**Lecture 5: Lifts of polytopes and non-negative rank**
CSE 599S: Entropy optimality, Winter 2016
**Instructor:** James R. Lee **Last updated:** January 24, 2016

# 1 Lifts of polytopes

## 1.1 Polytopes and inequalities

Recall that the *convex hull* of a subset $X \subseteq \mathbb{R}^n$ is defined by

$$\mathrm{conv}(X) = \{\lambda x + (1-\lambda)x' : x, x' \in X, \lambda \in [0,1]\}.$$

A *d-dimensional convex polytope* $P \subseteq \mathbb{R}^d$ is the convex hull of a finite set of points in $\mathbb{R}^d$:

$$P = \mathrm{conv}(\{x_1, \ldots, x_k\})$$

for some $x_1, \ldots, x_k \in \mathbb{R}^d$.

Every polytope has a dual representation: It is a closed and bounded set defined by a family of linear inequalities

$$P = \{x \in \mathbb{R}^d : Ax \leqslant b\}$$

for some matrix $A \in \mathbb{R}^{m \times d}$.

Let us define a measure of complexity for $P$: Define $\gamma(P)$ to be the smallest number $m$ such that for some $C \in \mathbb{R}^{s \times d}$, $y \in \mathbb{R}^s$, $A \in \mathbb{R}^{m \times d}$, $b \in \mathbb{R}^m$, we have

$$P = \{x \in \mathbb{R}^d : Cx = y \text{ and } Ax \leqslant b\}.$$

In other words, this is the minimum number of *inequalities* needed to describe $P$. If $P$ is full-dimensional, then this is precisely the number of *facets* of $P$ (a facet is a maximal proper face of $P$).

Thinking of $\gamma(P)$ as a measure of complexity makes sense from the point of view of optimization: Interior point methods can efficiently optimize linear functions over $P$ (to arbitrary accuracy) in time that is polynomial in $\gamma(P)$.

## 1.2 Lifts of polytopes

Many simple polytopes require a large number of inequalities to describe. For instance, the *cross-polytope*

$$C_d = \{x \in \mathbb{R}^d : \|x\|_1 \leqslant 1\} = \{x \in \mathbb{R}^d : \pm x_1 \pm x_2 \cdots \pm x_d \leqslant 1\}$$

has $\gamma(C_d) = 2^d$. On the other hand, $C_d$ is the *projection* of the polytope

$$Q_d = \left\{(x, y) \in \mathbb{R}^{2d} : \sum_{i=1}^{n} y_i = 1,\ y_i \geqslant 0,\ -y_i \leqslant x_i \leqslant y_i\ \forall i\right\}$$

onto the $x$ coordinates, and manifestly, $\gamma(Q_d) \leqslant 3d$. Thus $C_d$ is the (linear) shadow of a much simpler polytope in a higher dimension.

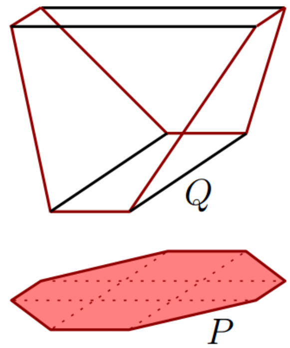


Figure 1: A lift $Q$ of a polytope $P$. [Source: Fiorini, Rothvoss, and Tiwary]

A polytope $Q$ is called a *lift* of the polytope $P$ if $P$ is the image of $Q$ under an affine projection, i.e. $P = \pi(Q)$, where $\pi : \mathbb{R}^N \to \mathbb{R}^n$ is the composition of a linear map and possibly a translation and $N \geqslant n$. By applying an affine map first, one can assume that the projection is merely coordinate projection to the first $n$ coordinates.

Again, from an optimization stand point, lifts are important: If we can optimize linear functionals over $Q$, then we can optimize linear functionals over $P$. For instance, if $P$ is obtained from $Q$ by projecting onto the first $n$ coordinates and $w \in \mathbb{R}^n$, then

$$\max_{x \in P} \langle w, x \rangle = \max_{y \in Q} \langle \bar{w}, y \rangle \,,$$

where $\bar{w} \in \mathbb{R}^N$ is given by $\bar{w} = (w, 0, 0, \ldots, 0)$.

This motivates the definition

$$\bar{\gamma}(P) = \min\{\gamma(Q) : Q \text{ is a lift of } P\} \,.$$

The value $\bar{\gamma}(P)$ is sometimes called the *(linear) extension complexity* of $P$.

**Exercise (1 point) 1.1.** Prove that $\gamma(C_d) = 2^d$.

### 1.2.1 The permutahedron

Here is a somewhat more interesting family of examples where lifts reduce complexity. The *permutahedron* $\Pi_n \subseteq \mathbb{R}^n$ is the convex hull of the vectors $(i_1, i_2, \ldots, i_n)$ where $\{i_1, \ldots, i_n\} = \{1, \ldots, n\}$. It is known that $\gamma(\Pi_n) = 2^n - 2$.

Given a permutation $\pi : [n] \to [n]$, the corresponding *permutation matrix* is defined by

$$P_\pi = \begin{pmatrix} e_{\pi(1)} \\ e_{\pi(2)} \\ \vdots \\ e_{\pi(n)} \end{pmatrix},$$

where $e_1, e_2, \ldots, e_n$ are the standard basis vectors.

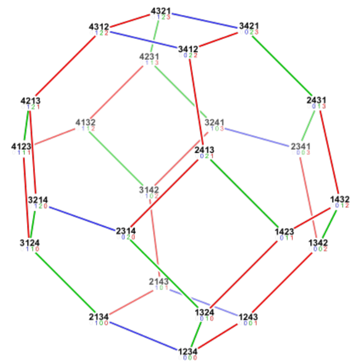


Figure 2: The permutahedron of order 4. [Source: Wikipedia]

Let $B_n \subseteq \mathbb{R}^{n^2}$ denote the convex hull of the $n \times n$ permutation matrices. The Birkhoff-von Neumann theorem tells us that $B_n$ is precisely the set of doubly stochastic matrices:

$$B_n = \left\{ M \in \mathbb{R}^{n \times n} : \sum_i M_{ij} = \sum_j M_{ij} = 1, M_{ij} \geqslant 0 \quad \forall i, j \right\},$$

thus $\gamma(B_n) \leqslant n^2$ (corresponding to the non-negativity constraints on each entry).

Observe that $\Pi_n$ is the linear image of $B_n$ under the map $A \mapsto (1, 2, \ldots, n)A$, i.e. we multiply a matrix $A \in B_n$ on the left by the row vector $(1, 2, \ldots, n)$. Thus $B_n$ is a lift of $\Pi_n$, and we conclude that $\bar{\gamma}(\Pi_n) \leqslant n^2 \ll \gamma(\Pi_n)$.

### 1.2.2 The cut polytope

If $P \neq NP$, there are certain combinatorial polytopes we should not be able to optimize over efficiently. A central example is the *cut polytope:* $\mathrm{CUT}_n \subseteq \mathbb{R}^{\binom{n}{2}}$ is the convex hull of all all vectors of the form

$$v^S_{\{i,j\}} = |\mathbf{1}_S(i) - \mathbf{1}_S(j)| \qquad \{i, j\} \in \binom{[n]}{2}$$

for some subset $S \subseteq \{1, \ldots, n\}$. Here, $\mathbf{1}_S$ denotes the characteristic function of $S$.

Note that the MAX-CUT problem on a graph $G = (V, E)$ can be encoded in the following way: Let $W_{ij} = 1$ if $\{i, j\} \in E$ and $W_{ij} = 0$ otherwise. Then the value of the maximum cut in $G$ is precisely the maximum of $\langle W, A \rangle$ for $A \in \mathrm{CUT}_n$. Accordingly, we should expect that $\bar{\gamma}(\mathrm{CUT}_n)$ cannot be bounded by any polynomial in $n$ (lest we violate a basic tenet of complexity theory).

Our goal in this lecture and the next will be to show that the cut polytope does not admit lifts with $n^{O(1)}$ facets.

### 1.2.3 Exercises

**Exercise (1 point) 1.2.** Define the *bipartite perfect matching polytope* $\mathrm{BM}_n \subseteq \mathbb{R}^{n^2}$ as the convex hull of all the indicator vectors of edge sets of perfect matchings in the complete bipartite graph $K_{n,n}$. Show that $\gamma(\mathrm{BM}_n) \leqslant n^2$.

**Exercise (1 point) 1.3.** Define the *subtour elimination polytope* $\mathrm{SEP}_n \subseteq \mathbb{R}^{\binom{n}{2}}$ as the set of points $x = (x_{ij}) \in \mathbb{R}^{\binom{n}{2}}$ satisfying the inequalities

$$\begin{aligned} x_{ij} &\geqslant 0 \qquad && \{i,j\} \in \binom{[n]}{2} \\ \sum_{i=1}^{n} x_{ij} &= 2 \qquad && j \in [n] \\ \sum_{i \in S} \sum_{j \notin S} x_{ij} &\geqslant 2 \qquad && S \subseteq [n], 2 \leqslant |S| \leqslant n-2 \,. \end{aligned}$$

Show that $\bar{\gamma}(\mathrm{SEP}_n) \leqslant O(n^3)$ by think of the $x_{ij}$ variables as edge capacities, and introducing new variables to enforce that the capacities support a flow of value 2 between every pair $i, j \in [n]$.

**Exercise (1 point) 1.4** (Goemans)**.** Show that for any polytope $P$,

$$\#\text{ faces of } P \leqslant 2^{\#\text{ facets of } P} \,.$$

Recall that a facet of $P$ is a face of largest dimension. (Thus if $P \subseteq \mathbb{R}^n$ is full-dimensional, then a facet of $P$ is an $(n-1)$-dimensional face.) Use this to conclude that $\bar{\gamma}(\Pi_n) \geqslant \log(n!) \geqslant \Omega(n \log n)$.

**Exercise (1 point) 1.5** (Martin, 1991)**.** Define the *spanning tree polytope* $\mathrm{ST}_n \subseteq \mathbb{R}^{\binom{n}{2}}$ as the convex hull of all the indicator vectors of spanning trees in the complete graph $K_n$. Show that $\bar{\gamma}(\mathrm{ST}_n) \leqslant O(n^3)$ by introducing new variables $\{z_{uv,w} : u, v, w \in \{1, 2, \ldots, n\}\}$ meant to represent whether the edge $\{u, v\}$ is in the spanning tree $T$ and $w$ is in the component of $v$ when the edge $\{u, v\}$ is removed from $T$.

## 2 Non-negative matrix factorization

The key to understanding $\bar{\gamma}(\mathrm{CUT}_n)$ comes from Yannakakis' factorization theorem.

Consider a polytope $P \subseteq \mathbb{R}^d$ and let us write in two ways: As a convex hull of vertices

$$P = \mathrm{conv}\left(\{x_1, x_2, \ldots, x_n\}\right) ,$$

and as an intersection of half-spaces: For some $A \in \mathbb{R}^{m \times d}$,

$$P = \left\{ x \in \mathbb{R}^d : Ax \leqslant b \right\} .$$

Given this pair of representations, we can define the corresponding *slack matrix* of $P$ by

$$S_{ij} = b_i - \langle A_i, x_j \rangle \qquad i \in \{1, 2, \ldots, m\}, j \in \{1, 2, \ldots, n\} \,.$$

Here, $A_1, \ldots, A_m$ denote the rows of $A$.

We need one more definition. If we have a non-negative matrix $M \in \mathbb{R}_+^{m \times n}$, then a *rank-r non-negative factorization of M* is a factorization $M = AB$ where $A \in \mathbb{R}_+^{m \times r}$ and $B \in \mathbb{R}_+^{r \times n}$. We then define the *non-negative rank of M*, written $\mathrm{rank}_+(M)$, to be the smallest $r$ such that $M$ admits a rank-$r$ non-negative factorization.

**Exericse (0.5 points) 2.1.** Show that $\mathrm{rank}_+(M)$ is the smallest $r$ such that $M = M_1 + \cdots + M_r$ where each $M_i$ is a non-negative matrix satisfying $\mathrm{rank}_+(M_i) = 1$.

The next result gives a precise connection between non-negative rank and extension complexity.

**Theorem 2.2** (Yannakakis Factorization Theorem)**.** *For every polytope $P$, it holds that $\bar{\gamma}(P) = \mathrm{rank}_+(S)$ for any slack matrix $S$ of $P$.*

The key fact underlying this theorem is Farkas' Lemma (see Section Section 2.1 for a proof). Recall that a function $f : \mathbb{R}^d \to \mathbb{R}$ is *affine* if $f(x) = \langle a, x\rangle - b$ for some $a \in \mathbb{R}^d$ and $b \in \mathbb{R}$. Given functions $f_1, \ldots, f_k : \mathbb{R}^d \to \mathbb{R}$, denote their non-negative span by

$$\mathrm{cone}\left(\{f_1, f_2, \ldots, f_k\}\right) = \left\{\sum_{i=1}^{k} \lambda_i f_i : \lambda_i \geqslant 0\right\}.$$

**Lemma 2.3** (Farkas Lemma)**.** *Consider a polytope $P = \{x \in \mathbb{R}^d : Ax \leqslant b\}$ where $A$ has rows $A_1, A_2, \ldots, A_m$. Let $f_i(x) = b_i - \langle A_i, x\rangle$ for each $i = 1, \ldots, m$. If $f$ is any affine function such that $f|_P \geqslant 0$, then*

$$f \in \mathrm{cone}(\{f_1, f_2, \ldots, f_m\}).$$

The lemma asserts if $P = \{x \in \mathbb{R}^d : Ax \leqslant b\}$, then every valid linear inequality over $P$ can be written as a non-negative combination of the defining inequalities $\langle A_i, x\rangle \leqslant b_i$.

**Exericse (0.5 points) 2.4.** Use Farkas' Lemma to prove that if $S$ and $S'$ are two different slack matrices for the same polytope $P$, then $\mathrm{rank}_+(S) = \mathrm{rank}_+(S')$.

There is an interesting connection here to proof systems. The theorem says that we can interpret $\bar{\gamma}(P)$ as the minimum number of axioms so that every valid linear inequality for $P$ can be proved using a conic (i.e., non-negative) combination of the axioms.

To conclude this section, let us now prove the Yannakakis Factorization Theorem.

*Proof of Theorem 2.2.* Let us write $P = \{x \in \mathbb{R}^d : Ax \leqslant b\} = \mathrm{conv}(V)$ where $V = \{x_1, \ldots, x_N\}$ and $A \in \mathbb{R}^{m\times d}$. Let $M_{ij} = b_i - \langle A_i, x_j\rangle$ denote the associated slack matrix.

First, let us suppose there is a lift $Q \subseteq \mathbb{R}^{d+d'}$ of $P \subseteq \mathbb{R}^d$ given by $r$ inequalities. We may assume that

$$Q = \left\{(x, y) \in \mathbb{R}^d \times \mathbb{R}^{d'} : Rx + Sy = t, Ux + Vy \leqslant c\right\},$$

and $P$ is the projection of $Q$ to the first $d$ coordinates, and where $U \in \mathbb{R}^{r\times d}$ and $V \in \mathbb{R}^{r\times d'}$.

Now observe that the inequalities $Ax \leqslant b$ are valid for $Q$ simply because if $(x, y) \in Q$ then $x \in P$. For every $x_j \in P$, let $y_j \in \mathbb{R}^{d'}$ be such that $(x_j, y_j) \in Q$. Let $Z \in \mathbb{R}^{(r+m)\times N}$ denote the matrix that records the slack of the $r$ inequalities of $Q$ at the points $(x_1, y_1), \ldots, (y_N, x_N)$, and then in the last $m$ rows the slack of the inequalities $Ax \leqslant b$.

Then we have: $\mathrm{rank}_+(M) \leqslant \mathrm{rank}_+(Z)$ (since $M$ is precisely the last $m$ rows of $Z$). But Lemma 2.3 tells us that the last $m$ rows of $Z$ are non-negative combinations of the first $r$ rows, hence $\mathrm{rank}_+(Z) \leqslant r$, and we conclude that $\mathrm{rank}_+(M) \leqslant r$.

Conversely, let us suppose there is a non-negative factorization $M = KL$ where $K \in \mathbb{R}_+^{m\times r}$ and $L \in \mathbb{R}_+^{r\times N}$. We claim that the $x$-coordinate projection of

$$Q = \{(x, y) \in \mathbb{R}^{d+r} : Ax + Ky = b, y \geqslant 0\}$$

is precisely $P$, which will imply that $\bar{\gamma}(P) \leqslant r$. This is not quite true: One should also verify that $Q$ is a polytope, which means it should be bounded. For that to be true, it should be true that no column of $K$ is identically zero. But this is easy to enforce: If not, we can find a factorization of smaller rank by deleting that column and the corresponding row of $L$.

Note that $\text{proj}_x(Q) \subseteq P$ because $Ky \geqslant 0$; this is where we use the fact that $K$ is non-negative. For the other direction $P \subseteq \text{proj}_x(Q)$, we need to find for every vertex $x_j$ of $P$ a point $y_j \in \mathbb{R}^r$ such that $(x_j, y_j) \in Q$. We simply take $y_j$ to be the $j$th column of $L$, noting that

$$Ax_j + Ky_j = Ax_j + (b - Ax_j) = b$$

and also $y_j \geqslant 0$ (this is where we use that $L$ is non-negative). □

### 2.1 Proof of Farkas' Lemma

**Exercise (2 points) 2.5.** Prove Farkas' Lemma by completing each of the following steps. Recall that $P = \{x \in \mathbb{R}^d : Ax \leqslant b\}$ is a polytope and $A \in \mathbb{R}^{m \times d}$. Let $A_1, \ldots, A_m$ denote the rows of $A$.

1. Let $\mathcal{A} = \{f : \mathbb{R}^d \to \mathbb{R} \mid f \text{ is affine}\}$. Give a natural interpretation of $\mathcal{A}$ as a $(d+1)$-dimensional vector space; addition of functions should have the natural meaning $(f+g)(x) = f(x) + g(x)$.

2. Let $f_1, f_2, \ldots, f_m : \mathbb{R}^d \to \mathbb{R}$ be the $m$ affine functions given by $f_i(x) = b_i - \langle A_i, x\rangle$. Show that for $x \in \mathbb{R}^d$,
$$x \in P \iff f(x) \geqslant 0 \quad \forall f \in \text{cone}(\{f_1, \ldots, f_m\}).$$

3. Consider the following fundamental fact.

   **Theorem 2.6** (Hyperplane separation theorem)**.** *For any $n \geqslant 1$, if $K \subseteq \mathbb{R}^n$ is a non-empty, closed convex set and $y \notin K$, then there is a vector $v \in \mathbb{R}^n$ and value $b \in \mathbb{R}$ such that $\langle v, z\rangle \geqslant b$ for all $z \in K$, but $\langle v, y\rangle < b$.*

   Use this theorem in conjunction with (i) and (ii) to prove that if $f \notin \text{cone}(\{f_1, \ldots, f_m\})$ then there is a point $x \in P$ such that $f(x) < 0$. [Hint: This will be the tricky part. One needs to use the fact that $P$ is bounded.] Conclude that Lemma 2.3 is true.

4. We are left to prove Theorem 2.6. Without loss of generality, we can assume that $y = 0$. Argue that the optimization $\min_{z \in K} \|z\|^2$ has a unique solution. Let $z^*$ be the optimizer, and show that one can take $v = z^*$ to prove the theorem.

### 2.2 Slack matrices and the correlation polytope

Thus to prove a lower bound on $\bar{\gamma}(\text{CUT}_n)$, it suffices to find a valid set of linear inequalities for $\text{CUT}_n$ and prove a lower bound on the non-negative rank of the corresponding slack matrix.

Toward this end, consider the correlation polytope $\text{CORR}_n \subseteq \mathbb{R}^{n^2}$ given by

$$\text{CORR}_n = \text{conv}\left(\{xx^T : x \in \{0,1\}^n\}\right).$$

**Exericse (0.5 points) 2.7.** Show that for every $n \geqslant 1$, $\text{CUT}_{n+1}$ and $\text{CORR}_n$ are linearly isomorphic.

Now we identify a slack matrix for $\mathrm{CORR}_n$. Denote by

$$\mathbb{R}_2[x_1, \ldots, x_n] = \left\{ a_0 + \sum_i a_i x_i + \sum_{i,j} a_{ij} x_i x_j \right\} .$$

the set of quadratic polynomials on $\mathbb{R}^n$. Let

$$\mathsf{QML}^n = \{ f : \{0,1\}^n \to \mathbb{R} : f = g|_{\{0,1\}^n} \text{ for some } g \in \mathbb{R}_2[x_1, \ldots, x_n] \}$$

be the functions given by restricting quadratic polynomials to the discrete cube.

Observe that every $f \in \mathsf{QML}^n$ can be written as a multi-linear function

$$f(x) = a_0 + \sum_i a_i x_i + \sum_{i \neq j} a_{ij} x_i x_j$$

since $x_i^2 = x_i$ for $x_i \in \{0,1\}$. Finally, define the set of non-negative quadratic multi-linear functions

$$\mathsf{QML}^n_+ = \{ f \in \mathsf{QML}^n : f(x) \geqslant 0 \quad \forall x \in \{0,1\}^n \} .$$

**Lemma 2.8.** *Define the (infinite) matrix $\mathcal{M}_n : \mathsf{QML}^n_+ \times \{0,1\}^n \to \mathbb{R}_+$ by*

$$\mathcal{M}_n(f, x) = f(x) .$$

*Then $\mathcal{M}_n$ is a slack matrix for* $\mathrm{CORR}_n$.

*Proof.* Consider $f \in \mathsf{QML}^n_+$. Recalling that $x_i = x_i^2$, we can write

$$f(x) = b - \sum_i A_{ii} x_i^2 - \sum_{i \neq j} A_{ij} x_i x_j$$

for some symmetric matrix $A \in \mathbb{R}^{n \times n}$ and $b \in \mathbb{R}$.

Define the Frobenius inner product on matrices $A, B \in \mathbb{R}^{n \times n}$ by

$$\langle A, B \rangle = \mathrm{Tr}(A^T B) = \sum_{i,j} A_{ij} B_{ij} ,$$

and observe that

$$f(x) = b - \langle A, xx^T \rangle .$$

Since $f(x) \geqslant 0$ for all $x \in \{0,1\}^n$, we have $b - \langle A, xx^T \rangle \geqslant 0$ for all $x \in \{0,1\}^n$, hence by convexity $\langle A, Y \rangle \leqslant b$ holds for all $Y \in \mathrm{CORR}_n$. The quantity $f(x)$ is precisely the slack of this inequality at the vertex $x$. □

**Exercise (0.5 points) 2.9.** Complete the preceding proof by showing that the family of linear inequalities underlying $\mathcal{M}_n$ characterize $\mathrm{CORR}_n$.

Combining Exercise 2.7 and Lemma 2.8 yields the following.

**Theorem 2.10.** *For all $n \geqslant 1$, it holds that $\bar{\gamma}(\mathrm{CUT}_{n+1}) \geqslant \mathrm{rank}_+(\mathcal{M}_n)$.*