Cutting planes for MIPs

Definition 1 (Valid inequality)**.** *Let $\alpha \in \mathbb{R}^n$ and $\beta \in \mathbb{R}$. The inequality $\alpha^T u \leq \beta$ is said to be valid for a set $K \subseteq \mathbb{R}^n$ if, for every $u^* \in K$, we have $\alpha^T u^* \leq \beta$.*

Consider an mixed integer (linear) program of the form:

$$z := \max \left\{ c^T x + h^T y \;\middle|\; (x, y) \in S \right\}, \tag{1}$$

where S is a mixed integer set defined by

$$S := \left\{ (x, y) \in \mathbb{Z}_+^n \times \mathbb{R}_+^d \;\middle|\; Ax + Gy \leq b \right\}. \tag{2}$$

If we solve the LP relaxation for this MIP, we maximize $c^T x + h^T y$ over the set

$$P_0 := \left\{ (x, y) \in \mathbb{R}_+^n \times \mathbb{R}_+^d \;\middle|\; Ax + Gy \leq b \right\}. \tag{3}$$

Suppose we get a basic optimal solution $(x^0, y^0) \in P_0$ to the LP relaxation. If (x^0, y^0) also belongs to S, then we are done. Otherwise, we are not. In branch and bound, we would branch now. In cutting plane approaches, we instead "cut off" (x^0, y^0) from P_0 with a valid inequality and resolve.

Definition 2 (Cutting plane)**.** *Let $\alpha^T x + \gamma^T y \leq \beta$ be a valid inequality for a set S, and suppose $\alpha^T x^0 + \gamma^T y^0 > \beta$. The inequality $\alpha^T x + \gamma^T y \leq \beta$ is a cutting plane separating (x^0, y^0) from S.*

Repeatedly performing this procedure yields the cutting plane approach.

Solving MIPs via a cutting plane approach

In the following, P_0 represents feasible feasible region for the initial LP relaxation, and initialize the algorithm with $i = 0$.

1. solve the linear program $\max \left\{ c^T x + h^T y \mid (x, y) \in P_i \right\}$;
2. if the associated basic solution (x^i, y^i) belongs to S, then stop;
3. else, solve the *separation problem*:
   - find a cutting plane $ax + by \leq b$ separating (x^i, y^i) from S;
   - let $P_{i+1} := P_i \cap \{(x, y) \mid ax + gy \leq b\}$;
   - let $i := i + 1$;
   - go to step 1.

Some may say this is not an "algorithm" since some steps are not clearly specified. Like what?

A famous paper used a cutting plane approach to solve TSP. Can you guess the year and the authors? Here's the abstract:

"It is shown that a certain tour of 49 cities, one in each of the 48 states and Washington, D.C., has the shortest road distance."

Gomory fractional cuts

A cutting plane algorithm for solving general integer programs was developed by Gomory (1958). Its cutting planes can be generated from the simplex tableau. As an example, consider the following IP.

$$\begin{aligned}
\max \quad & 5.5x_1 + 2.1x_2 \\
& -x_1 + x_2 \leq 2 \\
& 8x_1 + 2x_2 \leq 17 \\
& x_1,\ x_2 \geq 0 \\
& x_1,\ x_2 \text{ integer.}
\end{aligned}$$

Introduce slack variables to write in equality form:

$$\begin{aligned}
\max \quad & 5.5x_1 + 2.1x_2 \\
& -x_1 + x_2 + x_3 = 2 \\
& 8x_1 + 2x_2 + x_4 = 17 \\
& x_1,\ x_2,\ x_3,\ x_4 \geq 0 \\
& x_1,\ x_2,\ x_3,\ x_4 \text{ integer.}
\end{aligned}$$

Why is it okay to enforce that x_3 and x_4 be integers?

Here is the optimal LP relaxation tableau (with objective function z):

$$\begin{aligned}
z \qquad\qquad\qquad & + 0.58x_3 + 0.76x_4 = 14.08 \\
x_2 & + 0.8x_3 + 0.1x_4 = 3.3 \\
x_1 \qquad & - 0.2x_3 + 0.1x_4 = 1.3.
\end{aligned}$$

The corresponding LP solution is $x^* = (1.3, 3.3, 0, 0)$ with objective $z = 14.08$, which is not IP feasible so we need to cut it off.

From the tableau, we have $x_2 + 0.8x_3 + 0.1x_4 = 3.3$. Since x_2 must be integer, we must have:

$$0.8x_3 + 0.1x_4 = 0.3 + k, \text{ where } k \in \mathbb{Z}.$$

Observe that $0.8x_3 + 0.1x_4$ must be nonnegative, so we can write the valid inequality:

$$0.8x_3 + 0.1x_4 \geq 0.3.$$

This is the Gomory fractional cut. How would you express it in the space of (x_1, x_2) variables?

After adding this cut, the objective decreases from $z = 14.08$ to $z = 13.8625$.

More generally, suppose that nonnegative integers $x_1, \ldots, x_n$ satisfy the equation

$$\sum_{i=1}^{n} a_i x_i = a_0,$$

where a_0 is fractional, i.e., $a_0 \notin \mathbb{Z}$. (Think of this equation as a row from the simplex dictionary.)

The associated *Gomory fractional cut* is:

$$\sum_{i=1}^{n} (a_i - \lfloor a_i \rfloor) x_i \geq a_0 - \lfloor a_0 \rfloor. \tag{4}$$

Write the Gomory fractional cut for last row of simplex tableau (from previous page) in this way.

Proposition 1. *The Gomory fractional cut* (4) *is valid for the set* $X = \left\{ x \in \mathbb{Z}_+^n \;\middle|\; \sum_{i=1}^{n} a_i x_i = a_0 \right\}$, *where we assume that* $a_0 \notin \mathbb{Z}$.

“Separation = Optimization”

The separation problem is central to the cutting plane method and to the ellipsoid method.

Problem: Separation problem for an arbitrary convex set $K \subseteq \mathbb{R}^n$.
Input: a point $x^* \in \mathbb{R}^n$.
Output: Determine whether $x^* \in K$. If $x^* \in K$, then simply return “yes”. Otherwise, return a linear inequality $\alpha x \leq \beta$ that is valid for K but that x^* violates.

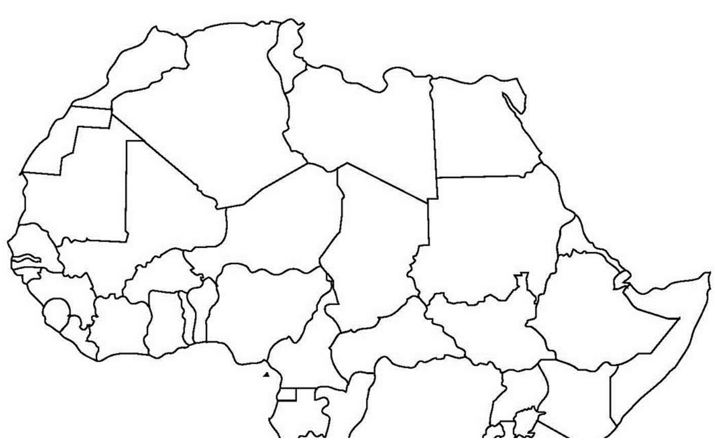
(Under some minor assumptions) If we can solve this problem in polynomial time, then we can optimize over a polyhedron K in polynomial time by the ellipsoid method!!!

This has **huge** consequences in integer and combinatorial optimization! For example, suppose that $K = \text{conv}(X)$ where, X is the set of integer feasible points. Or, imagine that K is the LP relaxation for the cut-based formulation for ATSP.

The ellipsoid method

If one wants to solve a convex (minimization) problem, it is enough to be able to test whether convex sets are nonempty. Why?

The ellipsoid method as a lion hunt in the Sahara.



The ellipsoid method was (essentially) developed by Shor (1970) for solving convex problems. It checks whether the convex set K is nonempty as follows.

1. Find ellipsoid E_0 with $K \subseteq E_0$.
2. Let x^* be the center of ellipsoid E_0.
3. Test if $x^* \in K$.

   If $x^* \in K$, then we conclude $K \neq \emptyset$ and stop.

   If $x^* \notin K$, then find a valid inequality $a^T x \leq b$ for K that x^* violates.
4. Find a new ellipsoid E_1 such that $E_{\text{half}} \subseteq E_1$, where $E_{\text{half}} := E_0 \cap \{x \mid a^T x < a^T x^*\}$, and
$$\frac{\text{vol}(E_1)}{\text{vol}(E_0)} \leq e^{-1/(2n)} < 1.$$
5. Let $E_0 \leftarrow E_1$ and go to step 2.

The ellipsoids shrink rather quickly, and in $O\left(n \log \frac{\text{vol}(E_0)}{\text{vol}(K)}\right)$ iterations we either:

- find a point in K, or
- determine that K cannot contain a sufficiently large ball.

Notice that the number of iterations does not depend on the number of constraints!

In the case of LP's with integer data, we can perform all the details in polynomial time (in the input size, represented in binary), so the algorithm runs in polynomial time. Also, we can perturb the problem so that K contains a sufficiently large ball if and only if $K \neq \emptyset$.

Theorem 1 (Khachiyan, 1979)**.** *LPs in polytime by properly implementing the ellipsoid method.*

What follows is the famous "separation=optimization" result. Almost the same results were obtained independently by Karp and Papadimitriou (1982) and Padberg and Rao (1981).

Theorem 2 (Grötschel, Lovász, and Schrijver, 1981)**.** *(Informal) A class of LPs can be optimized in polytime if and only if the associated separation problems can be solved in polytime.*

Corollary 1. *We can optimize a linear objective function over the exponentially-sized cut-based LP relaxation for TSP in polytime.*

The proof of this is left as an exercise. Remark: no one actually solves it this way in practice!