1   Cutting planes in general

The cutting plane method is really a family of strategies for solving integer programs.

The philosophy is that an integer program can have multiple linear programming formulations: different sets of linear inequalities describing the same set of integer points. When we solve the linear programming relaxation and get a fractional solution, that's a sign that our linear programming formulation wasn't very good.

So why not try to improve it?

Specifically, suppose that we have an integer program

$$
\begin{aligned}
\underset{\mathbf{x}\in\mathbb{Z}^n}{\text{maximize}} \quad & \mathbf{c}^\mathsf{T}\mathbf{x} \\
\text{subject to} \quad & A\mathbf{x} \le \mathbf{b} \\
& \mathbf{x} \ge \mathbf{0}
\end{aligned}
$$

We solve the LP relaxation, and get a fractional solution $\mathbf{x}^*$. If we want an improved formulation of this integer program, we want to generate a new inequality $\mathbf{p}^\mathsf{T}\mathbf{x} \le q$ such that:

- It's *valid* for the integer program: every point $\mathbf{x} \in \mathbb{Z}^n$ that satisfies $A\mathbf{x} \le \mathbf{b}$ and $\mathbf{x} \ge \mathbf{0}$ also satisfies $\mathbf{p}^\mathsf{T}\mathbf{x} \le q$.

  We don't want to change the problem, after all!

- It cuts off the fractional solution $\mathbf{x}^*$ we got previously: we have $\mathbf{p}^\mathsf{T}\mathbf{x}^* > q$.

  If this does not hold, then adding the new inequality won't help; we'll still get $\mathbf{x}^*$ as the optimal solution to the LP relaxation.

Such an inequality is called a **cutting plane** for the integer program. If we can come up with a cutting plane, then we can add it as an additional constraint, and solve the LP relaxation of the new integer program.

Of course, there's no guarantee that the new LP relaxation will have an integer solution, either. We might get another fractional solution, in which case we'll have to do this again. The hope is that after several steps, we'll get an integer solution.

One big question remains: where do we actually get these cutting planes to begin with?

There are many strategies, and each one results in a cutting plane method. They vary in how difficult they are (some require more or less work to come up with a cut) and how effective.

---

[1]This document comes from an archive of the Math 3272 course webpage: http://misha.fish/archive/3272-fall-2022

2 The Gomory fractional cut

The Gomory fractional cut is one strategy for coming up with cutting planes. It's quick to perform, and the cuts it produces are pretty good.

To make it work, we'll assume that we have a purely integer program: some cutting plane methods still work with a mix of integer and real variables, but this one isn't one of them. We'll also assume that all numbers in the constraints are integers. (If they're merely rational numbers, we can turn them into integers by clearing denominators.) This is important because:

Fact 1. *Given a purely integer linear program where all numbers in the constraints are integers, at every basic solution the slack variables also have integer values.*

This fact holds because, given the assumptions, every slack variable will be a difference between two *integer* quantities: the two sides of an inequality in integer variables with integer coefficients.

Normally, our slack variables are real numbers, and there is no reason to force them to be integers. However, the Gomory fractional cut cannot work unless all the variables are integer variables, and that includes the slack variables.

2.1 The general rule

The Gomory fractional cut takes an equation in integer variables, and uses it to come up with an inequality they must satisfy. The rule defining the fractional cut is:

Theorem 1. *Suppose that nonnegative integer variables $x_1, x_2, \ldots, x_n$ satisfy the equation*

$$a_1 x_1 + a_2 x_2 + \cdots + a_n x_n = b.$$

Then they also satisfy the inequality

$$\lfloor a_1 \rfloor x_1 + \lfloor a_2 \rfloor x_2 + \cdots + \lfloor a_n \rfloor x_n \leq \lfloor b \rfloor$$

*where $\lfloor r \rfloor$ denotes the **floor** of r: the greatest integer less than or equal to r. Moreover, the difference between the two sides of this inequality is an integer.*

Proof. The quantity $\lfloor a_1 \rfloor x_1 + \lfloor a_2 \rfloor x_2 + \cdots + \lfloor a_n \rfloor x_n$ has smaller or equal coefficients on every variable, compared to $a_1 x_1 + a_2 x_2 + \cdots + a_n x_n$. Since all the variables are nonnegative, it must be smaller. We conclude that
$$\lfloor a_1 \rfloor x_1 + \lfloor a_2 \rfloor x_2 + \cdots + \lfloor a_n \rfloor x_n \leq b.$$

However, in the inequality we just wrote down, the left-hand side is an integer. So if it is less than or equal to b, it is also less than or equal to $\lfloor b \rfloor$, giving us the inequality.

The difference between the two sides of the inequality is an integer simply because both sides are integers. □

It is often convenient to rewrite the inequality in this theorem by subtracting it from the equation we started with, getting

$$(a_1 - \lfloor a_1 \rfloor)x_1 + (a_2 - \lfloor a_2 \rfloor)x_2 + \cdots + (a_n - \lfloor a_n \rfloor)x_n \geq b - \lfloor b \rfloor.$$

Here, $a_i - \lfloor a_i \rfloor$ and $b - \lfloor b \rfloor$ are called the **fractional parts** of a_i and of b.

Be careful when taking fractional parts of negative numbers! If r is a positive real number, $r - \lfloor r \rfloor$ is just the part of r after the decimal, but this is no longer true if r is negative. For example, if $r = -1.23$, then its floor is $\lfloor r \rfloor = -2$, and its fractional part is $r - \lfloor r \rfloor = 0.77$.

There is a further detail we need to know: not just that this inequality is valid, but that it cuts off fractional optimal solutions. This will force the dual simplex method to find us a new solution, with a new opportunity for it to be an integer solution.

If the Gomory fractional cut is obtained from a row of our dictionary, then the fractional-part form of the inequality will contain only nonbasic variables: the single basic variable x_i will have $a_i = 1$ (since it appears on the left side with coefficient 1) and therefore $a_i - \lfloor a_i \rfloor = 0$. So the right-hand side of the inequality will be 0 at the current basic solution. Provided we pick a row where the constant term is a fraction, we will have $b - \lfloor b \rfloor > 0$, so the inequality will not hold!

2.2 An example

As an example, consider the following integer program:

$$
\begin{aligned}
\underset{x,y \in \mathbb{R}}{\text{maximize}} \quad & 3x + 2y \\
\text{subject to} \quad & 3x + y \le 6 \\
& y \le 2 \\
& x, y \ge 0
\end{aligned}
$$
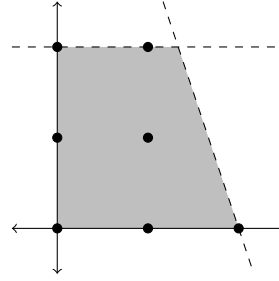
Before we can find a cutting plane, we should solve the linear programming relaxation.

$$
\begin{array}{c}
\max \zeta = 0 + 3x + 2y \\
\hline
w_1 = 6 - 3x - y \\
w_2 = 2 \qquad - y
\end{array}
\rightsquigarrow
\begin{array}{c}
\max \zeta = 4 - 2w_2 + 3x \\
\hline
w_1 = 4 + w_2 - 3x \\
y = 2 - w_2
\end{array}
\rightsquigarrow
\begin{array}{c}
\max \zeta = 8 - w_1 - w_2 \\
\hline
x = \frac{4}{3} - \frac{1}{3}w_1 + \frac{1}{3}w_2 \\
y = 2 \qquad - w_2
\end{array}
$$

If we did not find an integer solution, then by definition, one of the basic variables will have a fractional value in the optimal dictionary. In our case, x has a fractional value. Finding the Gomory fractional cut requires us to pick one such variable; if there are several, it doesn't matter which one we pick, but in this case, we can only pick x.

To apply Theorem 1 here, we need to write x's equation in the appropriate form. We move all the nonbasic variables to the left-hand side, getting

$$
x + \frac{1}{3}w_1 - \frac{1}{3}w_2 = \frac{4}{3}.
$$

The inequality in Theorem 1 is obtained by rounding every coefficient down to the nearest integer. This turns x into x, $\frac{1}{3}w_1$ into $0w_1$ or 0, $-\frac{1}{3}w_2$ into $-w_2$, and $\frac{4}{3}$ into 1, so we get the inequality

$$
x - w_2 \le 1.
$$

The reason we like to subtract this inequality from the previous inequality (or equivalently, take the fractional parts of the coefficients) is that this eliminates the basic variable x, giving us an inequality in the nonbasic variables:

$$\frac{1}{3}w_1 + \frac{2}{3}w_2 \geq \frac{1}{3}.$$

Our next step is to add this to our dictionary. We add a new variable w_3 representing the amount by which $\frac{1}{3}w_1 + \frac{2}{3}w_2$ exceeds $\frac{1}{3}$. Solving for w_3, this gives us the equation

$$w_3 = -\frac{1}{3} + \frac{1}{3}w_1 + \frac{2}{3}w_2.$$

The clause at the end of Theorem 1 is important here: it tells us that the new variable w_3 will be an integer at any integer solution to our program, so we will continue to have an integer program in which *all* variables are integers.

We can add the equation $w_3 = -\frac{1}{3} + \frac{1}{3}w_1 + \frac{2}{3}w_2$ as another row to our dictionary and solve the new linear program using the dual simplex method:

$$
\begin{array}{rl}
\max \zeta = & 8 - w_1 - w_2 \\
\hline
x = & \frac{4}{3} - \frac{1}{3}w_1 + \frac{1}{3}w_2 \\
y = & 2 \qquad\quad - w_2 \\
w_3 = & -\frac{1}{3} + \frac{1}{3}w_1 + \frac{2}{3}w_2
\end{array}
\qquad \rightsquigarrow \qquad
\begin{array}{rl}
\max \zeta = & \frac{15}{2} - \frac{1}{2}w_1 - \frac{3}{2}w_3 \\
\hline
x = & \frac{3}{2} - \frac{1}{2}w_1 + \frac{1}{2}w_3 \\
y = & \frac{3}{2} + \frac{1}{2}w_1 - \frac{3}{2}w_3 \\
w_2 = & \frac{1}{2} - \frac{1}{2}w_1 + \frac{3}{2}w_3
\end{array}
$$

(A brief synopsis of the dual simplex pivot we did: we know that w_3 is our leaving variable, and because both w_1 and w_2 have positive coefficients, they're both on our shortlist for entering variables. We compare ratios, and w_1's ratio $\frac{1}{1/3}$ is larger than w_2's ratio $\frac{1}{2/3}$, so w_2 is our entering variable. After pivoting, we end up with an optimal dictionary.)

We are still not done, because $(x, y) = (\frac{3}{2}, \frac{3}{2})$ is not a integer solution. But let's take a break from that to look at what is going on graphically.

To visualize the procedure of adding a cutting plane, we can rewrite our inequality $x - w_2 \leq 1$ in yet a third form: in terms of x and y. To do this, substitute $w_2 = 2 - y$, getting $x - (2 - y) \leq 1$ or $x + y \leq 3$. In the diagram of the feasible region, here is what this looks like:
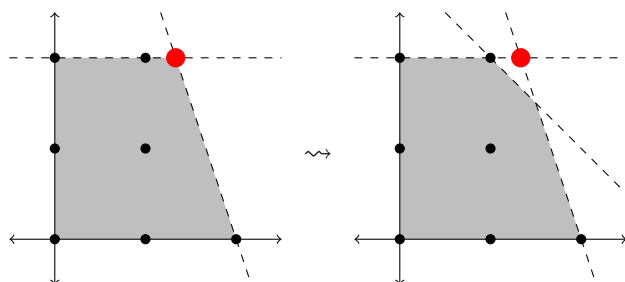
The newly added inequality separates all the integer solutions (in black) from the fractional solution to the LP relaxation (the large point in red). Unfortunately, the optimal solution to the new feasible region happens to be the *one* corner of that region that doesn't have integer coordinates. We have the worst luck!

To continue, we go back to our dictionary. All three basic variables have a fractional value, so we could pick any of them to deal with, but let's pick x again. (In this example, it turns out that we'll get the same cut no matter which equation we get it from.)

Moving all the variables to the left, we get $x + \frac{1}{2}w_1 - \frac{1}{2}w_3 = \frac{3}{2}$. Taking the fractional parts according to the alternate form of Theorem 1, we get the inequality $\frac{1}{2}w_1 + \frac{1}{2}w_3 \geq \frac{1}{2}$. We can add this to our dictionary, with a new slack variable w_4, if we define $w_4 = -\frac{1}{2} + \frac{1}{2}w_1 + \frac{1}{2}w_3$.

Applying the dual simplex method again takes us only one pivot step:

$$
\begin{array}{rl}
\max \zeta = & \frac{15}{2} - \frac{1}{2}w_1 - \frac{3}{2}w_3 \\
\hline
x = & \frac{3}{2} - \frac{1}{2}w_1 + \frac{1}{2}w_3 \\
y = & \frac{3}{2} + \frac{1}{2}w_1 - \frac{3}{2}w_3 \\
w_2 = & \frac{1}{2} - \frac{1}{2}w_1 + \frac{3}{2}w_3 \\
w_4 = & -\frac{1}{2} + \frac{1}{2}w_1 + \frac{1}{2}w_3
\end{array}
\qquad \leadsto \qquad
\begin{array}{rl}
\max \zeta = 7 - & w_3 - w_4 \\
\hline
x = 1 + & w_3 - w_4 \\
y = 2 - & 2w_3 + w_4 \\
w_2 = 0 + & 2w_3 - w_4 \\
w_1 = 1 - & w_3 + 2w_4
\end{array}
$$

The optimal solution is $(x, y) = (1, 2)$, which is an integer solution! After the second cutting plane, we are done.

The second cut we added, which we wrote as $\frac{1}{2}w_1 + \frac{1}{2}w_3{}' \geq \frac{1}{2}$, can be written in terms of x and y as $2x + y \leq 4$. Here is another diagram showing the evolution of our feasible region as we add the cutting planes:
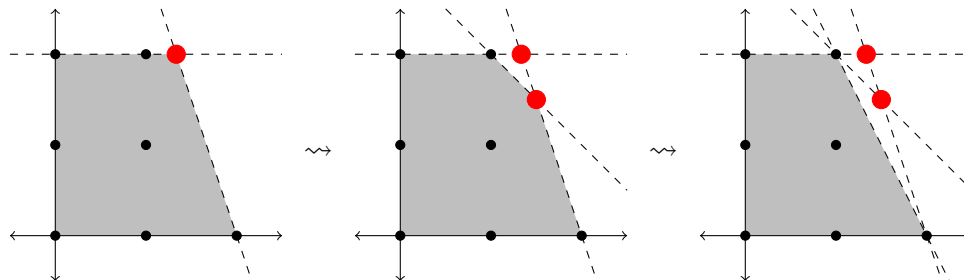
3   Extensions

The cutting plane method is often combined with the branch-and-bound method into a hybrid algorithm called "branch-and-cut". Here, when solving a linear program and getting a fractional solution, we make a choice between two options:

- Pick a variable x_i with a fractional value, and use it to branch out to two new linear programs, as usual in the branch-and-bound method.

- Add a cutting plane inequality to replace the linear program by a new one with a different solution.

It's a matter of heuristics (in other words, guesswork) to decide between these two options. These heuristics are only partially developed by mathematical reasoning; partially, we just check them on practical examples to see how well they behave.