Improving the LP bound of a MILP by dual concurrent branching and the relationship to cut generation methods

H. Georg Büsching
Händelstraße 9
32457 Porta Westfalica
Germany
lpbuesching@googlemail.com

June 30, 2013

Abstract

In this paper branching for attacking MILP is investigated. Under certain circumstances branches can be done concurrently. By introducing a new calculus it is shown there are restrictions for dual values. As a second result of this study a new class of cuts for MILP is found, which are defined by those values. This class is a superclass of all other classes of cuts. Furthermore the restrictions of the dual values can be used for studying the addition of arbitrary inequalities. This theory has similarities but also big differences to the theory of disjunctive programming.

1 Introduction

1.1 Advice for reading

As the first half of the paper uses non-standard calculations in the dual space of linear programming, some readers might not see the consequences of the thoughts. To have an idea of the aim, the author suggests to read directly 6.1 after this chapter and here especially the example, which should be understood in detail. Then it is time to attack the very heart of the article at 2, where the dual theory is developed from scratch.

1.2 History of the article

More than a decade ago the author was working at his first job having a degree in pure mathematics from the university. He had too much time at this job, so he was writing some heuristic programs for some combinatorial problems, which was mainly the TSP. Studying the backgrounds of better heuristics, he read some state of the art articles like [6]. Hearing from the world TSP, he just guessed that doing the branch-and-cut work when trying to prove the optimality of a tour there should be some improvement possible. In particular having a non-integer value on two variables representing one road in Germany and one in Australia it must be possible to do this branch concurrently, as the roads are too far away from each other. With this fixed idea he decided to understand the basics of linear programming. It came out that something similar is possible, but only in the dual space (2.5). The generalization to MILP was a natural step. He implemented it as described in 3. But he never found anything in all research papers where something was said about dual concurrent branching. He also understood that a good LP-solver would be quickly able to get non-optimal dual values, which would fit much better together. By making experiments with transforming 2.5 again into normal space he found the theorem 5.4. There cuts are created, which is reasoned by a pretty complex dual argumentation. To prove the theorem, that all cuts based on a case analysis should be recovered by this theory, the author found out that it must possible to do this in two steps via 6.1 and 6.3. Both theorems were proved with some effort, using a very natural technique for 6.1, which the author has also never found anywhere else in the literature. Trying to classify the new method by known methods, the author investigated the difference to the theory of disjunctive programming.

Encouraged by Mr. Balas and Mr. Cook he finally consolidated all results. He uses now one notation which is only non-standard in a few aspects. As the ideas had evolved in the course of the time the notations were not stable in earlier version from this article.

Other ideas during the study of the TSP were an unifying theory for heuristics and LP relaxation and cut generation methods and an idea how to separate a MILP into two real subproblems. There is nothing more in the paper about these topics than the last sentence.

1.3 Terminology

We only use partially the normal representations for LPs: $Ax \geq b$ with $\tau(x) = cx$, which has to minimized. In the following we study LPs on variables x_n and the inequalities on such a LP shall be called f_m. The function $f_m(x) = \sum a_{n,m}x_n - b_n$ represents the m-th inequality. Some of the inequalities might be equalities. The problem is well-defined, meaning that at least one optimal solution exists, only $\geq$-inequalities are used. Furthermore all x_n should positive, and the inequalities $x_n \geq 0$ are within the set of the $\{f_m\}$. Without loss of generality we consider only minimization problems. All theorems, which are written down can be used by simple transformations to such problems. The authors considers thus transformations as so simple that he will not even bother with them. We will later operate with a maximization problem as an example, which only has $\leq$ inequalities.

The biggest part of the paper are investigations in the dual space. We let y_m be the dual variable assigned to the inequality f_m, the dual inequalities themselves shall be named g_n. The objective in dual space is just $\omega(y)$, depending on the vector y. As the problem is well defined, also the dual problem has at least one optimal solution. It is also demanded that all dual inequalities g_n are seen as equalities. So you have a linear combination of the $\tau(x)$ by the $f_m(x)$:

$$\tau(x) = \sum_m y_m f_m(x) + \omega(y)$$

This demand could be seen from the point of literature as non-standard but understandable. As we have added the $x_n \geq 0$ to the set of $\{f_m\}$ this is possible. The index m will always go over all inequalities (dual variables), which are in the starting LP. Additional dimensions will not be included in such sums, they will have an index a or c. As announced we study analysis of cases concurrently. One analysis of cases shall be represented by superscript i, an underlying case of such a case distinction shall have a superscript j.

Seen the whole problem as a MILP or Disjunctive LP, we also let all the more restrictive LP, where a case of a case distinction is represented be well-defined, meaning that an optimal solution exists. Cases without solution are never considered in the theoretical part.

Especially for this paper we create two new objects: The first object is a branch $\mathfrak{b}^{i,j}$ which is defined by looking at the reduction of the dual variables from one dual solution of the starting problem and the dual problem of the more restrictive problem of a certain case j of a case distinction i. The space of all such points shall be named the branching space $\mathfrak{B}^{i,j}$.

Last we also consider the vector $\mathfrak{k}^i \in \mathfrak{K}^i$ of all cases for a case distinction which has the form $(\mathfrak{b}^{i,1}, \mathfrak{b}^{i,2}, \ldots)$. This shall be called knot with a reference to the botanical meaning of the word knot. For each branch we'll define an objective ω and for each knot we'll define a function Δ and also an objective ω. The definition of that functions will be natural in those spaces. We'll use those functions with leaving out the index of source space.

We could avoid these last definitions as like in the general theorem in 9.1, which extends the results of this paper as far as possible. Anyhow for studying the efficiency of concurrent branching and thereby the efficiency of cuts exactly the vectors of branches and the knots should be investigated. As they are that central we decided to give these objects names.

2 Description of dual concurrent branching

2.1 Central theorem

In the following we examine a problem P, which can be partially represented as a minimal linear problem P^0. The description of the entire problem needs some additional analysis of cases. It should be remarked that the set of MILP-instances is a real subset of this problem class. The dual solution space of P^0 will be noted as V^0. Furthermore we chose an arbitrary $y^0 \in V^0$. By looking at one case j of a case distinction i the we call the dual solution space $V^{i,j}$.

Now let the branch $\mathfrak{B}^{i,j}$ be the set $\{\mathfrak{b}^{i,j} |\ y^0 - \mathfrak{b}^{i,j} \in V^{i,j}\}$. So $\mathfrak{B}^{i,j}$ is a linear tranformation of $V^{i,j}$ with the fixed starting point y^0 as the new zero vector. Our objective function ω can easily be expanded to $\mathfrak{b}^{i,j}$ in an almost natural way just by setting it to $\omega(\mathfrak{b}^{i,j}) = \omega(y^0 - \mathfrak{b}^{i,j}) - \omega(y^0)$.

On the space $\mathfrak{B}^{i,j}$ we have both an additive structure and the multiplication with a scalar, so it is a subspace of a vector space. During those operations we might leave $V^{i,j}$ as some associated dual values might become negative. In the later subsection 7.2 we'll see that such temporarily forbidden results should not be avoided.

We make a break at this point and use an easy example to illustrate the concept:

$$f_1 : x_1 + x_2 \geq 1 \text{ and } f_2 : x_2 + x_3 \geq 1 \text{ and } f_3 : x_3 + x_1 \geq 1$$

$$f_4 : x_4 + x_5 \geq 1 \text{ and } f_5 : x_5 + x_6 \geq 1 \text{ and } f_6 : x_6 + x_4 \geq 1$$

Minimize $\tau(x) = x_1 + x_2 + x_3 + x_4 + x_5 + x_6$ and all variables have to be integer. All variables are positive, so we write also this bounds down.

$$f_7 : x_1 \geq 0 \text{ and } f_8 : x_2 \geq 0 \text{ and } f_9 : x_3 \geq 0$$

$$f_{10} : x_4 \geq 0 \text{ and } f_{11} : x_5 \geq 0 \text{ and } f_{12} : x_6 \geq 0$$

Just by viewing at it you immediately see that the optimal solution is $x_1 = x_2 = x_3 = x_4 = x_5 = x_6 = \frac{1}{2}$ with objective 3. But when integers are required the objective has to be at least 4. We'll consider the linear functions $f_m(x)$ as announced in the previous section. As an example $f_1(x)$ is defined by $x_1 + x_2 - 1$. Transformed to the dual space in the textbook standard way, we get:

$$g_1 : y_1 + y_3 \leq 1 \text{ and } g_2 : y_1 + y_2 \leq 1 \text{ and } g_3 : y_2 + y_3 \leq 1$$

$$g_4 : y_4 + y_6 \leq 1 \text{ and } g_5 : y_4 + y_5 \leq 1 \text{ and } g_6 : y_5 + y_6 \leq 1$$

Maximize $\tau(x) = y_1 + y_2 + y_3 + y_4 + y_5 + y_6$.

Anyhow we use the dual problem written down in another way, where $\tau(x)$ has a linear representation by the $f_1(x), \ldots f_{12}(x)$:

$$g_1 : y_1 + y_3 + y_7 = 1 \text{ and } g_2 : y_1 + y_2 + y_8 = 1 \text{ and } g_3 : y_2 + y_3 + y_9 = 1$$

$$g_4 : y_4 + y_6 + y_{10} = 1 \text{ and } g_5 : y_4 + y_5 + y_{11} = 1 \text{ and } g_6 : y_5 + y_6 + y_{12} = 1$$

The optimal solution y^0 is $y_1 = y_2 = y_3 = y_4 = y_5 = y_6 = \frac{1}{2}$ and all other coordinates are 0. We now consider the cases $f_a^{1,1} : x_1 \leq 0$ or $f_a^{1,2} : x_1 \geq 1$. For the first case we get the following dual inequalities, which describes $V^{1,1}$:

$$g_1 : y_1 + y_3 - y_a^{1,1} + y_7 = 1 \text{ and } g_2 : y_2 + y_3 + y_8 = 1 \text{ and } g_3 : y_3 + y_1 + y_9 = 1$$

$$g_4 : y_4 + y_5 + y_{10} = 1 \text{ and } g_5 : y_5 + y_6 + y_{11} = 1 \text{ and } g_6 : y_6 + y_4 + y_{12} = 1$$

The optimal solution $y^{1,1}$ of the first dual subproblem is $y_a^{1,1} = -1$ and $y_1 = y_3 = 1$ and $y_2 = 0$ and $y_4 = y_5 = y_6 = \frac{1}{2}$. The value of the vector for the bounds for this dual point is still 0. We see that the value of vector $\mathfrak{b}^{1,1}$

is $(-\frac{1}{2}, \frac{1}{2}, -\frac{1}{2}, 0, 0, 0, 0, 0, 0, 0, 0, -1)$, where the last coordinate represents the additional dual variable and $\omega(\mathfrak{b}^{1,1}) = \omega(y^{1,1}) - \omega(y^0) = \frac{7}{2} - 3 = \frac{1}{2}$.

After this illustration we continue with the definition of the basic concept: As $y^0 - \mathfrak{b}^{i,j}$ is a dual solution, all coordinates y_m most be positive or null, in case f_m is not an equality.

Remark 2.1 *For those m we get that $\mathfrak{b}^{i,j}_m \leq y^0_m$ holds, which already has the structure of our main statement 2.5.*

In 2.1 we call those inequalities, where the right hand side is greater 0 the main inequalities.

Remark 2.2 *If $\mathfrak{b}^{i,j}$ relates to an optimal solution of $V^{i,j}$, then one main inequality is sharp.*

To see this we assume the contrary. We introduce ϵ as a very small and positive number. We consider $y - (1+\epsilon)\mathfrak{b}^{i,j}$, which has higher objective value and the inequalities in 2.1 still hold. We conclude that $(1+\epsilon)\mathfrak{b}^{i,j}$ is in $\mathfrak{B}^{i,j}$. This is a contradiction to the optimality of $\mathfrak{b}^{i,j}$. So one of the main inequalities must be sharp.

Remark 2.3 *Furthermore we state that for each n, $\sum_m a_{n,m}\mathfrak{b}^{i,j}_m = 0$.*

This follows from the two linear representations of $\tau(x) = y^0 f = (y^0 - \mathfrak{b}^{i,j})f$. The next step of our thoughts is to go from a case to one analysis of cases. The corresponding object will be named as a knot. Its space is defined by:

$$\mathfrak{K}^i = \bigoplus_j \mathfrak{B}^{i,j}$$

An element out of our construct $\mathfrak{K}^i$ is defined by $\mathfrak{k}^i = \oplus_j \mathfrak{b}^{i,j}$. We further define:

$$\omega(\mathfrak{k}^i) = \min_j \omega(\mathfrak{b}^{i,j})$$

$$\Delta_m(\mathfrak{k}^i) = \max_j \mathfrak{b}^{i,j}_m$$

The delta represents the largest reduction for each dual dimension for a whole analysis of cases (knot).
For the case of $x_1 \geq 1$ in our easy example we again choose the optimal point, which is: $y^{1,2}_a = 1$ and $y_1 = y_3 = 0$ and $y_2 = 1$ and $y_4 = y_5 = y_6 = \frac{1}{2}$ and all other variables are 0.
So we get for $\mathfrak{b}^{1,2} = (\frac{1}{2}, -\frac{1}{2}, \frac{1}{2}, 0, 0, 0, 0, 0, 0, 0, 0, -1)$ with objective $\frac{1}{2}$. This leads to:

$$\mathfrak{k}^1 = ((-\tfrac{1}{2}, \tfrac{1}{2}, -\tfrac{1}{2}, 0, 0, 0, 0, 0, 0, 0, 0, -1)(\tfrac{1}{2}, -\tfrac{1}{2}, \tfrac{1}{2}, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1))$$
$$\Delta(\mathfrak{k}^1) = (\tfrac{1}{2}, \tfrac{1}{2}, \tfrac{1}{2}, 0, 0, 0, 0, 0, 0, 0, 0, 0)$$
$$\omega(\mathfrak{k}^1) = \tfrac{1}{2}$$

Another knot $\mathfrak{k}^2$ could be for $x_2 \leq 0$ or $x_2 \geq 1$, when choosing again in both cases the optimal point you again get:

$$\Delta(\mathfrak{k}^2) = (\tfrac{1}{2}, \tfrac{1}{2}, \tfrac{1}{2}, 0, 0, 0, 0, 0, 0, 0, 0, 0)$$
$$\omega(\mathfrak{k}^2) = \tfrac{1}{2}$$

Considering the knot $\mathfrak{k}^3$ as the branches $x_4 \leq 0$ or $x_4 \geq 1$, we pick up again the optimal points for each solution and receive:

$$\Delta(\mathfrak{k}^3) = (0, 0, 0, \tfrac{1}{2}, \tfrac{1}{2}, \tfrac{1}{2}, 0, 0, 0, 0, 0, 0)$$
$$\omega(\mathfrak{k}^3) = \tfrac{1}{2}$$

Let $\mathfrak{k}^i, \mathfrak{l}^i \in \mathfrak{K}^i$. We have the following equalities and inequalities.

Remark 2.4

$$\begin{array}{lcl} \Delta_m(\lambda\mathfrak{k}^i) & = & \lambda\Delta_m(\mathfrak{k}^i) \\ \Delta_m(\mathfrak{k}^i + \mathfrak{l}^i) & \leq & \Delta_m(\mathfrak{k}^i) + \Delta_m(\mathfrak{l}^i) \\ \omega(\lambda\mathfrak{k}^i) & = & \lambda\omega(\mathfrak{k}^i) \\ \omega(\mathfrak{k}^i + \mathfrak{l}^i) & \geq & \omega(\mathfrak{k}^i) + \omega(\mathfrak{l}^i) \end{array}$$

We conclude that the deltas are still convex and ω is concave.

Now we build an even more complex space $\mathfrak{T}$, which will be the sum of all knots and our final object. This space represents parallel knots. If needed you could name this object "tree".

$$\mathfrak{T} = \bigoplus_i \mathfrak{K}^i$$

We have the following functions for $\mathfrak{t} \in \mathfrak{T}$:

$$\Delta_m(\mathfrak{t}) = \sum_i \Delta_m(\mathfrak{k}^i)$$

$$\omega(\mathfrak{t}) = \omega(y^0) + \sum_i \omega(\mathfrak{k}^i)$$

Now let $\mathfrak{t} \in \mathfrak{T}$, then we have chosen in all cases of all knots a solution vector. Remember that we are always talking about dual solutions and variables. If we choose for each case distinction i a case $j(i)$, then we have a new problem $P^{\{j(i)\}}$, which is in fact P^0 together with the inequalities from all $P^{i,j(i)}$. Looking at the dual solution space of this problem we define a vector $\hat{y}$ by the means of the $\mathfrak{b}^{i,j}$:

$$\hat{y} = y^0 - \sum_i \mathfrak{b}^{i,j(i)}$$

As we have seen in 2.3 that for each i and j the equality $\mathfrak{b}^{i,j} A_{i,j} = 0$ holds, we see that the m-th dual equality is fulfilled. Anyhow the condition that this dual solution is valid, is that for all inequalities the associated variable is positive.

$$\hat{y}_m = y^0_m - \sum_i \mathfrak{b}^{i,j(i)}_m \geq y^0_m - \sum_i \Delta_m(\mathfrak{k}^i) = y^0_m - \Delta_m(\mathfrak{t})$$

So by $\Delta_m(\mathfrak{t}) \leq y^0_m$ we can make sure, that the coordinates of $\hat{y}$ are positive. Notice that the condition is independent of our choice $j(i)$. The objective value of $\hat{y}$ is $\omega(y^0) + \sum_i \omega(\mathfrak{b}^{i,j}) \geq \omega(\mathfrak{t})$. Putting these thoughts together we get our central statement.

Theorem 2.5 (Central theorem) *If $\Delta_m(\mathfrak{t}) \leq y^0_m$ holds for all m, then the original problem P must have an optimal solution that is greater than $\omega(\mathfrak{t})$.*

Via this theorem we see in our example that the vector $(\mathfrak{k}^1, \mathfrak{k}^3)$ gives a lower limit of 4. It is clear that it can't be applied to $(\mathfrak{k}^1, \mathfrak{k}^2)$. An analysis of the structure of our problem shows, that it consists of two pieces of a very trivial MILP. When the two branches like 1 and 3 are made in different subproblems, they can be combined.

2.2 Building the little combining LP

The last statement seems to be rather abstract, but by an easy trick, we'll get a different form, that can be used in algorithms. We just exchange $\mathfrak{k}^{i,j}$ by $\lambda_i \mathfrak{k}^i$ with $\lambda_i \geq 0$. As the deltas and the objectives ω_i are linear on a scalar (2.4), we get the main result of this section:

Theorem 2.6 (Central theorem on concurrent branching) *The ability to combine case distinctions can be assured by the following inequalities:*

$$\sum_i \lambda_i \Delta_m(\mathfrak{k}^i) \leq y^0_m \ \forall m$$

The new lower limit is $\sum_i \lambda_i \omega(\mathfrak{k}^i) + \omega(y^0)$, which might be seen as the objective on a linear program on the λ_i.

So by solving this LP-instance in λ_i we get a better lower limit for our problem P. By looking at all $P^{i,j}$ all values in this LP-instance can be calculated, first $y^0_m - \mathfrak{b}^{i,j}_m$, secondly $\omega(\mathfrak{b}^{i,j})$ and lastly $\Delta_m(\mathfrak{k}^i)$ and $\omega(\mathfrak{k}^i)$. For another good example of the usage of this theorem we refer to 4.

By the definition of ω for some i you can choose the $\mathfrak{b}^{i,j}$ in such a way that for all j the equation $\omega(\mathfrak{k}^i) = \omega(\mathfrak{b}^{i,j})$ holds. This can be achieved by substituting $\mathfrak{b}^{i,j}$ with $\omega(\mathfrak{k}^i)\omega(\mathfrak{b}^{i,j})^{-1}\mathfrak{b}^{i,j}$. When $\omega(\mathfrak{b}^{i,j}) = 0$ then we get no progress on the objective function of P from this case distinction, so it is not an interesting branch. By this substitution in 2.6, the objective function $\omega_i(\mathfrak{k}^i)$ remains the same, but normally the Δ-values will decrease, leading to higher values. We call this *normalization*.

We seem to have presented in this section a theory on a problem, which can only partly described by a LP. But in truth we studied one LP, the dual, where the problem is described too sharp and can be weakened by case distinctions. Looking at the dual combining LP of 2.6, for a given i, you could weaken the LP for different j. Via this you again make a new case distinctions, where you can again try to make case distinctions in parallel as in the presented theory. We will sketch one manual example later at 4.

Although the mathematical formulation to combine case distinctions (knots) has been explained broadly in this section, some details are still not covered. The problem is that those details might be not too easy to attack at all. When you think of a fast implementation of this idea you want to have an effective, numerically stable and fast algorithm to find good elements $\mathfrak{b}^{i,j} \in \mathfrak{B}_{i,j}$, where most $\mathfrak{b}^{i,j}$ are zero. From these you get good $\mathfrak{k}^i$ for a given solution y^0. We'll see later in 3 that the normal approach to use optimal solutions of the $P^{i,j}$ leads in some examples to problems. So later in 8.1 and 8.2 we will attack these problems by using non-optimal $y^0 - \mathfrak{b}^{i,j}$ even before normalization, where most $\mathfrak{b}^{i,j}_m$ should be zero.

2.3 Rough description of the huge combining LP

In this section we will follow again the made definitions and results. We had started with one solution $y^0 \in V_0$. For each branch of a knot we have also a $y^{i,j} \in V^{i,j}$, so $\mathfrak{b}^{i,j}$ is defined in all coordinates in inequalities.
Via the restrictions $\Delta_m(\mathfrak{k}^i) \geq -\mathfrak{b}^{i,j}_m$ and $\omega(\mathfrak{k}^i) \leq \omega(\mathfrak{b}^{i,j})$ for all j we have defined the $\Delta(\mathfrak{k}^i)$ and $\omega(\mathfrak{k}^i)$ as linear inequalities. Like before we define $\Delta_m(\mathfrak{t})$ as a sum of the $\Delta_m(\mathfrak{k}^i)$ and $\omega(\mathfrak{t})$ as the sum of the $\omega(\mathfrak{k}^i)$ plus the objective $\omega(y^0)$, which is also a linear term of y^0. Using also the restrictions of 2.5 we have defined now a huge LP $P_\mathfrak{t}$. We can now postulate a theorem, which describes the problem of doing case distinctions in parallel in a mathematically satisfactory way:

Theorem 2.7 (Central theorem - complete form) *Each solution of $P_\mathfrak{t}$ represents a lower bound of P.*

The optimal value is the optimal lower bound possible via our combining technique. It should be noticed that the definitions in 2.6 and 5.1 are more tightened inequalities systems, which are also less complex in comparison with this huge inequality system.

In the last section 9.1 of this paper we will really write down the complex LP in detail and discuss its properties. In the author eyes it is playing a central role for studying MILP in general.

3 Implementation with usage of optimal solutions of the subproblems

Putting the thoughts from the previous section together you get the following algorithm described as pseudo-code to get higher objective values of a binary LP-instance:

1: derive P^0 from given MILP-instance
2: load P^0 into LP-solver
3: solve P^0 and save one optimal solution x and the fitting dual solution y^0
4: **for** all i, where x_i is not integer **do**
5: case $j = 1$:
6: add inequality $x_i \leq 0$ to P^0 (so getting $P^{i,1}$)
7: solve this LP by usage of y^0 as start, get $y^0 - \mathfrak{b}^{i,1}$ as dual solution
8: case $j = 2$:
9: add inequality $-x_i \leq -1$ to P^0 (so getting $P^{i,2}$)
10: solve this LP by usage of y^0 as start, get $y^0 - \mathfrak{b}^{i,2}$ as dual solution
11: $\omega(\mathfrak{k}^i) = \min(\omega(\mathfrak{b}^{i,1}), \omega(\mathfrak{b}^{i,2}))$
12: **if** $\omega(\mathfrak{k}^i) > 0$ **then**
13: mark i
14: **if** $\omega(\mathfrak{b}^{i,1}) \leq \omega(\mathfrak{b}^{i,2})$ **then**
15: for all m: $\mathfrak{b}^{i,2}_m = \omega(\mathfrak{b}^{i,1})(\omega(\mathfrak{b}^{i,2}))^{-1}\mathfrak{b}^{i,2}_m$
16: **else**
17: for all m: $\mathfrak{b}^{i,1}_m = \omega(\mathfrak{b}^{i,2})(\omega(\mathfrak{b}^{i,1}))^{-1}\mathfrak{b}^{i,1}_m$
18: **end if**
19: for all m: $\Delta_m(\mathfrak{k}^i) = \max(\mathfrak{b}^{i,1}_m, \mathfrak{b}^{i,2}_m)$
20: **end if**
21: **end for**
22: build new LP R with all marked i in λ_i as described in 2.6
23: solve R

In his implementation the author was not able to use the old solution in lines 7 and 10 effectively. This has some impact because of the structure of the optimal dual solution in most of the prominent problems.

To understand this let's consider you have chosen an optimal y^0 with $y^0_m > 0$ for some m. But also another optimal $\bar{y}^0$ exists with $\bar{y}^0_m = 0$. Now for all i one j could exist where $y^0_m - \mathfrak{b}^{i,j}_m = 0$. Taking y^0 as starting point leads to the situation that never two knots can be combined, ensured by 2.5. But if you had chosen the other $\bar{y}^0$ as starting point, you would have less problems. As a side-remark it should be noticed that if P^0 has a multiple (so non-trivial) optimal dual solution space this normally holds also for the $P^{i,j}$.

The above algorithm has been implemented with the Open-Source package glpk. In this program all MILP are transformed to be Minimum-problems by exchanging the sign of the objective. The problem library MIPLIB2003 [4] has been processed partially getting the results on the following page.

In the given table the column *Branches* measures the number of variables, where branching took place. The actual number of calculated LPs is 2 times more plus the initial LP and the combining LP. The column *Degree* is equal to $\sum_i \lambda_i$ of the optimal value of the combining LP. It gives an idea how much knots can be used at the same time but also in an effective way. Furthermore also the total time in seconds for the calculation of all $P^{i,j}$ is presented.

The given table only includes those instances, where the program finished within 1 hour. Also only those instances were included where at least one case distinction increased the objective. On the other hand for the instance tr12-30 a quite high lower bound was reached: Starting from 14210 the bound 79695 is reached where the real value is 130596. Also the running times were on an old version of glpk with an computer which is out of date, so the running time should not be compared too much to other experiments.

Instance	Pure LP	Bound Inc	Branches	Degree	Total
a1c1s1	997.53	1195.33	173	53.92	24
aflow30a	983.17	14.28	31	5.19	2
aflow40b	1005.66	7.16	38	1.80	17
air04	55535.44	84.61	292	1.00	1040
air05	25877.61	72.54	223	1.00	329
arki001	7579599.81	126.65	81	6.63	12
danoint	62.64	0.05	34	1.00	3
fiber	156082.52	15734.31	47	6.90	3
fixnet6	1200.88	210.51	60	21.33	2
gesa2	25476489.68	81043.25	58	35.91	5
gesa2-o	25476489.68	81891.56	73	36.70	4
liu	346.00	214.00	536	1.00	16
mas74	10482.80	42.52	12	1.19	1
mas76	38893.90	24.86	11	1.62	0
modglob	20430947.62	69955.22	29	8.31	0
net12	17.25	11.40	429	1.30	3115
p2756	2688.75	10.20	30	2.00	4
pp08a	2748.35	762.82	51	11.41	0
pp08aCUTS	5480.61	166.85	46	6.47	1
roll3000	11097.13	5.44	214	4.32	36
rout	981.86	2.34	35	1.00	1
set1ch	32007.73	3904.90	138	64.56	2
seymour	403.85	1.50	632	3.30	291
sp97ar	652560391.11	241502.97	194	2.00	522
swath	334.50	0.40	45	5.71	19
timtab1	28694.00	137970.93	136	16.46	1
timtab2	83592.00	106311.17	233	27.02	4
tr12-30	14210.43	65484.48	348	322.01	8
vpm2	9.89	0.48	31	7.41	1

4 Another example of the presented technology

$$\begin{aligned} f_1: &\quad x_1 + x_2 + x_3 \le 1 \\ f_2: &\quad x_2 + x_3 + x_4 \le 1 \\ f_3: &\quad x_3 + x_4 + x_5 \le 1 \\ f_4: &\quad x_4 + x_5 + x_1 \le 1 \\ f_5: &\quad x_5 + x_1 + x_2 \le 1 \end{aligned}$$

Maximize $\tau(x) = x_1 + x_2 + x_3 + x_4 + x_5$ and all variables have to be integer.

The optimal dual solution of the LP is simply $y_m^0 = \frac{1}{3}$ with $\omega = \frac{5}{3}$ and $m \in \{1; \ldots; 5\}$, all dual variables representing the trivial inequalities (reduced costs) are 0. The optimal solution of the MILP has objective of 1. As an example for our technique we branch on the two cases $x_1 = 0$ and $x_1 \ge 1$. We get then the following dual solutions leaving out the other dual variables which are still 0:

$$\begin{array}{lllllll} x_1 = 0: & y = & (0, & \frac{1}{2}, & \frac{1}{2}, & 0, & \frac{1}{2}) \text{ with } \omega = \frac{3}{2} \\ x_1 = 1: & y = & (1, & 0, & 0, & 1, & 0) \text{ with } \omega = 1 \\ x_1 = 1: & y = & (\frac{1}{2}, & \frac{1}{4}, & \frac{1}{4}, & \frac{1}{2}, & \frac{1}{4}) \text{ with } \omega = \frac{3}{2} \text{ (Normalization)} \end{array}$$

$$\Rightarrow \Delta(\mathfrak{k}^1) = (\tfrac{1}{3}, \tfrac{1}{12}, \tfrac{1}{12}, \tfrac{1}{3}, \tfrac{1}{12});\ \text{with } \omega = \tfrac{1}{6}$$

Via using the symmetry of the problem we get for the combination of the knots the following LP:

$$\begin{array}{ccccccccccc} \frac{1}{3}\lambda_1 & + & \frac{1}{12}\lambda_2 & + & \frac{1}{3}\lambda_3 & + & \frac{1}{12}\lambda_4 & + & \frac{1}{12}\lambda_5 & \le & \frac{1}{3} \\ \frac{1}{12}\lambda_1 & + & \frac{1}{3}\lambda_2 & + & \frac{1}{12}\lambda_3 & + & \frac{1}{3}\lambda_4 & + & \frac{1}{12}\lambda_5 & \le & \frac{1}{3} \\ \frac{1}{12}\lambda_1 & + & \frac{1}{12}\lambda_2 & + & \frac{1}{3}\lambda_3 & + & \frac{1}{12}\lambda_4 & + & \frac{1}{3}\lambda_5 & \le & \frac{1}{3} \\ \frac{1}{3}\lambda_1 & + & \frac{1}{12}\lambda_2 & + & \frac{1}{12}\lambda_3 & + & \frac{1}{3}\lambda_4 & + & \frac{1}{12}\lambda_5 & \le & \frac{1}{3} \\ \frac{1}{12}\lambda_1 & + & \frac{1}{3}\lambda_2 & + & \frac{1}{12}\lambda_3 & + & \frac{1}{12}\lambda_4 & + & \frac{1}{3}\lambda_5 & \le & \frac{1}{3} \end{array}$$

The objective is $\frac{1}{6}\lambda_1 + \frac{1}{6}\lambda_2 + \frac{1}{6}\lambda_3 + \frac{1}{6}\lambda_4 + \frac{1}{6}\lambda_5$.

The optimal solution is $\lambda_i = \frac{4}{11}$ and objective is $\frac{10}{33}$. So we have shown that the maximum in our original MILP is less or equal $\frac{5}{3} - \frac{10}{33}$.

At this point it is again possible to make a case distinction on $x_1 = 0$ or $x_1 \ge 1$. If we assume $x_1 = 0$ the above LP would have the following form:

$$\begin{array}{ccccccccccc} \frac{1}{3}\lambda_1 & + & \frac{1}{12}\lambda_2 & + & \frac{1}{3}\lambda_3 & + & \frac{1}{12}\lambda_4 & + & \frac{1}{12}\lambda_5 & \le & \frac{1}{3} \\ -\frac{1}{6}\lambda_1 & + & \frac{1}{3}\lambda_2 & + & \frac{1}{12}\lambda_3 & + & \frac{1}{3}\lambda_4 & + & \frac{1}{12}\lambda_5 & \le & \frac{1}{3} \\ -\frac{1}{6}\lambda_1 & + & \frac{1}{12}\lambda_2 & + & \frac{1}{3}\lambda_3 & + & \frac{1}{12}\lambda_4 & + & \frac{1}{3}\lambda_5 & \le & \frac{1}{3} \\ \frac{1}{3}\lambda_1 & + & \frac{1}{12}\lambda_2 & + & \frac{1}{12}\lambda_3 & + & \frac{1}{3}\lambda_4 & + & \frac{1}{12}\lambda_5 & \le & \frac{1}{3} \\ -\frac{1}{6}\lambda_1 & + & \frac{1}{3}\lambda_2 & + & \frac{1}{12}\lambda_3 & + & \frac{1}{12}\lambda_4 & + & \frac{1}{3}\lambda_5 & \le & \frac{1}{3} \end{array}$$

And for $x_1 \ge 1$:

$$\begin{array}{ccccccccccc} -\frac{2}{3}\lambda_1 & + & \frac{1}{12}\lambda_2 & + & \frac{1}{3}\lambda_3 & + & \frac{1}{12}\lambda_4 & + & \frac{1}{12}\lambda_5 & \le & \frac{1}{3} \\ \frac{1}{3}\lambda_1 & + & \frac{1}{3}\lambda_2 & + & \frac{1}{12}\lambda_3 & + & \frac{1}{3}\lambda_4 & + & \frac{1}{12}\lambda_5 & \le & \frac{1}{3} \\ -\frac{2}{3}\lambda_1 & + & \frac{1}{12}\lambda_2 & + & \frac{1}{3}\lambda_3 & + & \frac{1}{12}\lambda_4 & + & \frac{1}{3}\lambda_5 & \le & \frac{1}{3} \\ \frac{1}{3}\lambda_1 & + & \frac{1}{12}\lambda_2 & + & \frac{1}{12}\lambda_3 & + & \frac{1}{3}\lambda_4 & + & \frac{1}{12}\lambda_5 & \le & \frac{1}{3} \\ -\frac{2}{3}\lambda_1 & + & \frac{1}{3}\lambda_2 & + & \frac{1}{12}\lambda_3 & + & \frac{1}{12}\lambda_4 & + & \frac{1}{3}\lambda_5 & \le & \frac{1}{3} \end{array}$$

We'll stop the calculation at this point. We could now calculate some iterated Δ via checking the differences for the resulting normal variables λ_i and build a new LP, which would represent the possibility of combining the changes of λ_i when doing the case distinctions.

The examples show that non-trivial combining is possible, and that the method can be iterated in a surprising way. But it also shows some limitations. The above MILP is easily solved by doing the 4 case distinctions on x_1 and x_2. Furthermore it is even possible to make another case distinctions on one inequality. It is clear via the first inequality that $x_1 = 1$ or $x_2 = 1$ or $x_3 = 1$ or $x_1 = x_2 = x_3 = 0$ holds. Via this case distinction it is seen most quickly that the optimal value of the MILP is 1.

5 Application of the theory to produce cuts for the original MILP

When thinking of integrating the previous results of the concurrent branching into an existing branch and cut framework, the dual combining inequality of 2.6 does not fit easily. It is striking that instead of that additional dual LP you would just like to have more restrictions in normal space instead. Generation of cuts should be the aim. We will see that this is possible in the end of the chapter.

We start at the point that we have an optimal dual solution y^0 with only one knot $\mathfrak{k}^1$, which describes a case distinction. We now use a new special form of theorem 2.7 . For this we freeze the $\Delta(\mathfrak{k}^1)$ as linear factors of a scalar λ. Contrary to the special form 2.6 we really let $y = y^0$ play the role of dual variables and do not fix it. So our variables are the vector y and the scalar λ. Transformed back via dual-dual correspondence this will give us more or less the normal inequalities plus one additional equality, which will be our cut. We have the following dual inequalities for this special model:

Definition 5.1

$$g_m : \lambda \Delta_m(\mathfrak{k}^1) \leq y_m \ \forall m$$

$$g_n : \sum_m a_{m,n} y_m = c_n \ \forall n$$

The objective function of this dual problem is $\sum_m b_m y_m + \lambda \omega(\mathfrak{k}^1)$.

We now construct from this problem its dual problem. Contrary from the transformation from normal to dual we use the for dual of the dual the textbook form so that we have no equations:

Remark 5.2

$$y_m : \sum_n a_{m,n} x_n - x_m \geq b_m \forall m$$

$$\lambda : \sum_m \Delta(\mathfrak{k}^1) x_m \geq \omega(\mathfrak{k}^1)$$

The objective is just $\sum_n c_n x_n$ as the normal objective. This looks already promising, but prior to the final transformation to get a cut we must first prove that this system is really valid for all integer values.

Before doing the proof we must first study the reuse of knots for other basic solutions than the starting one. In 2.6 we had some knots which we tried to add to some basic solution. If we would use another solution $\bar{y}_0$, we could naturally use the methodology also. Adding the knots might still be possible. The only thing which might happen is that all λ_i in 2.6 have to be 0. The same holds if we have a stricter LP. Then the dual solution has only some more variables, but the original ones are still there. We call a linear Program P^1 compatible to another problem P^2, when it is a stricter version of the other. So the dual problem of P^1 can be seen as a loosened version than the dual of P^2, as this problem will have in general more variables. For this case have a natural mapping of a dual solution point y^0 into the dual solution set of the problem P^2.

Remark 5.3 *The values of knots can still be applied to a stricter version of the starting normal LP.*

So the knot info of a LP remains valid for all descendants.

We can now prove indirectly that the inequalities of 5.2 are valid for the LP. Consider you have an integer solution of the LP. Then it is clear, that this integer solution is the only optimal solution of a version of the LP, which has been made stricter via adding more inequalities. This more restrictive LP is represented in the dual space by a loosened LP. We can still try to add our knot in the dual space. Via this we get the special model 5.1 for the loosened dual inequality. For this model the optimal objective has to be identical to the dual LP and the normal LP. Otherwise we would prove that the optimal integer solution of the more strict LP has to have a bigger objective than the already existing integer solution, which is a contradiction.

The optimal dual solution of our loosened LP 5.1 in the dual space is a solution of a more restrictive LP 5.2 than the original one. So we have found an optimal normal solution, which also holds for the more restrictive inequalities. As we have said that the original solution was the only optimal solution, it must be identical to the new one. So the original solution has to fulfill 5.2. So all integer solutions fulfill 5.2.

As a final step we state that the normal inequality system is equivalent to:

$$f_m : \sum_n a_{m,n} x_n \geq b_m \forall m$$

$$\lambda : \sum_m (\Delta_m(\mathfrak{k}^i)(\sum_n a_{m,n} x_n - b_m) \geq \omega(\mathfrak{k}^i)$$

Or written with our functions $f_m(x) = \sum_n a_{m,n} x_n - b_m$:

$$f_m : \sum_n a_{m,n} x_n \geq b_m \ \forall m$$

$$\lambda : \sum_m \Delta(\mathfrak{k}^i) f_m \geq \omega(\mathfrak{k}^i)$$

The cut λ in its last form is surprisingly short and that is where we aimed to go. The dual LP 5.1 has at least an increase to the value of the optimal solution of $\omega(\mathfrak{k}^i)$. So the same holds for its dual which is equivalent to the last inequality. We put all our results together getting:

Theorem 5.4 (Generation of branching cuts) *The following inequality is true for all integer solutions:*

$$\lambda : \sum_m \Delta(\mathfrak{k}^i) f_m \geq \omega(\mathfrak{k}^i)$$

When the chosen starting point y^0 is optimal, the increase of the objective by adding one cut of this kind is at least $\omega(\mathfrak{k}^i)$.

Finally we apply this cut to the problem in 4 and we get:

$$\begin{array}{rcl} \frac{1}{3}f_1 + \frac{1}{12}f_2 + \frac{1}{12}f_3 + \frac{1}{3}f_4 + \frac{1}{12}f_5 & \geq & \frac{1}{6} \\ \frac{9}{12}x_1 + \frac{6}{12}x_2 + \frac{6}{12}x_3 + \frac{6}{12}x_4 + \frac{6}{12}x_5 & \leq & \frac{11}{12} - \frac{5}{6} \\ \frac{3}{4}x_1 + \frac{1}{2}x_2 + \frac{1}{2}x_3 + \frac{1}{2}x_4 + \frac{1}{2}x_5 & \leq & \frac{3}{4} \end{array}$$

When applying $x_1 = 0$ as case in the original problem, you get the solution vector $(0, \frac{1}{2}, \frac{1}{4}, \frac{1}{4}, \frac{1}{2})$. This solution is equalizing the above cut. And for the other case $x_1 = 1$ with solution vector $(1, 0, 0, 0, 0)$, this is also sharp. If we would have chosen the knot without normalization, we would have got:

$$x_1 + x_2 + x_3 + x_4 + x_5 \leq \frac{3}{2}$$

At this cut $(0, \frac{1}{2}, \frac{1}{4}, \frac{1}{4}, \frac{1}{2})$ is equalizing the cut, but $(1, 0, 0, 0, 0)$ isn't. So by this procedure the cut produced with normalization is stricter than the cut without normalization.

6 Resulting theory of cuts for branching

6.1 Adding inequalities in the dual view

We start with an inequality system $Ax - b < 0$ with some objective $\tau(x)$ and a valid dual point y^0, which doesn't have to be optimal. To this system we add the inequality (cut) f_c, introducing the new coordinate c in the dual space. Handling the inequality f_c like a branch 1 at the case distinction c in the preceding chapters, we get from the (non-)optimal dual solution $y^{c,1} = y^0 - \mathfrak{b}^{c,1}$ an objective increase $\omega(y^{c,1}) - \omega(y^0) = \omega(\mathfrak{b}^{c,1})$. Looking at the objective function for both systems as a linear combinations of $\{f_m(x)\}$ and $f_c(x)$, we find:

$$\tau(x) = \sum_m f_m(x) y^0_m + \omega(y^0)$$

$$\tau(x) = \sum_m f_m(x)(y_m^0 - \mathfrak{b}_m^{c,1}) + f_c(x)y_c + \omega(y^0 - \mathfrak{b}^{c,1})$$
$$\Rightarrow y_c f_c(x) = \sum_m f_m(x)\mathfrak{b}_m^{c,1} + \omega(\mathfrak{b}^{c,1})$$

When y_c is unequal 0 we get via this procedure a formula of the inequality f_c, which is by definition equal to the inequality from the usage of f_c as a knot consisting only of one case by our cuts of 5.4. As an inequality is only defined up to a factor, we can demand $y_c = 1$.

Theorem 6.1 (Recovering Inequalities by usage of Dual Deltas) *All inequalities f_c can be transformed to the delta form.*

$$f_c : \sum_m \mathfrak{b}_m^{c,1} f_m(x) \geq \omega(\mathfrak{b}^{c,1})$$

At this point it is possible to give an easy direct proof of the validity of 5.4. All the underlying inequalities of a case distinction can be transformed to a form like in 6.1. Via doing the maximization on the $\mathfrak{b}_m^{i,j}$ and minimization of $\omega(\mathfrak{b}^{i,j})$ all inequalities of the underlying branches are weakened, so the cut 5.4 becomes valid for all cases. To see that the objective increase in 5.4 is at least $\omega(\mathfrak{k}^i)$, we just give a fitting dual solution $\bar{y}$, which has this objective value. We set $\bar{y}_c$ to 1 and $\bar{y}_m$ to $y_m^0 - \Delta_m(\mathfrak{k}^i)$ for all m yields a representation of the objective. So this is dual solution.

It should be noted that the above form is not unique for a given inequality. Both y^0 and $y^0 - \mathfrak{b}^{c,1}$ can be chosen freely. This surprisingly easy argumentation will be again described by an example, we look again at 5.

$$\begin{aligned} f_1 :&\quad x_1 + x_2 + x_3 \leq 1 \\ f_2 :&\quad x_2 + x_3 + x_4 \leq 1 \\ f_3 :&\quad x_3 + x_4 + x_5 \leq 1 \\ f_4 :&\quad x_4 + x_5 + x_1 \leq 1 \\ f_5 :&\quad x_5 + x_1 + x_2 \leq 1 \end{aligned}$$

Maximize $\tau(x) = x_1 + x_2 + x_3 + x_4 + x_5$ and all variables have to be integer. We add now a cut produced by our method to this problem.

$$f_a : \tfrac{3}{4}x_1 + \tfrac{1}{2}x_2 + \tfrac{1}{2}x_3 + \tfrac{1}{2}x_4 + \tfrac{1}{2}x_5 \leq \tfrac{3}{4}$$

An optimal solution of the new dual LP is:

$$y_2 = y_3 = y_5 = \tfrac{1}{4} \text{ and } y_c = 1 \text{ with } \omega(y) = \tfrac{3}{2}$$

All other variables are 0. A solution of the original dual LP was:

$$y_1 = y_2 = y_3 = y_4 = y_5 = \tfrac{1}{3} \text{ with } \omega(y^0) = \tfrac{5}{3}$$

Using the representation of the objective we get for the cut this representation:

$$\begin{aligned} f_c = &\quad \tfrac{1}{3}(x_1 + x_2 + x_3 - 1) + \tfrac{1}{12}(x_2 + x_3 + x_4 - 1) + \tfrac{1}{12}(x_3 + x_4 + x_5 - 1) + \\ &\quad \tfrac{1}{3}(x_4 + x_5 + x_1 - 1) + \tfrac{1}{12}(x_5 + x_1 + x_2 - 1) \leq -(\tfrac{5}{3} - \tfrac{3}{2}) \\ f_c = &\quad (\tfrac{2}{3} + \tfrac{1}{12})x_1 + (\tfrac{1}{3} + \tfrac{2}{12})x_2 + (\tfrac{1}{3} + \tfrac{2}{12})x_3 + \\ &\quad (\tfrac{1}{3} + \tfrac{2}{12})x_4 + (\tfrac{1}{3} + \tfrac{2}{12})x_5 \leq \tfrac{2}{3} + \tfrac{3}{12} - \tfrac{1}{6} \end{aligned}$$

Multiplying the last step out we recover the original from of f_c. We want also want to prove the validity of f_c by this method. We make a branch on x_1, for case $f_{a_1} : \frac{1}{2}x_1 \leq 0$ we choose the following dual values.

$$y_{c_1} = 1 \quad y_1 = 0 \quad y_2 = \tfrac{1}{2} \quad y_3 = \tfrac{1}{2} \quad y_4 = 0 \quad y_5 = \tfrac{1}{2}$$

$$\omega(y^{1,1}) = \tfrac{3}{2}$$

Via these values we get the following representation for f_{a_1}:

$$\begin{aligned} f_{a_1} : \quad & \tfrac{1}{3}(x_1 + x_2 + x_3 - 1) - \tfrac{1}{6}(x_2 + x_3 + x_4 - 1) - \tfrac{1}{6}(x_3 + x_4 + x_5 - 1) + \\ & \tfrac{1}{3}(x_4 + x_5 + x_1 - 1) - \tfrac{1}{6}(x_5 + x_1 + x_2 - 1) \leq -(\tfrac{5}{3} - \tfrac{3}{2}) \end{aligned}$$

The other branch is the case $f_{a_2} : -\frac{1}{4}x_1 \leq -\frac{1}{4}$. As in 4 we use this non-optimal solution:

$$y_{c_2} = 1 \quad y_1 = \tfrac{1}{2} \quad y_2 = \tfrac{1}{4} \quad y_3 = \tfrac{1}{4} \quad y_4 = \tfrac{1}{2} \quad y_5 = \tfrac{1}{4}$$

$$\omega(y^{1,2}) = \tfrac{3}{2}$$

So we get the representation:

$$\begin{aligned} f_{a_2} = \quad & -\tfrac{1}{6}(x_1 + x_2 + x_3 - 1) + \tfrac{1}{12}(x_2 + x_3 + x_4 - 1) + \tfrac{1}{12}(x_3 + x_4 + x_5 - 1) + \\ & -\tfrac{1}{6}(x_4 + x_5 + x_1) + \tfrac{1}{12}(x_5 + x_1 + x_2) \leq -(\tfrac{5}{3} - \tfrac{3}{2}) \end{aligned}$$

We know that either $x \leq 0$ or $x \geq 1$ must hold, so if we produce from both inequalities the same inequality, this must be valid. This is easily done: Add the valid inequalities $\frac{1}{4}(x_2 + x_3 + x_4 - 1) \leq 0$ and $\frac{1}{4}(x_3 + x_4 + x_5 - 1) \leq 0$ and $\frac{1}{4}(x_5 + x_1 + x_2 - 1) \leq 0$ to the given representation of $x_1 \leq 0$. For the other representation just add the valid inequalities $\frac{1}{2}(x_1 + x_2 + x_3 - 1) \leq 0$ and $\frac{1}{2}(x_4 + x_5 + x_1) \leq 0$. Both representations end up in f_c, so this is a valid cut.

6.2 The relationship between our branching cuts and other cuts defined by case distinctions

In this section we prove that all cuts can be produced our dual delta method.

For the inequality f_a and the inequality (cut) f_c, both added separately to the same inequality system $\{f_m\}$, f_a is called harder than f_c in the context of $\{f_m\}$ when the solution space of the first system is contained in the second solution space.

In all of the proofs of cut generation schemes proofs by cases are used. So taking one of the cases of the proof, the set f_a of that case is harder than an arbitrary cut f_c, which will come from the method. To understand this we consider a new problem where from the additional non LP-conditions only the condition from that case distinction is to be fullfilled. The cut f_c should still be valid for that problem, so it will also be valid for the LP which represents the case related to the set f_a. So the definition just made is sensitive.

The inequality $f_a(x) \geq 0$ is harder than $f_c(x) \geq 0$ in the context $\{f_m\} \geq 0$. So the inequality $f_c(x)$ has to be a linear combination $r^{c,a}$ of $f_a(x)$ and $\{f_m\}$ with some correction for the hardness in the linear term which is a positive number:

Remark 6.2

$$f_c(x) = r_a^{c,a} f_c(x) + \sum_m r_m^{c,a} f_m(x) - (-b^{c,a})$$

Only for the ease of the notations we don't consider cases of case distinction like $x_1 = 0 \wedge x_2 = 0$. As an inequality is only defined by a factor, we can for sure demand $r_a^{c,a} = 1$. The aim is now to find a representation of $f_a(x)$ in the manner of 6.1 where all factors are smaller than the representation of $f_c(x)$ in the same manner. As $f_a(x)$ is representing an arbitrary underlying case j of a case distinction i, we could recover $f_c(x)$ by maximization of all cases. This would show that any cut by any case distinction schema can be found via our method. We have:

$$f_c(x) = f_a(x) + \sum_m r_m^{c,a} f_m(x) + b^{c,a}$$

$$f_c(x) = \sum_m \mathfrak{b}_m^{c,1} f_m(x) - \omega(\mathfrak{b}^{c,1})$$

$$\Rightarrow f_a(x) = \sum_m (\mathfrak{b}_m^{c,1} - r_m^{c,a}) f_m(x) - (\omega(\mathfrak{b}^{c,1}) + b^{c,a})$$

So we have already a desired representation with $\mathfrak{b}^{i,j} = \mathfrak{b}^{c,1} - r_m^{c,a}$, anyhow it has to checked, that $y^0 - \mathfrak{b}^{i,j}$ is a valid dual solution with the desired objective. To show this we have to verify that the objective is the linear combination of the functions assigned to each inequality.

$$\begin{array}{rrcl} & \tau(x) & = & \sum_m (y_m^0 - \mathfrak{b}_m^{i,j}) f_m(x) + f_a(x) - (\omega(y^0) + \omega(\mathfrak{b}^{c,1}) + b^{c,a}) \\ \Leftrightarrow & & & \sum_m (y_m^0 - \mathfrak{b}_m^{c,1}) f_m(x) + f_c(x) - \omega(y^0 + \mathfrak{b}^{c,1}) \\ & & = & \sum_m (y_m^0 - \mathfrak{b}_m^{c,1} + r_m^{c,a} f_m(x) + f_b(x) + \omega(y^0) + \omega(\mathfrak{b}^{c,1}) + b^{c,a} \\ \Leftrightarrow & f_c(x) & = & \sum r_m^{c,a} f_m(x) + f_b(x) + b^{c,a} \end{array}$$

As the last statement is true, also the equivalent statement is true, so $y^0 - \mathfrak{b}^{i,j}$ is a valid dual solution.

Remark 6.3 (Property of a harder inequality) *For some arbitrary inequality* $f_c : \sum_m \mathfrak{b}_m^{c,1} f_m(x) \geq \omega(\mathfrak{b}^{c,1})$ *the harder inequality* f_a *(underlying case) gives a cut* $\sum_m \mathfrak{b}_m^{i,j} f_m(x) \geq \omega(\mathfrak{b}^{i,j})$ *with* $\mathfrak{b}_m^{c,1} \geq \mathfrak{b}_m^{i,j}$ *and* $\omega(\mathfrak{b}^{c,1}) \leq \omega(\mathfrak{b}^{i,j})$.

Via doing the maximization and minimization for all cases of the case distinction we get at once the final theorem:

Theorem 6.4 (Property of cuts defined by proof by cases) *An arbitrary cut* f_c *is produced by some cut method. Then the following form reproduces* f_c :

$\sum_m \mathfrak{b}^{c,1}_m f_m(x) \geq \omega(\mathfrak{b}^{c,1})$. *Furtermore it exists a branching cut inequality of our form* $\sum_m \mathfrak{k}^i_m f_m(x) \geq \omega(\mathfrak{k}^i)$ *with* $\mathfrak{k}^i_m \leq \mathfrak{b}^{c,1}_m$ *and* $\omega(\mathfrak{k}^i) \geq \omega(\mathfrak{b}^{c,1})$. *By definition this new cut is stronger than* f_c. *So the defined class dominates all cut generation schemas for MILP.*

As last we want to investigate the values of $r^{c,a}$ in case $f_c = \sum_m \Delta_m(\mathfrak{k}^i) f_m(x)$ has been produced already by our method. By easily calculation you verify that for a given case j of the case distinction i:

$$r^{c,a}_m = \Delta_m(\mathfrak{k}^i) - \mathfrak{b}^{i,j}_m$$

7 Merging case distinctions

7.1 Merging cuts on the same case distinction

For one case differentiation you might have different knot values. Following 2.4 you might expect that using a sum of the two knots $\mathfrak{k}^i$ and $\mathfrak{l}^i$ might lead to linear inequalities, which are not the sum of the corresponding inequalities. By the following two statement we see that this is not possible for good computational implementations of the method.

Remark 7.1 (Property of normalization) *When one of the underlying cases of* $\mathfrak{k}^i$ *and* $\mathfrak{l}^i$ *are normalized, you have that* $\omega(\mathfrak{k}^i + \mathfrak{l}^i) = \omega(\mathfrak{k}^i) + \omega(\mathfrak{l}^i)$.

The proof of this remark is straightforward, anyhow this is not the case for the following statement, which will not be written down here. This statement already denies the effectiveness of idea of merging inequalities on the same case distinction.

Remark 7.2 (Property of good dual solutions) *When the underlying cases are optimal in a sense of a method similar to 8.1, then the following equality holds:*

$$\Delta(\mathfrak{k}^i + \mathfrak{l}^i) = \Delta(\mathfrak{k}^i) + \Delta(\mathfrak{l}^i)$$

7.2 Merging 2 knots

We will again study the LP from 4. Making a case distinction on $x_1 = 0 \wedge x_2 = 0$ or $x_1 = 0 \wedge x_2 = 1$ or $x_1 = 1 \wedge x_2 = 0$ or $x_1 = 1 \wedge x_2 = 1$ you get immediately the bound of 1. But the knots $\mathfrak{k}^1$ and $\mathfrak{k}^2$ cannot be combined at all, as $\Delta(\mathfrak{k}^1) = (\frac{1}{3}, \frac{1}{12}, \frac{1}{12}, \frac{1}{3}, \frac{1}{12}, 0, \ldots)$ with $\omega = \frac{1}{6}$ and $\Delta(\mathfrak{k}^2) = (\frac{1}{12}, \frac{1}{3}, \frac{1}{12}, \frac{1}{12}, \frac{1}{3}, \ldots)$ with $\omega = \frac{1}{6}$ are blocking each other.

So what's wrong with our technique? Let's investigate the underlying critical weak cases $x_1 \geq 0$ and $x_2 \geq 0$.

$$\begin{array}{lll} \mathfrak{b}^{1,1} = (\frac{1}{3}, -\frac{1}{6}, -\frac{1}{6}, \frac{1}{3}, -\frac{1}{6}, \ldots) & \text{with} & \omega(c_{1,1}) = -\frac{1}{6} \\ \mathfrak{b}^{2,1} = (-\frac{1}{6}, \frac{1}{3}, -\frac{1}{6}, -\frac{1}{6}, \frac{1}{3}, \ldots) & \text{with} & \omega(c_{2,1}) = -\frac{1}{6} \end{array}$$

So the two underlying cases can be combined. Neglecting the cases $x_1 = 1$ or $x_2 = 1$ and using 2.6, we get optimal $\lambda_1 = 2 = \lambda_2$, with objective of $-\frac{2}{3}$, which is already giving us a sharp limit for the MILP. So only in combination we reach the high λ_i-values. Both cases are helping each other to become more strict.

Let $i = 3$ the case of the case distinction on both x_1 and x_2. We identify the case $x_1 = 0$ and $x_2 = 0$ with $j = 1$, $x_1 = 0$ and $x_2 = 1$ with $j = 2$, $x_1 = 1$ and $x_2 = 0$ with $j = 3$ and $x_1 = 1$ and $x_2 = 1$ with $j = 4$. Projecting $\mathfrak{b}^{1,1}$ and $\mathfrak{b}^{2,1}$ into the dual space of $i = 3$, we see from above $\mathfrak{b}^{3,1} = 2\mathfrak{b}^{1,1} + 2\mathfrak{b}^{2,1} = (\frac{1}{3}, \frac{1}{3}, -\frac{1}{3}, \frac{1}{3}, \frac{1}{3}, \ldots)$ with $\omega(\mathfrak{b}^{3,1}) = -\frac{2}{3}$.

As the cases $x_1 = 1$ or $x_2 = 1$ also have objective decrease of $-\frac{2}{3}$, we set $\mathfrak{b}^{3,2} = \mathfrak{b}^{1,2}$ and $\mathfrak{b}^{3,3} = \mathfrak{b}^{2,2}$. Finally we choose $\mathfrak{b}^{3,4} = \mathfrak{b}^{1,2}$, also $\mathfrak{b}^{3,4} = \mathfrak{b}^{2,2}$ or $\mathfrak{b}^{3,4} = \frac{1}{2}\mathfrak{b}^{1,2} + \frac{1}{2}\mathfrak{b}^{2,2}$ would have been good choices. Putting all together we receive $\Delta(\mathfrak{k}^3) = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3}, \frac{1}{3}, \frac{1}{3})$ with $\omega(\mathfrak{k}^3) = -\frac{2}{3}$. As the resulting cut we get $x_1 + x_2 + x_3 + x_4 + x_5 \leq 1$.

The idea from the example can be used when having a set of knots. These should not be normalized, so most of the knots should have a weaker branch than the other. All of those weak branches should now be investigated, if a pair of them fits in such a way like above together. When this is the case, all resulting new branches for the double case distinction have to be defined. These have to be normalized now getting a merged knot (cut).

The above example also shows that even without helping weak-weak branchings, there is a choice how to calculate the other branches. So the cut from merged knots can have different forms.

8 Ideas for Effectiveness to find good dual values in the branching LPs

8.1 Measurement of good dual values

We go back to the dual model in 2.6. Looking at the inequality there you see that each knot blocks up certain inequalities (dual values). So to find good values, you have to search for knots and hereby for dual variables who block not much of our space but still give a good improvement in the objective function. First we have to define how to measure blocking up the space of inequalities.

Suppose again you have made a case j of case distinction i with a dual variables set, which gives a higher objective. Some of the $-\mathfrak{b}_m^{i,j}$ might be negative, but when the knot is glued together by maximizing we suspect that the value $\Delta_m(\mathfrak{k}^i)$ will be positive. Anyhow even if it is negative, quite likely no other case distinction will need the negative value. So we have argued to measure all negative $-\mathfrak{b}_m^{i,j}$ as 0. It is now a natural approach to measure the blocking up by a distance D to the starting point y^0:

$$D(\mathfrak{b}^{i,j}) = \sum_m \max(0, -\mathfrak{b}_m^{i,j})$$

We start now with having a dual solution $y^0 - \mathfrak{b}^{i,j}$ in $V^{i,j}$. We set ${}^0\mathfrak{b}^{\mathrm{i,j}} = \mathfrak{b}^{i,j}$ and want to find better a ${}^1\mathfrak{b}^{\mathrm{i,j}}$ where the ratio of $\omega({}^1\mathfrak{b}^{\mathrm{i,j}})$ to $D({}^1\mathfrak{b}^{\mathrm{i,j}})$ is better

than the corresponding values of ${}^{0}\mathfrak{b}^{\mathrm{i,j}}$. This ratio will be considered as a measure for the effectiness of a branch. The distance D can be modelled as additional inequalities with helping variables into the dual LP.
The idea is now to consider a new objective, where $\tilde{\omega}(y^0) = \tilde{\omega}(y^0 + {}^{0}\mathfrak{b}^{\mathrm{i,j}})$:

$$\tilde{\omega}(y) = \omega(y) - \frac{D({}^{1}\mathfrak{b}^{\mathrm{i,j}})}{D({}^{0}\mathfrak{b}^{\mathrm{i,j}})}\omega({}^{0}\mathfrak{b}^{\mathrm{i,j}})$$

Via our definitions we have assured that D is always positive. When the new LP is now solved to optimality getting the point $y_0 - {}^{1}\mathfrak{b}^{\mathrm{i,j}}$, it is clear that its optimal value is in the dual solution space $V^{i,j}$. Furthermore the following can easily be proved:

$$\frac{\omega({}^{1}\mathfrak{b}^{\mathrm{i,j}})}{D({}^{1}\mathfrak{b}^{\mathrm{i,j}})} \geq \frac{\omega({}^{0}\mathfrak{b}^{\mathrm{i,j}})}{D({}^{0}\mathfrak{b}^{\mathrm{i,j}})}$$

So in terms of efficiency of blocking up the space the new point is better or equal than the first point. It is visible already by our definitions, that we can iterate further. With the same technique it is possible in principle to measure knots and so cuts.

It should be not forgotten, that a prerequisite for this idea is that the set of the inequalities is well scaled, so that the dual variables have comparable value ranges.

8.2 Searching for fitting dual solutions

Suppose you have already made a branching with a knot $\mathfrak{k}^1$. Then to combine a second branching with the first you just demand additionally for the dual problem:

$$y^0 \geq \Delta(\mathfrak{k}^1) + \Delta(\mathfrak{b}^{2,j})\ \forall j$$

The benefit of this approach is that using those additional restrictions it is clear that the two branching can be fully combined. In terms of the theorem 2.6 this means that $\lambda_1 = \lambda_2 = 1$. Naturally the idea can easily be iterated via demanding:

$$y^0 \geq \sum_{\hat{i}<i} \Delta(\mathfrak{k}^{\hat{i}}) + \Delta(\mathfrak{b}^{\hat{i},j})\ \forall m, j$$

The author has made some experiments with such additional inequalities, often the produced inequalities became more and more numerical unstable.

It should be pointed out, that the order of the i can be freely chosen.

8.3 Local search for good dual solutions

Linear equations describe the nature. When a butterfly flies up in Brazil the emerging circulations won't normally influence the weather in Europe. Speaking in terms of LP an introduction of a new variable in a dual equality has often only effect in those equalities, which are strongly bound to the related inequalities. So it is striking a thought to define the neighbourhood of a new variable, and to freeze all other variables, which are not in the neighbourhood. This should have a big reduction of the running time as a result.

Clearly defining neighbourhood by the graph of the inequality system or other means is a complex story. The definition of the neighborhood should also be dependent on the type of the MILP.

Suppose you have a good neighbourhood definition. Then the technique of freezing most of the dual variables might also be an alternative to the strong branching method, which determines in a branch and cut framework the next variable to branch on as described in [6]. The strong branching relies on a good and steep implementation of the dual simplex, where you use the values of the objective after only some iteration of the Simplex algorithm. It should be noticed that such a steep dual Simplex algorithm is not a prerequisite of the algorithm. So my approach can be used in more simple LP-packages like glpk to do something similar.

9 Abstraction and conclusion

9.1 Description of the huge combining LP

The author has claimed in 2.7, that all models are special forms of a very huge LP. In this section we really write down this LP and show the embeddings. We will try to present the whole thing self-contained by leaving out the two new objects branch and knot. Let $Ax \geq b$ be the problem with additional conditions like integrity and objective cx. Let i a case distinction on one of the additional conditions (= knots), and j an underlying case. Then that problem is again $A^{i,j}x^{i,j} \geq b^{i,j}$. Naturally we are interested in the dual problem $yA = cy$ with objective by and the cases $y^{i,j}A^{i,j} = c^{i,j}y^{i,j}$.

Theorem 9.1 (big dual combining LP) *The combining LP P^c is defined now by:*

$$y^i \leq y^{i,j} \ \forall j$$

$$y + \sum_i (y^i - y) \geq 0$$

$$\omega^i \leq y^{i,j}b^{i,j} - yb \ \forall j$$

$$\omega = \sum_i \omega^i + yb$$

The objective ω which has to be maximized is a lower limit for the problem, the inequality $(y - y^i)Ax \geq (y - y^i)b + \omega^i$ is a cut for all i. The addition of all those cuts to the normal problem gives a lower bound of ω to the problem.

We leave out the very technical proof and look instead to the more simple versions of this theorem and show the embeddings. As first we look at 2.5, here we choose a fixed solution y^0 and fixed solutions $y^{i,j}$. The inequalities were just:

$$\Delta(\mathrm{t}) \leq y^0$$

$$\Leftrightarrow y_0 - \sum_i \max_j \mathfrak{b}^{i,j} \geq 0$$

$$\Leftrightarrow y_0 + \sum_i \min_j (y^{i,j} - y) \geq 0$$

So a solution to these inequalities corresponds to a solution of 9.1.

Next we analyse 2.6, we have chosen some y_0 and $y^{i,j}$ ($\mathfrak{b}^{i,j}$) and the λ_i. One problem is that some coordinates of $y^0 - \lambda_i \mathfrak{b}^{i,j}$ might be negative, when λ_i is greater 1. Via a scaling technique the last restriction can be circumvented.

Last we look at 5.1, we don't fix y_0 anymore but still $\mathfrak{b}^{i,j}$. There is no reason that $y_m - \mathfrak{b}^{i,j}_m$ must be positive. So the direct embedding for a y is not possible. But with a dual-dual argument as in 5 it is possible to show that there is some $\bar{y}$ which gives the same inequality and the advantage in the objectives. With this $\bar{y}$ all $\bar{y}_m - \mathfrak{b}^{i,j}_m$ are well defined.

In the authors view this huge LP plays a central role in studying the theory MILPs or Disjunctive Programming, as it produces cuts from disjunctions which fit optimal together to gain a good advance in the objective. So findings its good solutions is the same like creating effective cuts for the original problem.

9.2 Comparison to the theory of disjunctive cuts

A similar theory has been mainly developed by Balas. In the survey article [7] also there have defined the dual problem of Disjunctive Programming in chapter 4. For all practical cases there is a inequality system $Ax \geq b$ which dual solution space is included in all solution spaces of the problems of the disjunction (= one case distinction). It should be noted that his example on page 14 has not this property, but all MILP have this property by definition.

A mapping to 2.5 is straightforward: Take any dual solution y^0 of the stronger LP. Identify u_h with $y^{1,j}$ as we have only one case distinction.
The normal approach in this theory is to use CGLPs (Cut Generation LPs). Fixing one i the generation of a cut via solving all $y^{i,j} A^{i,j} = c^{i,j}$ is very similar to our technique. But there are big differences in the approaches, suggesting that there exists no unifying theory on both:

1. In the author's theory we can maximize the increase of the objective without getting the objective function as a result.

2. The minimum from the case distinction is taken before multiplication with A. For the CGLP-approach for the creation of disjunctive cuts the dual value u_q is multiplied with A_q and then maximum (or minimum) is taken.

3. The values of $\Delta(\mathfrak{k}^i)$ can become negative, this is not possible for the u_q.

It is already known that also disjunctive cuts can produce any cut of case differentiations. So both theories can produce the same cuts. The differences in 1, 2 and 3 are equalizing themselves. But in the author's theory the working together of cuts can be examined and understood. So this theory should explain better problems of the efficiency with cut generations schemes. Furthermore the determination of the dual variables and so the difference to a dual starting point at each branch seems to be easily implemented. This starting point can be reused for warming each of dual problems.

Studying some known ideas on disjunctive programming he found out that a lot ideas there can be reused for his theory. Creation of CGLP within the theory should be possible, as projection from a subspace (= subproblem) to the original problem. Already known limits of the usage disjunctive cuts and related methods seem to exists also for our cuts. Anyhow with the presented theory those limits can be understood. When a cut has too big coefficients there are too much and too high dual variables in the representation, so other complex inequalities are already blocked.

The theory of Mr. Balas is motivated geometrically in the normal space. The new theory is motivated algebraically in the dual space.

9.3 Conclusion

In this paper we have built a powerful theory to study cut generations methods and their efficiency, which has some similarity to disjunctive cuts. Starting from one point y^0, before adding cuts, you try out all cuts in question in a row. For each cut f_i you got an representation by the $\Delta(\mathfrak{k}^i)$. For those representations you have the big combining inequality system from 2.7. By simplifications of it, you can do a lot of things with the cuts.

You can get a quick view on where there are problems via the starting inequalities 2.6. This will give hints where collisions between cuts are situated. Furthermore you can merge cuts (knots) as in 7.2 to get other usefull cuts which are not a linear combination of the f_i. Having iterated our cuts method twice or more times, it is also possible to calculate the underlying cases of the iterated cuts.

The application of all our ideas like local search for good dual solutions should be studied at different types of problems. It might be possible that with this study there will be improvements in the state of the art of cut generators at least for some problem types.
With the represented theory it is possible to define the effectiveness of cut, see 8.1. This criteria can also be used to reject a cut, meaning that the cut is blocking too much space in the theorem 2.6 and other models. This non-effective cuts can also be described as having to big coefficients or being too similar to the objective function.
The author has implemented the cut generation method from 5.1. It worked perfectly and produced better values than the method 2.6. As this dual model is weaker than the second model, it should have indeed in general a higher objective. The author also did a few comparisons to other methods implemented in glpk. This textbook cut generation schemes were superior in quality than our general method. Anyhow these have been developed by quite a lot of people and for normal problems they are optimized by a lot experience. So it is no wonder that the effectiveness is not reached by a straightforward implementation of our

method.

References

[1] V. Chvátal. Linear Programming, W.H. Freeman, New York, 1983. Discrete Mathematics Volume 4, Issue 4, April 1973, Pages 305-337

[2] V. Chvátal. Edmonds Polytopes and a hierarchy of combinatorical problems.

[3] Michael Perregaard. Generating Disjunctive Cuts for Mixed Integer Programs, Doctoral Dissertation, Carnegie Mellon University (2003)

[4] R.E. Bixby, S. Ceria, C.M. McZeal and M.W.P. Savelsbergh. An updated mixed integer programming library: MIPLIB 3.0, Optima 58 (1998) 12-15.

[5] M. Conforti, G. Cornuejols, G. Zambelli. Polyhedral Approaches to Mixed Integer Linear Programming. p. 343 - 385 in 50 Years of Integer Programming 1958–2008
Edited by M. Juneger at al.

[6] D. Applegate, R. E. Bixby, V. Chvátal, and W. Cook. Finding cuts in the TSP.

[7] Egon Balas. Disjunctive Programming in Annals of Discrete Mathematics 5 (1979) 3-51. Technical Report 95-05, DIMACS, March 1995