Some lower bounds on sparse outer approximations of polytopes

Santanu S. Dey Andres Iroume Marco Molinaro

December 11, 2014

Abstract

Motivated by the need to better understand the properties of sparse cutting-planes used in mixed integer programming solvers, the paper [2] studied the idealized problem of how well a polytope is approximated by the use of sparse valid inequalities. As an extension to this work, we study the following "less idealized" questions in this paper: (1) Are there integer programs, such that sparse inequalities do not approximate the integer hull well even when added to a linear programming relaxation? (2) Are there polytopes, where the quality of approximation by sparse inequalities cannot be significantly improved by adding a budgeted number of arbitrary (possibly dense) valid inequalities? (3) Are there polytopes that are difficult to approximate under every rotation? (4) Are there polytopes that are difficult to approximate in all directions using sparse inequalities? We answer each of the above questions in the positive.

1 Introduction

The paper [2] studied how well one can expect to approximate polytopes using valid inequalities that are sparse. The motivation for this study came from the usage of cutting-planes in integer programming (IP) solvers. In principle, facet-defining inequalities of the integer hull of a polytope can be dense, i.e. they can have non-zero coefficients for a high number of variables. In practice, however, most state-of-the-art IP solvers bias their cutting-plane selection towards the use of sparse inequalities. This is done, in part, to take advantage of the fact that linear programming solvers can harness sparsity well to obtain significant speedups.

The paper [2] shows that for polytopes with a polynomial number of vertices, sparse inequalities produce very good approximations of polytopes. However, when the number of vertices increase, the sparse inequalities do not provide a good approximation in general; in fact with high probability the quality of approximation is poor for random 0-1 polytopes with super polynomial number of vertices (see details in [2]).

However the study in [2] is very "idealized" in the context of cutting-planes for IPs, since almost always some dense cutting-planes are used or one is interested in approximating the integer only only along certain directions. In this paper, we consider some natural extensions to understand the properties of sparse inequalities under more "realistic conditions":

1. All the results in the paper [2] deal with the case when we are attempting to approximate the integer hull using only sparse inequalities. However, in practice the LP relaxation may have dense inequalities. Therefore we examine the following question: Are there integer programs, such that sparse inequalities do not approximate the integer hull well when added to a linear programming relaxation?

2. More generally, we may consider attempting to improve the approximation of a polytope by adding a few dense inequalities together with sparse inequalities. Therefore we examine the following question: Are there polytopes, where the quality of approximation by sparse inequalities cannot be significantly improved by adding polynomial (or even exponential) number of *arbitrary* valid inequalities?

3. It is clear that the approximations of polytopes using sparse inequalities is not invariant under affine transformations (in particular rotations). This leaves open the possibility that a clever reformulation of the polytope of interest may vastly improve the approximation obtained by sparse cuts. Therefore a basic question in this direction: Are there polytopes that are difficult to approximate under *every* rotation?

4. In optimization one is usually concerned with the feasible region in the direction of the objective function. Therefore we examine the following question: Are there polytopes that are difficult to approximate in *almost all* directions using sparse inequalities?

We are able to present examples that answer each of the above questions in the positive. This is perhaps not surprising: an indication that sparse inequalities do not always approximate integer hulls well even in the more realistic settings considered in this paper. Understanding when sparse inequalities are effective in all the above settings is an important research direction.

The rest of the paper is organized as follows. Section 2 collects all required preliminary definitions. In Section 3 we formally present all the results. In Sections 4-7 we present proofs of the various results.

2 Preliminaries

2.1 Definitions

For a natural number n, let $[n]$ denote the set $\{1, \ldots, n\}$ and, for non-negative integer $k \leq n$ let $\binom{[n]}{k}$ denote the set of all subsets of $[n]$ with k elements. For any $x \in \mathbb{R}^n$, let $||x||_1$ denote the l_1 norm of x and $||x||$ or $||x||_2$ denote the l_2 norm of x.

An inequality $\alpha x \leq \beta$ is called *k-sparse* if α has at most k non-zero components. Given a polytope $P \subset \mathbb{R}^n$, P^k is defined as the intersection of all k-sparse cuts valid for P (as in [2]), that is, the best outer-approximation obtained from k-sparse inequalities.

Given two polytopes $P, Q \subset \mathbb{R}^n$ such that $P \subseteq Q$ we consider the Hausdorff distance $d(P, Q)$ between them:

$$d(P, Q) := \max_{x \in Q} \left(\min_{y \in P} ||x - y|| \right).$$

When $P, Q \subset [-1, 1]^n$, we have that $d(P, Q)$ is upper bounded by $2\sqrt{n}$, the largest distance between two points in $[-1, 1]^n$. In this case, if $d(P, Q) \propto \sqrt{n}$ the error of approximation of P by Q is basically as large as it can be and smaller $d(P, Q)$ (for example constant or of the order of $\sqrt{\log n}$) will indicate better approximations.

Given a polytope $P \subseteq \mathbb{R}^n$ and a vector $c \in \mathbb{R}^n$, we define

$$gap_P^k(c) = \max_{x \in P^k} cx - \max_{x \in P} cx,$$

namely the "gap" between P^k and P in direction c. We first note that $d(P, P^k)$ equals the worst directional gap between P^k and P (the proof is presented in Appendix A).

Lemma 1. For every polytope $P \subseteq \mathbb{R}^n$, $d(P, P^k) = \max_{c: ||c||=1} gap_P^k(c)$.

For a set $\mathcal{D} = \{\alpha_1 x \leq \beta_1, \ldots, \alpha_d x \leq \beta_d\}$ of (possibly dense) valid inequalities for P, let $P^{k,\mathcal{D}}$ denote the outer-approximation obtained by adding all k-sparse cuts and the inequalities from $\mathcal{D}$:

$$P^{k,\mathcal{D}} = \left(\bigcap_{i=1}^{d} \{x \in \mathbb{R}^n : a_i x \leq b_i\} \right) \bigcap P^k. \tag{1}$$

Since $P^{k,\mathcal{D}} \subseteq P^k$ we have that $d(P, P^{k,\mathcal{D}}) \leq d(P, P^k)$ for any set $\mathcal{D}$ of valid inequalities for P.

2.2 Important Polytopes

Throughout the paper, we will focus our attention on the polytopes $\mathcal{P}_{t,n} \subseteq [0,1]^n$ defined as

$$\mathcal{P}_{t,n} = \left\{ x \in [0,1]^n : \sum_{i=1}^{n} x_i \leq t \right\}. \tag{2}$$

Notice that for $t = 1$ we obtain a simplex and for $t = n/2$ we obtain half of the hypercube. Moreover different values to t yield very different properties regarding approximability using sparse inequalities, as discussed in [2].

Proposition 2. The following hold:

1. $d(\mathcal{P}_{1,n}, \mathcal{P}^k_{1,n}) = \frac{\sqrt{n}}{k} - \frac{1}{\sqrt{n}}$.
2. $d(\mathcal{P}_{n/2,n}, \mathcal{P}^k_{n/2,n}) = \begin{cases} \sqrt{n}/2 & \text{if } k \leq n/2 \\ \frac{n\sqrt{n}}{2k} - \frac{\sqrt{n}}{2} & \text{if } k > n/2 \end{cases}$.

We will also consider symmetrized versions of the polytopes $\mathcal{P}_{t,n}$. To define this symmetrization, for $x \in \mathbb{R}^n$ and $I \subset [n]$ let x^I denote the vector obtained by switching the sign of the components of x not in I:

$$x^I_i = \begin{cases} x_i & \text{if } i \in I \\ -x_i & \text{if } i \notin I. \end{cases}$$

More generally, for a set $P \subseteq \mathbb{R}^n$ we define $P^I = \left\{ x^I \in \mathbb{R}^n : x \in P \right\}$.

Definition 3. For a polytope $P \subseteq \mathbb{R}^n_+$, we define its symmetrized version $\overline{P} = \operatorname{conv}\left(\bigcup_{I \subseteq [n]} P^I \right)$.

Note that $\overline{\mathcal{P}_{1,n}}$ is the cross polytope in dimension n; more generally, we have the following external description of the symmetrized versions of $\mathcal{P}_{t,n}$ and $\mathcal{P}^k_{t,n}$ (proof presented in Appendix B).

Lemma 4.

$$\overline{\mathcal{P}_{t,n}} = \left\{ x \in [-1,1]^n : \ \forall I \subset [n], \ \sum_{i \in I} x_i - \sum_{i \in [n] \setminus I} x_i \leq t \right\} \tag{3}$$

$$\overline{\mathcal{P}_{t,n}}^k = \left\{ x \in [-1,1]^n : \ \forall I \in \binom{[n]}{k}, \ \forall I^+, I^- \text{ partition of } I, \right.$$
$$\left. \sum_{i \in I^+} x_i - \sum_{i \in I^-} x_i \leq t \right\}. \tag{4}$$

3 Main results

In our first result (Section 4), we point out that in the worst case LP relaxations plus sparse inequalities provide a very weak approximation of the integer hull.

Theorem 5. For every even integer n there is a polytope $Q_n \subseteq [0,1]^n$ such that:

1. $\mathcal{P}_{n/2,n} = \operatorname{conv}(Q_n \cap \mathbb{Z}^n)$
2. $d(\mathcal{P}_{n/2,n}, (\mathcal{P}_{n/2,n})^k \cap Q_n) = \Omega(\sqrt{n})$ for all $k \leq n/2$.

In Section 5 we consider the second question: How well does the approximation improve if we allowed a budgeted number of dense valid inequalities. Notice that for the polytope $\mathcal{P}_{\frac{n}{2},n}$, while Proposition 2 gives that $d(\mathcal{P}_{\frac{n}{2},n}, \mathcal{P}^k_{\frac{n}{2},n}) \geq \Omega(\sqrt{n})$, adding exactly *one* dense cut ($ex \leq n/2$) to the k-sparse closure (even for $k = 1$) would yield the original polytope $\mathcal{P}_{\frac{n}{2},n}$.

We consider instead the symmetrized polytope $\overline{\mathcal{P}_{\frac{n}{2},n}}$. Notice that while this polytope needs 2^n dense inequality to be described *exactly*, it could be that a small number of dense inequalities, together with sparse cuts, is already enough to provide a good approximation; we observe that in higher dimensions valid cuts for $\overline{\mathcal{P}_{\frac{n}{2},n}}$ can actually cut off significant portions of $[-1,1]^n$ in *multiple orthants*. We show, however, that in this even *exponentially* many dense inequalities do not improve the approximation significantly.

Theorem 6. Consider an even integer n and the polytope $P = \overline{\mathcal{P}_{\frac{n}{2},n}}$. For any $k \leq n/100$ and any set $\mathcal{D}$ of valid inequalities for P with $|\mathcal{D}| \leq \exp\left(\frac{n}{600^2}\right)$, we have

$$d\left(P, P^{k,\mathcal{D}}\right) \geq \frac{1}{6}\sqrt{n}.$$

In the proof of this theorem we use a probabilistic approach to count in how many orthants an inequality can significantly cut off the box $[-1,1]^n$.

In Section 6 we consider the question of sparse approximation of a polytope when rotations are allowed. We show that again $\overline{\mathcal{P}_{n/2,n}}$ cannot be approximated using sparse inequalities after *any* rotation is applied to it.

Theorem 7. Consider an even integer n and the polytope $P = \overline{\mathcal{P}_{\frac{n}{2},n}}$. For every rotation $R : \mathbb{R}^n \to \mathbb{R}^n$ and $k \leq \frac{n}{200^3}$, we have

$$d\big(R(P), (R(P))^k\big) = \Omega(\sqrt{n}).$$

The proof of this theorem relies on the intuition given by Theorem 6: since $\overline{\mathcal{P}_{\frac{n}{2},n}}$ required exponentially many dense inequalities in order to be well approximated, no rotation is able to align all of them with the axis so that they can be captured by sparse inequalities.

Finally, in Section 7 we show that $\overline{\mathcal{P}_{\frac{n}{10},n}}$ and its k-sparse approximation have a large gap in almost every direction.

Theorem 8. Let $n \geq 1000$ be an integer divisible by 10 and consider the polytope $P = \overline{\mathcal{P}_{n/10,n}}$. If $C \in \mathbb{R}^n$ is a random direction uniformly distributed on the unit sphere, then for $k \leq \frac{n}{10}$ we have

$$\Pr\left(gap^k_P(C) \geq \frac{\sqrt{n}}{20}\right) \geq 1 - \frac{4}{n}.$$

To prove this theorem we rely on the concentration of the value of Lipschitz functions on the sphere (actually we work on the simpler Gaussian space).

4 Strengthening of LP relaxation by sparse inequalities

We now present a short proof of Theorem 5. Consider the polytope

$$Q_n = \left\{ x \in [0,1]^n : \sum_{i \in I} x_i \leq \frac{n}{2} \quad \forall I \in \binom{[n]}{\frac{n}{2}+1} \right\}.$$

It is straightforward to verify that $\mathcal{P}_{n/2,n} = \mathrm{conv}(Q_n \cap \mathbb{Z}^n)$.

From Part (2) of Proposition 2, $\mathcal{P}^k_{n/2,n} = [0,1]^n$ thus $Q_n \cap \mathcal{P}^k_{n/2,n} = Q_n$. Now $x = \frac{n}{n+2}e$ belongs to Q_n and its projection onto $\mathcal{P}_{n/2,n}$ corresponds to $y = \frac{1}{2}e$. Therefore,

$$d\left(\mathcal{P}_{n/2,n}, \mathcal{P}^k_{n/2,n} \cap Q_n\right) = \frac{n-2}{2n+4}\sqrt{n} = \Omega(\sqrt{n}).$$

This concludes the proof of the theorem.

5 Strengthening by general dense cuts

Now we turn to the proof of Theorem 6. For that we will need Bernstein's concentration inequality (stated in a slightly weaker but more convenient form).

Theorem 9 ([3], Appendix A.2)**.** Let $X_1, X_2, \ldots, X_n$ be independent random variables such that $\mathbb{E}[X_i] = 0$ and $|X_i| \leq M \ \forall i$. Let $X = \sum_{i=1}^n X_i$ and $\sigma^2 = \mathrm{Var}(X) \leq U$. Then:

$$\Pr(|X| > w) \leq \exp\left(-\min\left\{\frac{w^2}{4U}, \frac{3w}{4M}\right\}\right).$$

Notice that to prove the theorem it suffices to consider the case $k = \frac{n}{100}$, which is what we do. Recall that $P = \mathcal{P}_{n/2,n}$, consider any set $\mathcal{D}$ of valid inequalities for P with $\|\mathcal{D}| \leq \exp(\frac{n}{600^2})$; for convenience let $d = |\mathcal{D}|$. From Lemma 4 we know P^k contains all the points in $\{-1,1\}^n$. Also note that for any $\bar{x} \in \{-1,1\}^n$ achieves the maximal distance in P^k from P, namely $d(P, P^k) = d(P, \bar{x}) = \frac{1}{2}\sqrt{n}$. We then consider a random such "bad" point X, namely X is uniformly distributed in $\{-1,1\}^n$ (equivalently, the X_i's are independent and uniformly distributed over $\{-1,1\}$). We will show that there exist an instantiation of the scaled random $\frac{2X}{3}$ which belongs to $P^{k,\mathcal{D}}$, which will then lower bound the distance $d(P, P^{k,\mathcal{D}})$ by $d(P, \frac{2\bar{x}}{3}) = \frac{1}{6}\sqrt{n}$ (for some $\bar{x} \in \{-1,1\}^n$) and thus prove the result.

To achieve this, consider a single inequality $ax \leq b$ from $\mathcal{D}$ (we assume without loss of generality that $\|a\|_1 = 1$). We claim that with probability more than $1 - \frac{1}{d}$, the point $\frac{2X}{3}$ satisfies this inequality. By symmetry of X, we can assume without loss of generality that $a \geq 0$. To prove this, let $\bar{a}$ be the vector obtained by keeping the k largest components of a and zeroing out the other components (ties are broken arbitrarily), and let $\underline{a} = a - \bar{a}$. Since $\bar{a}x \leq b$ is a k-sparse valid inequality for P and $X \in P^k$, we have that

$$aX = \bar{a}X + \underline{a}X \leq b + \underline{a}X. \tag{5}$$

Claim 10. $\mathrm{Var}(\underline{a}X) \leq \frac{b(n-k)}{k^2}$.

Proof. Since $\mathrm{Var}(X_i) = 1$ for all $i \in [n]$, we obtain that

$$\mathrm{Var}(\underline{a}X) \quad = \textstyle\sum_{i=1}^n \underline{a}_i^2 \,\mathrm{Var}(X_i) = ||\underline{a}||^2. \tag{6}$$

Note that the kth largest component of a is at most $1/k$ (otherwise $\|a\|_1 > 1$), hence $\underline{a}_i X_i \leq \frac{1}{k}$ for all i, so we have

$$||\underline{a}||^2 = \sum_{i=1}^n (\underline{a}_i X_i)^2 \leq \frac{1}{k}\sum_{i=1}^n \underline{a}_i X_i. \tag{7}$$

Moreover, by comparing averages of the components of $\bar{a}$ and $\underline{a}$ and then using $\bar{a}e \leq b$, we have that

$$\sum_{i=1}^n \frac{\underline{a}_i}{n-k} \leq \sum_{i=1}^n \frac{\bar{a}_i}{k} \leq \frac{b}{k}. \tag{8}$$

Now by using (6)-(8), we obtain the bound $\mathrm{Var}(\underline{a}X) \leq \frac{b(n-k)}{k^2}$, thus concluding the proof. ◇

Now using the fact that $|\underline{a}_i X_i| \leq \frac{1}{k}$, $\mathbb{E}(\underline{a}X) = 0$ and the above bound on $\mathrm{Var}(\underline{a}X)$, we obtain by an application of Bernstein's inequality (Theorem 9) with $w = 30b\frac{\sqrt{\log d}}{\sqrt{k}}$:

$$\Pr\left(\underline{a}X \geq 30b \cdot \frac{\sqrt{\log d}}{\sqrt{k}}\right) \leq \exp\left(-\min\left\{\frac{30^2 b \cdot k \cdot \log d}{4(n-k)}, \frac{30}{4} \cdot 3b \cdot \sqrt{k \log d}\right\}\right). \tag{9}$$

To upper bound the right-hand side of this expression, first we employ our assumption $d \leq \exp(\frac{n}{600^2})$ and $k = \frac{n}{100}$ to obtain

$$\sqrt{\log d} \leq \frac{1}{600}\sqrt{n} \leq \frac{3 \cdot 99}{30 \cdot 10}\sqrt{n} = \frac{3}{30}\left(\frac{n-k}{k}\right)\sqrt{k}.$$

With this at hand, we have that the minimum in the right-hand side of (9) is achieved in the first term. Moreover, notice that $b \geq 1/2$: the point $p = (\frac{1}{2}, \ldots, \frac{1}{2})$ belongs to P and hence $b \geq ap = \frac{1}{2}\|a\|_1 = 1/2$. Putting these observations together gives

$$\Pr\left(\underline{a}X \geq 30b \cdot \frac{\sqrt{\log d}}{\sqrt{k}}\right) \leq \exp\left(-\frac{30^2}{4 \cdot 99} b \cdot \log d\right) < \exp(-\log d) = \frac{1}{d}.$$

Then using (5) and the above inequality, we obtain that with probability more than $1 - \frac{1}{d}$ we have

$$\begin{aligned} aX &\leq b\left(1 + 30\frac{\sqrt{\log d}}{\sqrt{k}}\right) \\ &= b\left(1 + \frac{1}{2} \cdot 600\frac{\sqrt{\log d}}{\sqrt{n}}\right) \leq b\frac{3}{2}, \end{aligned} \tag{10}$$

where the first equality uses $k = \frac{n}{100}$ and the second inequality uses the assumption that $\sqrt{\log d} \leq \frac{1}{600}\sqrt{n}$. Now note that (10) implies that the point $\frac{2X}{3}$ satisfies $ax \leq b$ with probability more than $1 - \frac{1}{d}$.

Since $|\mathcal{D}| = d$, we can then take a union bound over the above argument to get that with strictly positive probability $\frac{2X}{3}$ satisfies *all* the inequalities in $\mathcal{D}$. Hence with strictly positive probability $\frac{2X}{3}$ belongs to $P^{k,\mathcal{D}}$ and in particular there is a point $\overline{x} \in \{-1, 1\}^n$ such that $\frac{2\overline{x}}{3} \in P^{k,\mathcal{D}}$.

This gives the lower bound $d(P, P^{k,\mathcal{D}}) \geq d\left(P, \frac{2\overline{x}}{3}\right)$; now we lower bound the right-hand side. It is easy to see that the closest point in $\overline{P}$ to $2\overline{x}/3$ is $\overline{x}/2$, the projection onto $\overline{P}$. Since $||2\overline{x}/3 - \overline{x}/2|| = \frac{1}{6}||\overline{x}||$, we obtain that $d(\overline{P}, \overline{x}) \geq \frac{1}{6}\sqrt{n}$ which concludes the proof.

6 Sparse approximation of rotations of a polytope

In this section we prove Theorem 7; for that we need to recall some standard definitions from convex geometry.

Definition 11. Given a set $P \subseteq \mathbb{R}^n$:

- We say that P is *centrally symmetric* if $\forall x \in P : -x \in P$.
- For any $\alpha \in \mathbb{R}$ we define the set $\alpha P := \{\alpha x : x \in P\}$.
- The *polar* of P is the set $P^\circ = \{z \in \mathbb{R}^n : zx \leq 1 \ \forall x \in P\}$.

We also need the following classical result about approximating convex set by polytopes with few vertices (see for instance Lemma 4.10 of [1] and [6])

Theorem 12. For every centrally symmetric convex set $S \subseteq \mathbb{R}^k$, there is a polytope S' with at most $(\frac{3}{\epsilon})^k$ vertices such that $S \subseteq S' \subseteq (1+\epsilon)S$

By applying this result to the polar we obtain approximations with bounded number of *facets* instead of vertices.

Lemma 13. For every centrally symmetric conver set $C \subseteq \mathbb{R}^k$, there is a polytope C' with at most $(\frac{3}{\epsilon})^k$ facets such that $C \subseteq C' \subseteq (1+\epsilon)C$.

Proof. Consider the (centrally symmetric) convex set $\frac{1}{1+\epsilon}C^\circ$; applying the above result, we get S with $(3/\epsilon)^k$ vertices and $\frac{1}{1+\epsilon}C^\circ \subseteq S \subseteq C^\circ$. Taking polars (and noticing that $(\lambda A)^\circ = (1/\lambda)A^\circ$), we get $C \subseteq S^\circ \subseteq (1+\epsilon)C$ and S° has at most $(3/\epsilon)^k$ facets. This concludes the proof. □

The key ideas used in our proof of Theorem 6 is twofold (recall that $P = \overline{\mathcal{P}_{n/2,n}}$):

1. Roughly speaking, $(RP)^k$ is the intersection of (rotations of) *k-dimensional* polytopes. This allow us to use Lemma 13 above (with n set to k) to get a good approximation H of $(RP)^k$ using fewer than $\exp(\frac{n}{600^2})$ inequalities.
2. Then argue that $d(RP,(RP)^k) \approx d(RP,H) = \Omega(\sqrt{n})$ since $d(P,R^{-1}(H)) = \Omega(\sqrt{n})$ due to the number of facets of H and Theorem 6.

Proof of Theorem 7. Note that it is sufficient to prove the result for $k = \frac{n}{200^3}$, which is what we do. To make the above ideas precise, observe that $(RP)^k = \bigcap_{K\in\binom{[n]}{k}} Q_K$, where $Q_K = RP + 0^K \times \mathbb{R}^{\bar{K}}$ (we use $\bar{K} := [n]\setminus K$). To approximate each Q_K, using Lemma 13, let $h_K \subseteq \mathbb{R}^k$ be a polytope such that $\mathrm{proj}_K Q_K \subseteq h_K \subseteq (1+\epsilon)\mathrm{proj}_K Q_K$ and h_K has at most $(3/\epsilon)^k$ facets. Let $H_k = h_K + 0^K \times \mathbb{R}^{\bar{K}}$; then $Q_K \subseteq H_K \subseteq (1+\epsilon)Q_K$ and H_K has at most $(3/\epsilon)^k$ facets.

Now notice that for convex sets A, B, we have $((1+\epsilon)A) \cap ((1+\epsilon)B) \subseteq (1+\epsilon)(A\cap B)$. This gives that if we look at the intersection $\bigcap_{K\in\binom{[n]}{k}} H_K$, we obtain

$$\begin{aligned}(RP)^k &= \bigcap_{K\in\binom{[n]}{k}} Q_K \subseteq \bigcap_{K\in\binom{[n]}{k}} H_K \subseteq (1+\epsilon)\bigcap_{K\in\binom{[n]}{k}} Q_K \\ &= (1+\epsilon)(RP)^k. \end{aligned} \tag{11}$$

Notice $\bigcap_{K\in\binom{[n]}{k}} H_K$ has at most $\binom{n}{k}\left(\frac{3}{\epsilon}\right)^k \le \left(\frac{en}{k}\right)^k\left(\frac{3}{\epsilon}\right)^k = \left(\frac{3en}{k\epsilon}\right)^k$ facets. Thus, setting $\epsilon = \frac{1}{10}$ we get

$$\begin{aligned}\left(\frac{3en}{k\epsilon^*}\right)^k &= \left(30\cdot e\cdot 200^3\right)^{\frac{n}{200^3}} \\ &= \left(\exp(\log(30\cdot e\cdot 200^3))\right)^{\frac{n}{200^3}} \\ &= \left(\exp\left(\log(30\cdot e\cdot 200^3)\cdot\frac{n}{200^3}\right)\right) \\ &< \exp\left(\frac{n}{601^2}\right). \end{aligned} \tag{12}$$

Then define $H := \bigcap_{K\in\binom{[n]}{k}} H_K$, so that $(RP)^k \subseteq H \subseteq (1+\epsilon)(RP)^k$.

In order to control the relationship between this multiplicative approximation and the distance $d(.,.)$, we introduce the set $C = R([-1,1]^n)$. Notice that by construction $RP \subseteq H \cap C$.

Claim 14. $d(RP, H\cap C) \ge \frac{1}{6}\sqrt{n}$

Proof. Assume by contradiction that $d(RP,H\cap C) < \frac{1}{6}\sqrt{n}$. Then since distances between points and number of facets of a polytope are invariant under rotation, we obtain that $d(P, R^{-1}(H\cap C)) < \frac{1}{6}\sqrt{n}$ where $R^{-1}(H\cap C)$ is defined using at most $\exp(n/(600)^2)$ inequalities (because C has $2n$ facets, using (12) H has at most $\exp(n/(601)^2)$ and for sufficiently large n, $\exp(n/(601)^2) + 2n \le \exp(n/(600)^2)$). However notice that this contradicts the result of Theorem 6, since $k = \frac{\sqrt{n}}{100} \le \frac{n}{100}$ and $R^{-1}(H\cap C)$ is defined using at most $2^{n/(600)^2}$ inequalities. ◇

But from (11) we have $(1+\epsilon)(RP)^k \cap C$ contains $H\cap C$, and hence

$$d(RP, (1+\epsilon)(RP)^k \cap C) \ge \frac{1}{6}\sqrt{n}. \tag{13}$$

Claim 15. $d(RP, (RP)^k \cap C) \geq d(RP, (1+\epsilon)(RP)^k \cap C) - \epsilon\sqrt{n}$

Proof. Take $\bar{x} \in (1+\epsilon)(RP)^k \cap C$ and $\bar{y} \in RP$ that achieve $d(\bar{x}, \bar{y}) = d((1+\epsilon)(RP)^k \cap C, RP)$. Look at the point $\frac{1}{1+\epsilon}\bar{x}$ and notice it belongs to $(RP)^k \cap C$; let $\tilde{y}$ be the point in RP closest to $\frac{1}{1+\epsilon}\bar{x}$. Then since $\bar{y}$ is the point in RP closest to $\bar{x}$,

$$d(RP, (1+\epsilon)(RP)^k \cap C) = d(\bar{x}, \bar{y}) \leq d(\bar{x}, \tilde{y}).$$

By triangle inequality, $d(\bar{x}, \tilde{y}) \leq d(\frac{1}{1+\epsilon}\bar{x}, \tilde{y}) + d(\frac{1}{1+\epsilon}\bar{x}, \bar{x}) \leq d(RP, (RP)^k \cap C) + d(\frac{1}{1+\epsilon}\bar{x}, \bar{x})$. To bound $d(\frac{1}{1+\epsilon}\bar{x}, \bar{x})$, notice it is equal to $\frac{\epsilon}{1+\epsilon}\|\bar{x}\|$; since $\bar{x}$ belongs to C, we can upper bound $\|\bar{x}\| \leq \sqrt{n}$ (this is why we introduced the set C in the argument). Putting these bounds together we obtain the result. ◇

Using (13) and Claim 2 we obtain that $d(RP, (RP)^k) \geq d(RP, (RP)^k \cap C) \geq d(RP, (1+\epsilon)(RP)^k \cap C) - \epsilon\sqrt{n} \geq (\frac{1}{6} - \frac{1}{10})\sqrt{n}$. This concludes the proof of the theorem. □

7 Lower bounds on approximation along most directions

We now prove Theorem 8. The main tool we use in this section is concentration of Lipschitz functions on Gaussian spaces.

Theorem 16 (Inequality (1.6) of [4])**.** Let $G_1, G_2, \ldots, G_n$ be independent standard Gaussian random variables, and let $f : \mathbb{R}^n \to \mathbb{R}$ be an L-Lipschitz function, namely for all $x, x' \in \mathbb{R}^n$, $|f(x) - f(x')| \leq L \cdot \|x - x'\|$. Then letting $Z = f(G_1, G_2, \ldots, G_n)$, for $t > 0$ we have

$$\Pr\left(|Z - \mathbb{E}(Z)| \geq t\right) \leq 2\exp\left(-\frac{t^2}{2L^2}\right)$$

To prove Theorem 8, recall that $P = \overline{\mathcal{P}_{n/10,n}}$. Let $G = (G_1, G_2, \ldots, G_n)$ be a random vector whose components are independent standard Gaussians. It is well-known that $\frac{G}{\|G\|_2}$ is uniformly distributed in the sphere (see for instance [4], page 55). Notice that $gap_P^k(\cdot)$ is positive homogeneous, so $gap_P^k\left(\frac{G}{\|G\|}\right) = \frac{1}{\|G\|} \cdot gap_P^k(G)$.

Our first step is to lower bound $gap_P^k(G)$ with high probability, starting by lower bounding the maximization of G over P^k.

Claim 17. With probability at least $1 - \frac{1}{n}$, $\max_{x \in P^k} Gx \geq 0.7n$.

Proof. Since $k = \frac{n}{10}$, we have that $P^k = [-1, 1]^n$ (Proposition 4). It then follows that

$$\max_{x \in P^k} Gx = \sum_{i=1}^{n} |G_i|. \tag{14}$$

The random variables $|G_i|$ have *folded normal* distribution [5], for which is known that $\mathbb{E}[|G_i|] = \sqrt{2/\pi} \geq 0.79$. Since the function $(x_1, \ldots, x_n) \mapsto \sum_{i=1}^n |x_i|$ is $\sqrt{n}$-Lipschitz, we can use Theorem 16 to obtain the bound

$$\Pr\left(\sum_{i=1}^{n} |G_i| < 0.7n\right) \leq 2\exp\left(-\frac{0.09^2 n}{2}\right) \leq \frac{1}{n},$$

where the last inequality holds if $n \geq 1000$. Equation (14) then concludes the proof. ◇

Next we upper bound the maximization of G over P.

Claim 18. With probability at least $1 - \frac{2}{n}$, $\max_{x \in P} Gx \leq 0.6n$.

Proof. Letting $ext(P)$ denote the set of extreme points of P, notice that $\max_{x\in P} Gx = \max_{v\in ext(P)} Gv$, so it suffices to upper bound the latter. Also notice that the extreme points of P are exactly the points in $\{-1,0,1\}^n$ with at most $\frac{n}{10}$ non-zero entries (Proposition 4).

Consider $v \in ext(P)$; we verify that $Gv \leq 0.6n$ with probability at least $1-2e^{-0.6n}$. One way of seeing this, is by noticing that since v has at most $\frac{n}{10}$ non-zero entries, $Gv = \sum_{i:v_i=1} G_i + \sum_{i:v_i=-1} -G_i$ is a function of G that has at most $\frac{n}{10}$ terms and is $\sqrt{\frac{n}{10}}$-Lipschitz, so Theorem 16 gives

$$\Pr\left(Gv > 0.6n\right) = \Pr\left(Gv - \mathbb{E}[Gv] > 0.6n\right) \leq 2e^{-0.6n}, \tag{15}$$

and the result follows. (Another way to see this is to use that fact that Gv is a centered Gaussian with variance at most $\frac{n}{10}$ and use a tail bound for the latter.)

Now notice that P has $\sum_{i=1}^{n/10}\binom{n}{i}2^i \leq \frac{n}{10}\binom{n}{n/10}2^{n/10}$ extreme points. Since $\binom{n}{t} \leq (\frac{en}{t})^t$ for all $0<t<n$, the number of extreme points of P can be upper bounded by

$$\exp\left(\ln\left(\frac{n}{10}\right) + \frac{n}{10}(\ln 10e + \ln 2)\right) \leq \frac{2}{n}e^{0.6n},$$

where the last inequality uses $n \geq 30$.

Then taking a union bound of (15) over all extreme points of P gives that with probability at least $1-\frac{2}{n}$ for all $v \in ext(P)$ we have $Gv \leq 0.6n$. This concludes the proof. ◇

Finally, standard results give that $\|G\|_2 \leq 2\sqrt{n}$ with probability at least $1-2e^{-0.5n}$ (for instance, notice by Jensen's inequality $\mathbb{E}[\|G\|]^2 \leq \mathbb{E}[\|G\|^2] = n$ and apply Theorem 16 to $\|G\|$). Using the fact $n \geq 30$, we then get $\Pr(\|G\| \leq 2\sqrt{n}) \geq 1-\frac{1}{n}$. Then taking a union bound over this event and the events $\max_{x\in P^k} Gx \geq 0.7n$ and $\max_{x\in P} Gx \leq 0.6n$ gives that with probability at least $1-\frac{4}{n}$ we have $gap_P^k\left(\frac{G}{\|G\|}\right) = \frac{1}{\|G\|}\cdot gap_P^k(G) \geq \frac{\sqrt{n}}{20}$. This concludes the proof of Theorem 8.

Acknowledgments

We thank Jon Lee for motivating some of the questions considered in this paper. Santanu S. Dey gratefully acknowledges the support by NSF under grant CCF-1415460.

References

[1] Alexander Barvinok. Thrifty approximations of convex bodies by polytopes. *International Mathematics Research Notices*, page rnt078, 2013.

[2] Santanu S Dey, Marco Molinaro, and Qianyi Wang. How good are sparse cutting-planes? In *Integer Programming and Combinatorial Optimization*, pages 261–272. Springer, 2014.

[3] Vladimir Koltchinskii. *Oracle Inequalities in Empirical Risk Minimization and Sparse Recovery Problems: Ecole dEté de Probabilités de Saint-Flour XXXVIII-2008*, volume 2033. Springer, 2011.

[4] M. Ledoux and M. Talagrand. *Probability in Banach Spaces: Isoperimetry and Processes.* Springer, New York, 1991.

[5] F. C. Leone, L. S. Nelson, and R. B. Nottingham. The folded normal distribution. *Technometrics*, 3(4):pp. 543–550, 1961.

[6] Gilles Pisier. *The volume of convex bodies and Banach space geometry*, volume 94. Cambridge University Press, 1999.

Appendix A

Proof of Lemma 1. It is not difficult to see that for $P^k = P$, the lemma holds, since $d(P, P^k) = gap_P^k(c) = 0\ \forall c : ||c|| = 1$. When, $P^k \neq P$, we have that $d(P, P^k) = d(x^0, y^0) > 0$ is attained at $x^0 \in ext(P^k)$ and $y^0 \in P$, the orthogonal projection of x^0 onto P (see [2]). Thus, $y^0 \in F = \{z \in \mathbb{R}^n : az = b\} \cap P$, a face of P such that $a = (x^0 - y^0)$, $b = (x^0 - y^0)y^0$ and $P \subseteq \{z \in \mathbb{R}^n : az \leq b\}$. Let $c = (x^0 - y^0)/||x^0 - y^0||$, we have: $\max_{x \in P} cx = cy^0$. On the other hand, $\max_{z \in P^k} cz = cx^0$, since otherwise, if $\exists \bar{x} \in P^k$ with $c\bar{x} > cx^0$, let $\bar{y}$ denote the orthogonal projection of $\bar{x}$ onto $\{z \in \mathbb{R}^n : az = b\}$. Then, for all $z \in P$ we have $d(\bar{x}, z) \geq d(\bar{x}, \bar{y}) > d(x^0, y^0)$ (the last inequality follows from the fact that $c\bar{x} > cx^0$, $c\bar{y} = cy^0$ and $\bar{x} - \bar{y}/||\bar{x} - \bar{y}|| = c$), a contradiction. So, we obtain

$$\begin{aligned} d(P, P^k) &= ||x^0 - y^0|| = c(x^0 - y^0) \\ &= \max_{x \in P^k} cx - \max_{x \in P} cx = gap_P^k(c). \end{aligned}$$

Now, assume by contradiction that $\exists c'$ s.t. $gap_P^k(c') > gap_P^k(c)$ and $||c'|| = 1$. Let $x' \in P^k, y' \in P$ denote the points at which $gap_P^k(c')$ is attained. Using the definition of $d(P, P^k)$ and the relation between c and c'

$$\begin{aligned} d(P, P^k) \geq ||x' - y'|| &= \frac{(x' - y')}{||x' - y'||}(x' - y') \\ &= \max_{c : ||c|| = 1} c(x' - y') \geq c'(x' - y') \\ &= gap_P^k(c') > gap_P^k(c) = d(P, P^k), \end{aligned}$$

a contradiction. Thus, we must have $d(P, P^k) = \max_{c : ||c|| = 1} gap_P^k(c)$. □

Appendix B

A polytope $P \subseteq \mathbb{R}_+^n$ is called *down-monotone* if whenever $x \in P$ and $0 \leq y \leq x$, we have $y \in P$. We begin with some preliminary results about the symmetrization we employ.

Lemma 19. For a down-monotone polytope $P \subseteq \mathbb{R}_+^n$ we have $\overline{P} = \bigcup_{I \subseteq [n]} P^I$.

Proof. It is sufficient to prove that the set $\bigcup_{I \subseteq [n]} P^I$ is convex. For that, consider $y^1, y^2 \in \bigcup_{I \subseteq [n]} P^I$; by definition, let $x^1, x^2 \in P$ be such that there are sets I_1, I_2 giving $(x^1)^{I_1} = y^1$ and $(x^2)^{I_2} = y^2$. For any $\lambda \in [0, 1]$, consider $y = \lambda y^1 + (1 - \lambda)y^2$; we show $y \in \bigcup_{I \subseteq [n]} P^I$.

By construction we have:

$$y_i = \begin{cases} \lambda x_i^1 + (1 - \lambda)x_i^2 & i \in I^1 \cap I^2 \\ \lambda x_i^1 - (1 - \lambda)x_i^2 & i \in I^1 \backslash I^2 \\ -\lambda x_i^1 + (1 - \lambda)x_i^2 & i \in I^2 \backslash I^1 \\ -\lambda x_i^1 - (1 - \lambda)x_i^2 & i \in [n] \backslash I^1 \cup I^2 \end{cases}$$

Now, let $\bar{I} = \{i \in [n] : y_i \geq 0\}$. Then define $x := y^{\bar{I}}$, which is nonnegative by construction. By non-negativity of the x^i's, we have $|\lambda x_i^1 - (1 - \lambda)x_i^2| \leq \lambda x_i^1 + (1 - \lambda)x_i^2$ and $| - \lambda x_i^1 + (1 - \lambda)x_i^2| \leq \lambda x_i^1 + (1 - \lambda)x_i^2$, thus $x \leq x^1 + (1 - \lambda)x^2 \in P$. Since P is down-monotone, we have that x belongs to P. Since $y = x^{\bar{I}}$, this gives that y belongs to $\bigcup_{I \subseteq [n]} P^I$, concluding the proof. □

Lemma 20. For a down-monotone polytope $P \subseteq \mathbb{R}_+^n$ we have $(\overline{P})^k = \overline{P^k}$.

Proof. We break the proof into a couple of claims.

Claim 21. $(\overline{P})^k \cap \mathbb{R}_+^n = P^k = \overline{P^k} \cap \mathbb{R}_+^n$.

Proof. For the first equality, notice that since $\overline{P}^k \supseteq P^k$ it suffices to prove $(\overline{P})^k \cap \mathbb{R}^n_+ \subseteq P^k$. For any $x \in (\overline{P})^k \cap \mathbb{R}^n_+$ and $I \subseteq \binom{[n]}{k}$, there exists $y \in \overline{P}$ such that $y|_I = x|_I$. Moreover, using the fact that $x \geq 0$ and the symmetry in the definition of $\overline{P}$, there is one such y which is non-negative, and hence $y \in P$. But again using $x|_I = y|_I$, we get that $x \in P^k$.

For the second equality, since P is down-monotone we have that P^k is down monotone. Therefore, from Lemma 19 $\overline{P^k} = \bigcup_{I \subseteq [n]} (P^k)^I$, which implies $\overline{P^k} \cap \mathbb{R}^n_+ = P^k$. ◇

Claim 22. Consider $z \in (\overline{P})^k$ and let $y = z^I$ for some $I \subseteq [n]$. Then $y \in (\overline{P})^k$.

Proof. First note that it is straight forward to verify that if $\alpha x \leq b$ is a valid inequality for $\overline{P}$, then for every $I \subseteq [n]$ the inequality $a^I x \leq b$ is also a valid inequality for $\overline{P}$. Then the point y must belong to $(\overline{P})^k$, since otherwise y would be separated by some k-sparse cut $ax \leq b$ and so z would be separated by the k-sparse cut $a^I x \leq b$. ◇

Now we conclude the proof of the lemma. For the direction $(\overline{P})^k \subseteq \overline{P^k}$, let $z \in (\overline{P})^k$ and let $I = \{i \in [n] : z_i \geq 0\}$ and $x = z^I$. Then using Claim 22 we get $x \in (\overline{P})^k \cap \mathbb{R}^n_+$. Thus by Claim 21 we have $x \in P^k$ and hence $z \in \overline{P^k}$, concluding this part of the proof. For the direction $\overline{P^k} \subseteq (\overline{P})^k$, let $z \in \overline{P^k}$. Let $I = \{i \in [n] : z_i \geq 0\}$ and $x = z^I$. The point $x \in \overline{P^k} \cap \mathbb{R}^n_+$. Thus, by Claim 21 we have that $x \in (\overline{P})^k \cap \mathbb{R}^n_+$. However, by Claim 22 we have that $z \in (\overline{P})^k$. This concludes the proof. □

The next result together with Lemma 20 implies Lemma 4.

Proposition 23. Consider non-negative vectors $a^1, \ldots, a^m \in \mathbb{R}^n_+$ and define the polyhedron $P = \{x \in \mathbb{R}^n_+ \mid a^i x \leq b_i \ \forall i \in [m]\}$. Then $\overline{P} = \{x \mid (a^i)^I x \leq b_i \ \ \forall I \subseteq [n], \ \forall i \in [m]\}$.

Proof. ($\overline{P} \subseteq \{x \mid (a^i)^I x \leq b_i \ \forall I \subseteq [n], \ \forall i \in [m]\}$) Consider $z \in \overline{P}$ and define $I = \{i \in [n] : z_i \geq 0\}$. Then $z^I \in \overline{P} \cap \mathbb{R}^n_+$ and thus $z^I \in P$ (from Lemma 19). Now observe that $(a^i)^I z = a_i z^I \leq b_i$ where the last inequality follows from that fact that $z^I \in P$. This concludes this part of the proof.

($\{x \mid (a^i)^I x \leq b_i \ \forall I \subseteq [n], \ \forall i \in [m]\} \subseteq \overline{P}$) Consider $z \in \{x \mid (a^i)^I x \leq b_i \ \forall I \subseteq [n], \ \forall i \in [m]\}$. Let $I = \{i \in [n] : z_i \geq 0\}$. Then observe that $a_i z^I = (a^i)^I z \leq b_i$ for all $i \in [m]$ and $z^I \in \mathbb{R}^n_+$. Thus, $z^I \in P$ or equivalently, $z \in \overline{P}$. This concludes the proof. □