Extended Formulations

Definition 1 (extension)**.** *Let* $P = \{x \mid Ax \le b\} \subseteq \mathbb{R}^n$ *be a polyhedron. A polyhedron* $Q \subseteq \mathbb{R}^d$ *is said to be an* extension *for* P *if* $\text{proj}_x(Q) = P$*, where* $\text{proj}_x(Q) := \{x \mid \exists y : (x, y) \in Q\}$.

The *size* of an extension is the number of its facets.

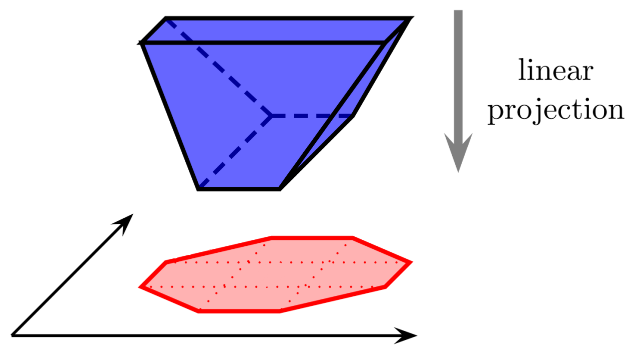


Figure 1: Extended formulations can be smaller.

Definition 2 (Extended formulation)**.** *A particular representation of an extension by linear inequalities is called an extended formulation, and its size is the number of inequalities.*

Disclaimer: Some people use "extended formulation" to refer to any MIP formulation that uses additional variables. This should be clear from the context.

If a research paper has polyhedral results, they are probably referring to Definition 2. If it gives MIP formulations but no polyhedral analysis, they are probably referring to the latter meaning.

Example: Uncapacitated Lot Sizing

The problem is to decide on a production plan for an n-period horizon for a single product. The basic model can be viewed as having data:

- f_t is the fixed cost of producing in period t.
- p_t is the unit production cost in period t.
- h_t is the unit storage cost in period t.
- d_t is the demand in period t.

We use the natural (or obvious) variables:

- x_t is the amount produced in period t.
- s_t is the stock at the end of period t.
- $y_t = 1$ if production occurs in period t, and $y_t = 0$ if otherwise.

An MIP formulation in these natural variables is as follows.

$$
\begin{aligned}
&\min \sum_{t=1}^{n} p_t x_t + \sum_{t=1}^{n} h_t s_t + \sum_{t=1}^{n} f_t y_t && \\
&s_{t-1} + x_t = d_t + s_t && t = 1, \ldots, n \\
&x_t \leq M y_t && t = 1, \ldots, n \\
&s_0 = 0 && \\
&s_t \geq 0 && t = 1, \ldots, n \\
&x_t \geq 0 && t = 1, \ldots, n \\
&y_t \in \{0, 1\} && t = 1, \ldots, n.
\end{aligned}
$$

An extended formulation uses additional variables of the form w_{it} representing the amount produced on day i to fulfill demand for day t. (Only define these variables for $i \leq t$.)

$$
\begin{aligned}
&\sum_{i=1}^{t} w_{it} = d_t && t = 1, \ldots, n \\
&w_{it} \leq d_t y_i && 1 \leq i \leq t \leq n \\
&x_i = \sum_{t=i}^{n} w_{it} && i = 1, \ldots, n \\
&s_{t-1} + x_t = d_t + s_t && t = 1, \ldots, n \\
&s_0 = 0 && \\
&w_{it} \geq 0 && 1 \leq i \leq t \leq n \\
&0 \leq y_t \leq 1 && t = 1, \ldots, n.
\end{aligned}
$$

Exercise. Let P_1 be the set of all (x, y, s) satisfying the first formulation. Further let Q_2 be the set of all (x, y, s, w) satisfying the extended formulation. Show that Q_2 is a stronger formulation than P_1. These polyhedra are in different variable spaces so you should actually compare the projection of Q_2 into the space of x, y, s variables, i.e., $P_2 := \text{proj}_{x,y,s}(Q_2)$.

Remark 1. *Q_2 is a perfect (extended) formulation, i.e., any basic feasible solution has y integer.*

Example: Spanning Tree Polytope

Let $G = (V, E)$ be a graph. Its spanning tree polytope is:

$$P_{ST}(G) := \operatorname{conv}\left\{x^F \in \{0,1\}^m \;\middle|\; (V, F) \text{ is a tree}\right\}.$$

By Edmonds (1971), an (exponentially-sized) perfect formulation for $P_{ST}(G)$ is as follows.

$$\begin{aligned}
&\sum_{e \in E} x_e = n - 1 \\
&\sum_{e \in E(S)} x_e \leq |S| - 1 && S \subseteq V,\ 2 \leq |S| \leq n - 1 \\
&x \geq 0.
\end{aligned}$$

There exist extended formulations for $P_{ST}(G)$ of size $O(nm)$. To do this, first bidirect the edges of G to get $D = (V, A)$ and arbitrarily pick a source node $s \in V$. Create the following variables:

- for each $a \in A$, the binary variable $y_a = 1$ iff arc a is chosen;
- for each arc $a \in A$ and each $t \in V \setminus \{s\}$, create a variable z_a^t that denotes the amount of flow sent across arc a to reach demand node t (sent from source node s).

In the formulation we should send one unit of flow from s to t. Moreover, the flow that we are sending to t should be of "commodity" t. This is thus a "multicommodity flow" formulation.

Proposition 1. *The above is an extended formulation for the spanning tree polytope, i.e., $x \in P_{ST}(G)$ if and only if there exist y and z such that (x, y, z) satisfies the above formulation.*

Remark 2. *The formulation in the y variable space is for directed spanning trees rooted at s.*

Example: Subtour Elimination Polytope for TSP

Recall our cut-based LP relaxation for asymmetric (i.e., directed) TSP over $G = (V, A)$.

$$\min \sum_{a \in A} c_a x_a \tag{1}$$

$$\sum_{a \in \delta^+(v)} x_a = 1 \qquad v \in V \tag{2}$$

$$\sum_{a \in \delta^-(v)} x_a = 1 \qquad v \in V \tag{3}$$

$$\sum_{a \in \delta^+(S)} x_a \geq 1 \qquad S \subseteq V,\ 2 \leq |S| \leq n-2 \tag{4}$$

$$x \geq 0. \tag{5}$$

Just like in the spanning tree formulation, we can use multicommodity flows to replace the exponentially many cut constraints (4) with a polynomial number of flow variables/constraints:

Proposition 2. *The multicommodity flow extended formulation for ATSP and the cut-based formulation for ATSP are equally strong.*

Remark 3. *In practice, the cut-based formulation (implemented via branch-and-cut) is computationally superior to the extended formulation.*

We have given a polysize extended formulation for the (directed) *subtour elimination* polytope, which is the LP relaxation for the cut-based ATSP formulation. Do there exist small extended formulations for its integer hull (the ATSP polytope)?

Theorem 1 (Fiorini et al., 2012)**.** *TSP polytopes have extension complexity* $2^{\Omega(\sqrt{n})}$.

Is it true that "easy" problems admit small extended formulations and "hard" problems do not?

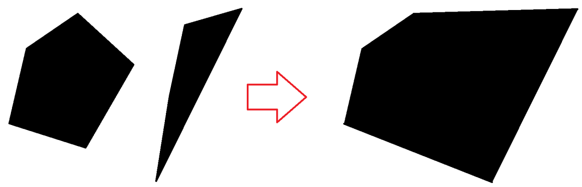
Theorem 2 (Rothvoß, 2014)**.** *Matching polytopes have extension complexity* $2^{\Omega(n)}$.

Theorem 3 (Rothvoß, 2014)**.** *TSP polytopes have extension complexity* $2^{\Omega(n)}$.

Extended Formulation for the Union of Polyhedra

A notable "meta" extended formulation is for the union of polyhedra.

Theorem 4 (Balas)**.** *If $P_1, P_2, \ldots, P_q$ are polytopes defined by $s_1, s_2, \ldots, s_q$ inequalities, respectively, then $P := \operatorname{conv}(P_1 \cup P_2 \cup \cdots \cup P_q)$ admits an extended formulation of size $q + s_1 + s_2 + \cdots + s_q$.*



Example. Consider the following "even" set:

$$S_n := \left\{ x \in \{0,1\}^n \;\middle|\; \sum_{i=1}^{n} x_i \text{ is even} \right\}$$

For any $S \subseteq [n]$ with $|S|$ odd we can write the valid inequality:

$$\sum_{i \in S} (1 - x_i) + \sum_{i \in [n] \setminus S} x_i \geq 1. \tag{6}$$

Theorem 5 (Jeroslow, 1975)**.** *The inequalities* (6) *induce facets of* $\operatorname{conv}(S_n)$*. Moreover, they and the bounds $0 \leq x_i \leq 1$ provide a perfect formulation for* $\operatorname{conv}(S_n)$*.*

Thus, this "even" set has exponential size inequality description in the original variables. To get a small *extended* formulation, we can take the union of the sets S_n^k where $0 \leq k \leq n$ and k even.

$$S_n^k := \left\{ x \in \{0,1\}^n \;\middle|\; \sum_{i=1}^{n} x_i = k \right\}$$

Lemma 1. *If k is an integer, then* $\operatorname{conv}(S_n^k) = \{x \in [0,1]^n \mid \sum_{i=1}^{n} x_i = k\}$*.*

Proposition 3. *The polytope* $\operatorname{conv}(S_n)$ *admits an extended formulation of size $O(n^2)$.*