(1)$. 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. 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. Similarly, whenever the variance is not too large, a protocol computing M in expectation can be turned into a randomized protocol computing M with high probability [11].

(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 [12] for a survey.

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 $\mathrm{UDISJ}(a,b) := 1 - a^\intercal b$ 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*.

As the communication complexity of UDISJ is $\Omega(n)$, 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.

B. Previous Work

In a recent paper, [13] 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. This solved a question left open in [8]. 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 an 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 [14]. In [15] the distributional complexity of unique disjointness problem was analyzed, which in particular implies the result of [14]. 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 [13] to establish lower bounds on

[1] We also assume some familiarity with communication complexity. See [10].

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 [16] and is based on information theoretic arguments. Here we derive a strong generalization dealing with perturbations for approximate EFs.

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 [17], strong lower bounds (of $2-\varepsilon$) 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 [18] for limits and tradeoffs in metric embeddings. For integrality gaps of linear (and also SDP) relaxations for the KNAPSACK problem see [19]. A nice overview of the differences and commonalities of the Sherali-Adams, the Lovász-Schrijver and the Lasserre hierarchies/relaxations can be found in [20]. Rank lower bounds of n for Lovász-Schrijver relaxations of CLIQUE have been obtained in [21]; a similar result for Sherali-Adams hierarchy can be found in [20]. Also, hierarchies based on SDPs have been widely studied, in particular formulations derived from the Lovász-Schrijver N_+ hierarchies (see [5]) and the Lasserre hierarchies (see [22]). For example, in [23] 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., [24] for the k-CSP in the Lasserre hierarchy or [25] 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, [26], [27], [28], [29]). In [13] a characterization of semidefinite EF via one-way quantum communication complexity is established.

Approximate EFs have been studied before, for specific problems, e.g., KNAPSACK in [30], or as a general tool, see [31]. The idea of considering a pair of polytopes P, Q first appeared in [32] and similar ideas appeared earlier in [33]. For recent results on computing the nonnegative rank see [34].

C. Contribution

(i) We develop a new framework for proving lower bounds on the sizes of approximate EFs. Through a generalization of Yannakakis's 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. We emphasize the fact that the results obtained within our framework are unconditional. In particular, they do not rely on $P \neq NP$.

(ii) We extend Razborov's rectangle corruption lemma to deal with perturbations 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 $2^{\Omega(n^{2\varepsilon})}$ when the offset is at most $n^{1/2-\varepsilon}$. To our knowledge, these are the first strong lower bounds on the nonnegative rank of matrices that contain no zeros. Our extension of Razborov's lemma allow us to recover known lower bounds for DISJ in the high-error regime of [16].

(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 $O(n^{1/2-\varepsilon})$-approximate EF for CLIQUE is $2^{\Omega(n^{2\varepsilon})}$. We see this as the first step in obtaining lower bounds on the sizes of approximate EFs for (faithful linear encodings of) other problems. 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. In particular we cannot expect to convert SDP-based approximation algorithms into LP-based ones by approximating the PSD-cone via linear programming.

Finally, we point out that our framework readily generalizes to SDPs by replacing nonnegative rank with PSD rank (see [35] or [13] for a definition of the PSD rank).

D. Outline

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

II. A Framework for Approximation Limits of LPs

In this section we establish the basics of our framework for studying approximation limits of LPs.

A. Linear Encodings and Approximate EFs

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

For every fixed dimension d, a linear encoding $(\mathscr{L},\mathscr{O})$ naturally defines a pair of nested convex sets $P \subseteq Q$ where $P := \operatorname{conv}\left(\{x \in \{0,1\}^d \mid x \in \mathscr{L}\}\right)$, $Q := \{x \in \mathbb{R}^d \mid \forall w \in \mathscr{O} \cap \mathbb{R}^d : w^\intercal x \leqslant \max\{w^\intercal x \mid x \in P\}\}$ if the goal is to maximize and $Q := \{x \in \mathbb{R}^d \mid \forall w \in \mathscr{O} \cap \mathbb{R}^d : w^\intercal x \geqslant \min\{w^\intercal x \mid x \in P\}\}$ otherwise. 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 $(\mathscr{L},\mathscr{O})$ 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 $\max\{w^\intercal x \mid Ex+Fy=g,\ y \geqslant \mathbf{0}\} \geqslant \max\{w^\intercal x \mid x \in P\}$ for all $w \in \mathbb{R}^d$ and $\max\{w^\intercal x \mid Ex+Fy=g,\ y \geqslant \mathbf{0}\} \leqslant \rho \max\{w^\intercal x \mid x \in P\}$ for all $w \in \mathscr{O} \cap \mathbb{R}^d$. 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 $\min\{w^\intercal x \mid Ex+Fy=g,\ y \geqslant \mathbf{0}\} \leqslant \min\{w^\intercal x \mid x \in P\}$ for all $w \in \mathbb{R}^d$ and $\min\{w^\intercal x \mid Ex+Fy=g,\ y \geqslant \mathbf{0}\} \geqslant \rho^{-1} \min\{w^\intercal x \mid x \in P\}$ for all $w \in \mathscr{O} \cap \mathbb{R}^d$. 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.

For example, consider the maximum k-SAT problem (Max k-SAT), where k is constant. We encode it in dimension $d = \Theta(n^k)$ with one variable x_C for each nonempty clause C with at most k literals. Given a truth assignment, we set x_C to 1 if C is satisfied and otherwise we set x_C to 0. Letting the number n of variables vary, this defines a language $\mathscr{L} \subseteq \{0,1\}^d$. We let $\mathscr{O} := \{0,1\}^d$: the set of weight vectors for finding the maximal number of satisfied clauses in a given subset. The pair $(\mathscr{L},\mathscr{O})$ defines a linear encoding of Max k-SAT.

Finally, let $u_1,\dots,u_n$ denote the variables of a Max k-SAT instance. Then 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 [36].

B. Factoring a Pair of Nested Polyhedra

Let P be a polytope and Q be a polyhedron 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 $\operatorname{xc}(P,Q)$ the minimum size of an EF of the pair P,Q. Note that the special case $P=Q$ reduces to the extension complexity $\operatorname{xc}(P) = \operatorname{xc}(P,P)$ of P.

Now, consider an inner description of P and an outer description of Q, say $P := \text{conv}(V)$ and $Q := \{x \in \mathbb{R}^d \mid Ax \leqslant b\}$ where $V := \{v_1, \ldots, v_n\}$ and $Ax \leqslant b$ has m inequalities denoted by $A_1 x \leqslant b_1$, $\ldots$, $A_m x \leqslant b_m$. The *slack matrix of the pair* P, Q w.r.t. these inner and outer descriptions is the $m \times n$ matrix $S^{P,Q}$ with $S^{P,Q}_{ij} = b_i - A_i v_j$ for $i \in [m]$ and $j \in [n]$.

Our first result gives an exact characterization of $\text{xc}(P,Q)$ in terms of the nonnegative rank of the slack matrix of the pair P, Q. It states that the minimum extension complexity of a polyhedron sandwiched between P and Q is exactly $\text{xc}(P,Q)$. The result readily generalizes Yannakakis's factorization theorem [8], which concerns the case $P = Q$. It first appeared in [32].

Theorem 1. *With the above notations, we have* $\text{xc}(P,Q) = \text{rank}_+(S^{P,Q})$ *for every slack matrix of the pair* P, Q*. The minimal value is realized by an EF where* K *is a polytope.*

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:

$$\begin{aligned} S^{P,\rho Q}_{ij} &= \rho b_i - A_i v_j = (\rho - 1) b_i + b_i - A_i v_j \\ &= S^{P,Q}_{ij} + (\rho - 1) b_i. \end{aligned}$$

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 polyhedral and 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 exactly* $\text{rank}_+(S^{P,\rho Q})$*. For a minimization problem, the minimum size of a* ρ*-approximate EF is* $\text{rank}_+(S^{P,\rho^{-1} Q})$.

Fixing $\rho \geqslant 1$, Theorem 2 characterizes the minimum number of inequalities in any LP providing a ρ-approximation for the problem under consideration.

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

Consider the maximum cut problem (Max CUT) with *arbitrary* weights, and its usual linear encoding. With this encoding we have $P = Q = \text{CUT}(n)$. 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 $\mathbf{0}$. These facets define the cut cone, which has high "extension complexity".

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

III. Extending Razborov's Lemma and Perturbing Unique Disjointness

In the first subsection we generalize Razborov's famous rectangle corruption lemma for the disjointness problem (see [15] or [10, Lemma 4.49] for the original version). In the following subsection we apply it to perturb 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 to ease application: optimized estimation and constants, incorporating rank-1 matrices, and nonnegative rank.

A. Extending Razborov's Lemma

For every $0 < p < 1$ and $1 \le \ell \le n/2$ we define the following distribution μ of random subsets a and b of size ℓ of $[n]$. We flip a biased coin and with probability p, we choose (a,b) uniformly among the pairs of subsets intersecting in exactly one element; with probability $1-p$, we choose (a,b) uniformly among the pairs of disjoint subsets.

Lemma 4. *For every $0 < p < 1$, $n \ge 3$, $1 \le \ell \le (n+1)/4$ let μ be the probability distribution above. Furthermore, let $A = \{(a,b) \mid a \cap b = \emptyset\}$ denote the event that a and b are disjoint, and $B = \{(a,b) \mid |a \cap b| = 1\}$ denote the event that a and b intersect in exactly one element. For every sequence of nonnegative functions $f_1, g_1, \ldots, f_r, g_r$ defined on the subsets of $[n]$, we introduce a random variable $X := \sum_{i=1}^r f_i(a)g_i(b)$. Then for every $0 < \varepsilon < 1$*

$$\mathbb{E}[XI_B] \geqslant (1-\varepsilon)\frac{p}{1-p}\mathbb{E}[XI_A] - rp\,\|XI_A\|_\infty 2^{-\frac{\varepsilon^2}{4\ln 2}\ell + O(\log \ell)}, \quad (1)$$

where $O(\log \ell)$ is a function only in ℓ and I_A, I_B are the indicators of the events A and B respectively. Recall that the uniform norm $\|Y\|_\infty$ of a random variable Y is the supremum of the values of $|Y|$.

A strengthened version of the original lemma is recovered by the choice $p = 1/4$, $r = 1$, $\ell = (n+1)/4$ and X the characteristic function of the rectangle $R = C \times D$.

B. Lower Bounds for Perturbations of UDISJ

Now we apply Lemma 4 to show that the nonnegative rank (and hence the complexity of computation in expectation) of any perturbed 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 $M_{ab} = \rho - 1$ whenever $|a \cap b| = 1$ and $M_{ab} = \rho$ whenever $a \cap b = \emptyset$ 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 [13] via nondeterministic communication complexity. We now extend this result for a wide range of ρ using Lemma 4.

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

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

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

Proof: Regarding the $2^n \times 2^n$ matrix M as a random variable over $2^{[n]} \times 2^{[n]}$, we apply Lemma 4 to $X := M$. Suppose that M has a rank-r nonnegative factorization. Therefore we can write X as $X(a,b) = \sum_{i=1}^r f_i(a)g_i(b)$ where f_i and g_i are nonnegative functions defined over $[2^n]$ with $i \in [r]$. Note that $MI_A = \rho I_A$ and $MI_B = (\rho-1)I_B$ and so (1) reduces to $p(\rho-1) \geqslant (1-\varepsilon)\frac{p}{1-p}(1-p) \cdot \rho - rp \cdot \rho \cdot 2^{-\frac{\varepsilon^2}{16\ln 2}n + O(\log n)}$ which gives the lower bound $r \geqslant \left(\frac{1}{\rho} - \varepsilon\right) 2^{\frac{\varepsilon^2}{16\ln 2}n + O(\log n)}$. If ρ is constant, this last expression is $2^{\Omega(n)}$ provided ε is chosen sufficiently close to 0. This proves part (i).

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

IV. Polyhedral Inapproximability of CLIQUE and SDPs

We will now use Theorem 5 and Theorem 2 to lower bound the sizes of certain approximate EFs.

A. 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^T where $b \in \{0,1\}^n$. In other words, $\mathrm{COR}(n) = \operatorname{conv}(\{bb^\intercal \mid b \in \{0,1\}^n\})$. This will be our inner polytope P. Next, let $Q = Q(n) := \{x \in \mathbb{R}^{n \times n} \mid \langle 2\operatorname{diag}(a) - aa^\intercal, x\rangle \leqslant 1,\ a \in \{0,1\}^n\}$, where $\langle \cdot, \cdot \rangle$ denotes the Frobenius inner product. This will be our outer polyhedron Q.

Then the following is known, see [13]. 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 polytope 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 $\mathrm{xc}(P, \rho Q) = 2^{\Omega(n^{1-2\beta})}$.

B. 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 $\mathscr{L} \subseteq \{0,1\}^*$ 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 $\mathscr{O} \subseteq \{-1,0,1\}^*$ 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 incidence vector of a clique of G. Indeed, if $b = \chi^X$ and ij a non-edge 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, $(\mathscr{L}, \mathscr{O})$ 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 $Q^{\text{all}} = \{x \in \mathbb{R}^{n \times n} \mid \forall$ graphs G s.t. $V(G) \subseteq [n]$: $\langle w^G, x \rangle \leqslant \omega(G),\ \forall i \neq j \in [n] : x_{ij} \geqslant 0\}$.

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

C. Polyhedral Inapproximability of SDPs

Recall that a spectrahedron is the projection to a subspace of the intersection of the SDP cone and an affine space. In this section we show that there exists a spectrahedron with small semidefinite extension complexity but high approximate extension complexity. This indicates that in general it is not possible to approximate SDPs arbitrarily well using LPs. (In contrast to SOCPs, see [37].) The result follows from Theorem 6 and [13].

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 $\mathrm{xc}_{SDP}(S)$.

Theorem 8 (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 polytope 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 $\mathrm{xc}(K) = 2^{\Omega(n^{1-2\beta})}$.

V. Conclusion

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. We obtain *unconditional* lower bounds on integrality gaps for small LP relaxations.We have proved that every polynomial-size LP relaxation for (a natural linear encoding of) CLIQUE has approximately an $\Omega(\sqrt{n})$ integrality gap.

Finally, our work sheds more light on the inherent limitations of LPs in the context of approximation algorithms.We provide strong evidence that certain approximation guarantees can only be achieved via non-LP-based techniques. We are convinced that our framework can be used to obtain strong approximation limits for (LP relaxations of) other well-known problems such as Max CUT, Max k-SAT and VERTEX COVER. The following important questions remain open.

(i) Is it possible to show a constant-factor polyhedral inapproximability for Max CUT with nonnegative 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.
(ii) So far no strong lower bounding technique for semidefinite EFs are known. Recent work by [38] provides hope to obtain such lower bounds. In fact the authors introduce a combinatorial lower bounding technique that they apply to relaxations of Max CUT and TSP. Although the details of their approach are not yet available, 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.

References

[1] V. V. Vazirani, *Approximation algorithms*. Berlin: Springer-Verlag, 2001.

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

[3] S. Arora, B. Bollobás, and L. Lovász, “Proving integrality gaps without knowing the linear program,” in *Proc. FOCS*, 2002, pp. 313–322.

[4] 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.*, vol. 3, pp. 411–430, 1990.

[5] L. Lovász and A. Schrijver, “Cones of matrices and set-functions and 0-1 optimization,” *SIAM J. Optim.*, vol. 1, pp. 166–190, 1991.

[6] M. Held and R. Karp, “The traveling salesman problem and minimum spanning trees,” *Oper. Res.*, vol. 18, pp. 1138–1162, 1970.

[7] L. Wolsey, “Heuristic analysis, linear programming and branch and bound,” *Math. Programming Stud.*, vol. 13, pp. 121–134, 1980.

[8] M. Yannakakis, “Expressing combinatorial optimization problems by linear programs,” *J. Comput. System Sci.*, vol. 43, no. 3, pp. 441–466, 1991.

[9] ——, “Expressing combinatorial optimization problems by linear programs (extended abstract),” in *Proc. STOC*, 1988, pp. 223–228.

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

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

[12] A. Chattopadhyay and T. Pitassi, “The story of set disjointness,” *SIGACT News*, vol. 41, pp. 59–85, 2010.

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

[14] B. Kalyanasundaram and G. Schnitger, "The probabilistic communication complexity of set intersection," *SIAM J. Discrete Math.*, vol. 5, pp. 545–557, 1992.

[15] A. A. Razborov, "On the distributional complexity of disjointness," *Theoret. Comput. Sci.*, vol. 106, no. 2, pp. 385–390, 1992.

[16] Z. Bar-Yossef, T. Jayram, R. Kumar, and D. Sivakumar, "An information statistics approach to data stream and communication complexity," *J. Comput. System Sci.*, vol. 68, no. 4, pp. 702–732, 2004.

[17] M. Charikar, K. Makarychev, and Y. Makarychev, "Integrality gaps for Sherali-Adams relaxations," in *Proc. STOC*. ACM, 2009, pp. 283–292.

[18] ——, "Local global tradeoffs in metric embeddings," *SIAM J. Comput.*, vol. 39, no. 6, pp. 2487–2512, 2010.

[19] A. R. Karlin, C. Mathieu, and C. T. Nguyen, "Integrality gaps of linear and semi-definite programming relaxations for knapsack," in *Proc. IPCO*, 2011, pp. 301–314.

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

[21] W. Cook and S. Dash, "On the matrix-cut rank of polyhedra," *Math. Oper. Res.*, vol. 26, pp. 19–30, 2001.

[22] J. B. Lasserre, "An explicit equivalent positive semidefinite program for nonlinear 0-1 programs," *SIAM J. Optim.*, vol. 12, pp. 756–769, 2002.

[23] S. Arora, S. Rao, and U. Vazirani, "Expander flows, geometric embeddings and graph partitioning," *J. ACM*, vol. 56, no. 2, p. 5, 2009.

[24] G. Schoenebeck, "Linear level Lasserre lower bounds for certain k-CSPs," in *Proc. FOCS*. IEEE, 2008, pp. 593–602.

[25] G. Schoenebeck, L. Trevisan, and M. Tulsiani, "A linear round lower bound for Lovász-Schrijver SDP relaxations of vertex cover," in *Proc. CCC*. IEEE, 2007, pp. 205–216.

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

[27] B. Barak, P. Raghavendra, and D. Steurer, "Rounding semidefinite programming hierarchies via global correlation," in *Proc. FOCS*. IEEE, 2011, pp. 472–481.

[28] B. Barak, P. Gopalan, J. Håstad, R. Meka, P. Raghavendra, and D. Steurer, "Making the long code shorter," 2012, manuscript.

[29] B. Barak, F. G. Brandão, A. Harrow, J. Kelner, D. Steurer, and Y. Zhou, "Hypercontractivity, sum-of-squares proofs, and their applications," in *STOC*, 2012, to appear.

[30] D. Bienstock, "Approximate formulations for 0-1 knapsack sets," *Oper. Res. Lett.*, vol. 36, no. 3, pp. 317–320, 2008.

[31] M. Vyve and L. Wolsey, "Approximate extended formulations," *Math. Program.*, vol. 105, no. 2, pp. 501–522, 2006.

[32] K. Pashkovich, "Extended formulations for combinatorial polytopes," Ph.D. dissertation, Magdeburg Universität, 2012.

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

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

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

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

[37] A. Ben-Tal and A. Nemirovski, "On polyhedral approximations of the second-order cone," *Math. Oper. Res.*, vol. 26, pp. 193–205, 2001.

[38] T. Lee and D. Theis, "Lower bounds for sizes of semidefinite formulations for some combinatorial optimization problems," 2012, arXiv:1203.3961.