An abstract model for branching and its application to mixed integer programming

Pierre Le Bodic, George Nemhauser

Abstract The selection of branching variables is a key component of branch-and-bound algorithms for solving mixed-integer programming (MIP) problems since the quality of the selection procedure is likely to have a significant effect on the size of the enumeration tree. State-of-the-art procedures base the selection of variables on their “LP gains”, which is the dual bound improvement obtained after branching on a variable. There are various ways of selecting variables depending on their LP gains. However, all methods are evaluated empirically. In this paper we present a theoretical model for the selection of branching variables. It is based upon an abstraction of MIPs to a simpler setting in which it is possible to analytically evaluate the dual bound improvement of choosing a given variable. We then discuss how the analytical results can be used to choose branching variables for MIPs, and we give experimental results that demonstrate the effectiveness of the method on MIPLIB 2010 “tree” instances where we achieve a 5% geometric average time and node improvement over the default rule of SCIP, a state-of-the-art MIP solver.

Keywords Branch and bound · Abstract model · Mixed integer programming · Computational complexity · Algorithm analysis
1 Introduction

Branch & Bound (B&B) [11] is currently the most successful and widely used algorithm to solve general Mixed Integer Programs (MIPs). B&B searches the solution space by recursively splitting it, which is traditionally represented by a tree, where the root node is associated with the entire solution space and where sibling nodes represent a partition of the solution space of their parent node. At each node, the subspace is encoded by a MIP, and its Linear Program (LP) relaxation is solved to provide a dual bound. If the LP is infeasible, or if the LP bound is no better than the primal bound, the node is pruned (the primal bound is the value of the best feasible solution found so far). Otherwise, the subspace at that node is partitioned and the two corresponding children nodes are recursively explored. The part of the B&B algorithm that decides how to partition the solution space is referred to as the branching rule. Typically, the branching rule selects one variable among the candidate variables, i.e. those that have a fractional value in the LP solution of the current node, but are required to be integer in the original MIP. Formally, suppose that at the current node the value of an integer variable \( x \) in the node’s LP solution is \( x_{lp} \notin \mathbb{Z} \). Branching on \( x \) would result in two children, each encoding a solution subspace, one in which \( x \) is upper-bounded by \( \lfloor x_{lp} \rfloor \), the other in which \( x \) is lower-bounded by \( \lceil x_{lp} \rceil \). The PhD thesis [1] provides a practical point of view of the latest advances in MIP solving and branching in particular. For a general overview of MIP solving and B&B, see [5,6,13].

This research is motivated by the desire to understand the fundamentals of state-of-the-art MIP branching rules (many of which are justified experimentally). The main contribution of this paper is the introduction of theoretical decision problems to study the branching component of B&B. Based on an analysis of these models, we introduce new scoring functions, and prove their efficiency on both simulated experiments and MIP instances. The paper is organized as follows. After defining the abstract B&B model in Sect. 2, we study the simplest of our problems in Sect. 3. More complex problems are introduced and analyzed in Sects. 4 and 5, respectively. Section 6 introduces two new scoring functions based on the theory developed in previous sections. Experimental results are presented in Sect. 7 and conclusions are given in Sect. 8.

2 Abstract Branch & Bound model

In this section we model an abstract version of B&B through a series of definitions. We start by defining what we consider to be a variable.

Definition 1 A variable \( x \) is encoded by a pair of non-negative (left and right) integer gains \( (l_x, r_x) \) with \( 1 \leq l_x \leq r_x \).

The \( l_x \) and \( r_x \) gains model the dual bound changes that occur when branching on the variable \( x \). We suppose that these two gains are known and fixed. (Note that when solving a MIP using B&B, these gains are not fixed for a variable.)

We now define a Branch & Bound tree.
Definition 2 Given a set of variables, a Branch & Bound tree is a binary tree (in the graph sense) with one variable \( x \) assigned to each inner node \( i \).

We then say that variable \( x \) is branched on at node \( i \). Note that, in Definition 2, the root node is considered as an inner node. Also note that, in our abstract model, the gains \((l_x, r_x)\) of a variable \( x \) do not change with the depth of the node where the variable is branched on, nor do they depend on the dual bound at this node.

These two definitions naturally lead to the definition of dual gap closed at a node \( i \) in a B&B tree. Note that throughout the article, the gap will always refer to the absolute gap (as opposed to the relative gap), i.e. the absolute difference between the primal bound and the dual bound.

Definition 3 The dual gap \( g_i \) closed at a node \( i \), referred to as \( g_i \), is given by the recursive formula

\[
g_i = \begin{cases} 
0 & \text{if } i \text{ is the root node} \\
gh + l_{x(h)} & \text{if } i \text{ is the left child of node } h \\
gh + r_{x(h)} & \text{if } i \text{ is the right child of node } h,
\end{cases}
\]

where \( x(h) \) is the variable branched on at node \( h \).

For instance, if at a node \( i \), the dual bound is \( g_i = 10 \), and variable \((2, 5)\) is branched on, the dual bound closed at the left and right children of \( i \) are 12 and 15, respectively.

The gap closed by a B&B tree is defined as follows:

Definition 4 A tree closes a gap \( G \) if for all leaves \( i \), \( g_i \geq G \) holds.

Throughout, the size of a tree refers to the number of nodes of the tree. Using these definitions, the following sections define problems that model the B&B algorithm in increasing complexity.

3 The Single Variable Branching problem

The first problem we define using the abstract model set up in Sect. 2 is the Single Variable Branching (SVB) problem, which is a tree-size measurement problem.

Problem: Single Variable Branching

Input: one variable encoded by \((l, r) \in \mathbb{Z}_{>0}^2, G \in \mathbb{Z}_{>0}, k \in \mathbb{Z}_{>0}, \text{such that } l \leq r \leq G\).

Question: Is the size of the Branch & Bound tree that closes the gap \( G \), repeatedly using variable \((l, r)\), at most \( k \)?

In the example given in Fig. 1, a B&B tree that closes a gap of \( G = 6 \) is built using only variables with gains \((2, 5)\). Observe that any tree closing a gap \( G \) contains the unique tree that closes \( G \) in a minimum number of nodes. As we will see, the analysis of SVB will prove more complex and insightful than it may seem at first glance.
Fig. 1  B&B tree with 9 nodes that closes a gap of 6 with variable gains (2, 5). The gap closed at each node is indicated by the value at the center of each node.

### 3.1 Motivation: state-of-the-art scoring functions and example

A state-of-the-art branching rule for MIP solving (as implemented in SCIP [2]) is a hybrid of mainly two methods: strong branching and pseudo-cost branching. Given a candidate variable $x$, strong branching computes the LP gains of the would-be children of the current node if $x$ is branched on. The candidate variable which provide the best LP bounds is then branched on. Pseudo-cost branching complements strong branching; it keeps track of the LP gains of branching variables for which the children have already been processed, and, given sufficient historical data, estimates the gains that would be computed by strong branching. Strong branching is computationally expensive, as it requires solving many LPs, while pseudo-cost branching only requires a few arithmetic operations at each node. However, pseudo-cost branching is an estimate, and it requires initialization. The state-of-the-art branching rule essentially consists of using strong branching early in the tree, i.e. at the root node and a few subsequent levels, while pseudo-cost branching is called at lower levels. In practice, there are many refinements to these methods. We direct the reader to Tobias Achterberg’s PhD thesis [1] and references therein for an in-depth review of branching rules.

At each node, strong branching (or pseudo-cost branching) provides the (estimated) LP gains $(l_x, r_x)$ resulting from branching up and down for each fractional variable $x$. Each candidate variable $x$ is then scored according to its gains $(l_x, r_x)$, and the highest scoring variable is selected for branching. The state-of-the-art scoring function used for this purpose is:

$$\max(\epsilon, l_x) \times \max(\epsilon, r_x),$$

where $\epsilon > 0$ is chosen close to 0 (e.g. $\epsilon = 10^{-6}$ in [1]) to break the ties if all variables satisfy $\min(l_x, r_x) = 0$, but plays no role otherwise. We simply refer to this scoring function as the product function. Prior to the introduction of the product function, the linear (or convex) function was the standard:

$$(1 - \mu) \times l_x + \mu \times r_x,$$

where $\mu$ is a parameter in $[0, 1]$ $(\mu = \frac{1}{6}$ in [1]). In [1], Achterberg reports that switching from the linear function to the product function yields an improvement of 14% in computing time, and 34% in number of nodes. The rationale behind the linear...
function with $\mu = \frac{1}{6}$ is that it should be preferable to improve the dual bound of both children by a little rather than only one of them. The product function prefers equal left and right gains, i.e. it is more likely to produce a balanced tree. However, there are types of instances for which the linear function performs better, so the product function does not systematically outperform the linear one.

To the best of our knowledge, both the linear and the product function have been established experimentally, and no theoretical evidence supports the use of one over the other, or over any other possible function. We now give a simple example suggesting that more complex functions are required to score variables. If the product or the linear functions are good scoring functions for solving MIPs, then they should also perform well on simple models such as SVB. Consider variables $(10, 10)$ and $(2, 49)$: both the product and the linear function (with default linear coefficient $\mu = \frac{1}{6}$) assign these variables an equal score. Observe that variable $(2, 49)$ receives a smaller score than $(2, 50)$ (and therefore than variable $(10, 10)$). We now consider the tree-sizes of two SVB instances, one that has variable $(2, 49)$ as input, the other having variable $(10, 10)$.

Figure 2 gives two SVB tree-size plots. For gaps no larger than 40, variable $(10, 10)$ requires fewer nodes than $(2, 49)$ to close the gap, but larger gaps are closed faster with $(2, 49)$. For a gap of 1000, using only $(10, 10)$ produces a tree 323 million times larger than using only $(2, 49)$, yet both scoring functions assign a higher score to variable $(10, 10)$. It is clear from this example that the state-of-the-art scoring functions are imperfect.

3.2 A polynomial-time algorithm

There exists a simple recurrence relation that solves SVB. For a given gap $G$, $t(G)$ is defined as the size of the SVB tree that closes $G$:

$$
t(G) = \begin{cases} 
1 & \text{if } G \leq 0 \\
1 + t(G - l) + t(G - r) & \text{otherwise.}
\end{cases}
$$

(1)
Hence, for a given $k$, the answer to SVB is YES if and only if $t(G) \leq k$. Unfortunately, computing $t(G)$ requires $O(G)$ operations. Therefore the running time of this algorithm is pseudo-polynomial (and thus exponential) in the encoding length of the input. Moreover, observe that we do not necessarily have $G = O(\log(k))$ for YES instances of SVB: if the given variables are all $(1, G)$, then $k = 2G + 1$ is the tree-size. However, there exists a closed-form formula that can be evaluated in polynomial time:

**Theorem 1** SVB is in P. Furthermore, a formula for the size of the B&B tree is

$$t(G) = 1 + 2 \times \sum_{k_r=1}^{\lceil \frac{G}{r} \rceil} \left( k_r + \left\lceil \frac{G-(k_r-1)r}{r} \right\rceil - 1 \right).$$

(2)

**Proof** The size of the input is $O(\log(G) + \log(k))$. Let $T$ be the minimal tree that closes the gap $G$, and let $d = \lceil \frac{G}{r} \rceil$ be the minimum depth of its leaf nodes. If $k < 2^{d+1} - 1$, the answer to the decision problem SVB is necessarily NO, therefore we can assume the opposite throughout the rest of proof. It follows that $d = O(\log(k))$. Starting from the root of $T$, we turn right at most $d$ times before finding a leaf. For a given path from the root to a leaf, let $k_l$ (resp. $k_r$) denote the number of times we turn left (resp. right). Observe that a pair $(k_l, k_r)$ does not uniquely identify a leaf. Furthermore, since the number of leaves is in $O(k)$, it is impossible to iterate over all of them in a time polynomial in the size of the input. However, since all leaves satisfy $k_r \leq d$, it is possible to iterate over all the values of $k_r$. From now on we suppose that $k_r \leq d$ is fixed, and restrict our attention to those leaves that are reached by turning right exactly $k_r$ times. The inequality

$$k_l \times l + k_r \times r < G + r$$

must hold for each leaf, since the gap closed at each leaf cannot exceed $G$ by $r$ or more, as otherwise the gap would already be closed at the leaf’s parent. We therefore have the bound $k_l \leq \left\lceil \frac{G-r(k_r-1)}{r} \right\rceil - 1$, and the depth of a leaf is bounded by $k_r + \left\lceil \frac{G-r(k_r-1)}{r} \right\rceil - 1$. Furthermore, observe that any $k_r$ elements chosen in the set $\{1, \ldots, k_r + \left\lceil \frac{G-r(k_r-1)}{r} \right\rceil - 1\}$ uniquely determine a path to a leaf, and that each leaf can be encoded using this scheme. This bijection thus ensures that the number of such leaves is

$$\left( k_r + \left\lceil \frac{G-r(k_r-1)}{r} \right\rceil - 1 \right).$$

The total number of leaves is therefore

$$\sum_{k_r=0}^{\lceil \frac{G}{r} \rceil} \left( k_r + \left\lceil \frac{G-(k_r-1)r}{r} \right\rceil - 1 \right).$$

(3)

Formula (3) can be computed in $O(d^2) = O((\log^2(k))$ operations, which is polynomial in the size of the input. Observe how the formula iterates over the possible $k_r$’s but avoids iterating over the range of $k_l$’s, as this may require $O(k)$ steps. Since each inner node has exactly two children, the theorem is proven.

\[ \square \]
Evaluating formula (2) becomes impractical as the gap $G$ becomes large. We present below an asymptotic study of the tree-size that can be used in practice when formula (2) is too expensive to compute.

### 3.3 Asymptotic study

Analyzing algorithms through a related recurrence relation similar to (1) (and its mathematical properties, such as the asymptotic behavior) is a standard technique (see [7, Section 4.5]). In this section and throughout the article we however systematically provide self-sufficient explanations and analyzes, as our readership may not be familiar with algorithm analysis techniques.

As the gap to close $G$ becomes large, we prove that the tree-size grows essentially linearly, i.e. there exists a fixed ratio determining the growth of the tree. Perhaps the most popular example of this phenomenon is the Fibonacci sequence, for which the golden ratio $\frac{1+\sqrt{5}}{2}$ corresponds to the asymptotic growth rate between two consecutive numbers. In fact, the Fibonacci sequence is given by the recursion formula (1) with variable $(1, 2)$ up to the additive constant $1$, which does not intervene in asymptotic results. This section therefore provides convergence results on sequences generalizing the Fibonacci sequence. Unfortunately, even though using the same notion for the ratio as in the Fibonacci sequence would be the most intuitive, the limit of $\frac{t(G+1)}{t(G)}$ when $G$ tends to infinity may not exist in our case. Indeed, consider variable $(2, 2)$, and the limit of these two subsequences:

$$\lim_{G \to \infty} \frac{t(G+1)}{t(G)} = \begin{cases} 1 & \text{if } G \text{ is odd} \\ 2 & \text{if } G \text{ is even} \end{cases}$$

In this example, the ratio $\frac{t(G+1)}{t(G)}$ does not converge as $G$ tends to infinity. However, there exists another definition of the ratio, that can be shown to converge. We use the following definition:

**Definition 5** The ratio of a variable is

$$\varphi = \lim_{G \to \infty} \sqrt[2]{\frac{t(G+1)}{t(G)}}.$$ 

With this definition, variable $(2, 2)$ has ratio $\varphi = \sqrt{2}$. This means that, asymptotically, the number of nodes doubles for every two additional units of gap to close.

Since the ratio $\varphi$ indicates the growth rate of a tree in the SVB setting, a variable with a small ratio requires a smaller SVB tree than one with a bigger ratio, given a large enough gap $G$. The ratios of the two variables taken as example in Sect. 3.1, namely $(10, 10)$ and $(2, 49)$, are approximately $1.071$ and $1.049$, respectively. This explains why variable $(2, 49)$ produces smaller trees than $(10, 10)$ for large enough gaps. Throughout this section we show that the limit defining the ratio $\varphi$ in Definition 5 always exists and how it can be computed. We first prove Proposition 1, which gives an equivalent way of defining $\varphi$. 

\[ \varphi = \lim_{G \to \infty} \sqrt[2]{\frac{t(G+1)}{t(G)}}. \]
Proposition 1 If the limit $\varphi$ exists, then we also have

$$\varphi = \lim_{G \to \infty} \sqrt[1]{\frac{t(G + r)}{t(G)}}.$$ 

Proof

$$\varphi^r = \left( \lim_{G \to \infty} \frac{t(G + l)}{t(G)} \right)^r$$

$$= \lim_{G \to \infty} \frac{t(G + l)}{t(G)} \times \frac{t(G + l)}{t(G)} \times \cdots \times \frac{t(G + l)}{t(G)}$$

$$= \lim_{G \to \infty} \frac{t(G + lr)}{t(G + l(r - 1))} \times \frac{t(G + l(r - 1))}{t(G + l(r - 2))} \times \cdots \times \frac{t(G + l)}{t(G)}$$

$$= \left( \lim_{G \to \infty} \frac{t(G + r)}{t(G)} \right)^l.$$ 

Proposition 2 If the limit $\varphi$ exists, then $\sqrt{2} \leq \varphi \leq \sqrt{2}$.

Proof Let $T$ (resp. $T^l$, $T^r$) be the tree that closes the gap $G$ (resp. $G + l$, $G + r$), and observe that the tree $T$ is a sub-tree of $T^l$ and $T^r$ (all three have the same root). First, each leaf of $T$ is an inner node of $T^r$. This means that the number of nodes at least asymptotically doubles, thus 2 is a lower bound on $\varphi^r$. Second, each leaf of $T$ is either a leaf or the parent of two leaves in $T^l$, hence 2 is an upper bound on $\varphi^l$. □

The limit $\varphi$ clearly exists in the case $l = r$, and the bounds in the above property provide the exact value. We now prove that the limit $\varphi$ exists in general:

Theorem 2 When $G$ tends to infinity, both sequences $\sqrt[1]{\frac{t(G + l)}{t(G)}}$ and $\sqrt{\frac{t(G + r)}{t(G)}}$ converge to $\varphi$, which is the unique root greater than 1 of the equation $p(x) = x^r - x^{r-l} - 1 = 0$.

Proof See “Proof of Theorem 2 and Corollary 1 in Appendix” and Theorem 3. □

Corollary 1 A numerical approximation of $\varphi^r$ is given by the fixed-point iteration

$$f(x) = 1 + \frac{1}{x^r - 1}$$

with starting point $x = 2$. □
Table 1  Ratios for some variables, truncated to 5 significant digits

| Variables | (6, 10) | (5, 11) | (4, 12) | (3, 13) | (2, 14) | (1, 15) |
|-----------|--------|--------|--------|--------|--------|--------|
| $\varphi$ | 1.0926 | 1.0955 | 1.1002 | 1.107  | 1.1204 | 1.1468 |

| Variables | (6, 12) | (4, 15) | (3, 20) | (2, 30) | (1, 60) |
|-----------|--------|--------|--------|--------|--------|
| $\varphi$ | 1.0926 | 1.0907 | 1.0873 | 1.0813 | 1.0709 | 1.0515 |

Proof  See “Proof of Theorem 2 and Corollary 1 in Appendix”.  

Having established Theorem 2, the definitions of $\varphi$ and $p$ should now make more sense, as they can be shown to originate more directly from the definition of the sequence:

\[ t(G) = 1 + t(G - l) + t(G - r) \]

\[ \Rightarrow \frac{t(G)}{t(G - r)} = \frac{1}{t(G - r)} + \frac{t(G - l)}{t(G - r)} + 1 \]

\[ \Rightarrow \lim_{G \to \infty} \frac{t(G)}{t(G - r)} = \lim_{G \to \infty} \left( \frac{1}{t(G - r)} + \frac{t(G - l)}{t(G)} + 1 \right) \]

\[ \Rightarrow \varphi^r = \varphi^{r-l} + 1. \]

The polynomial $p : x \to x^r - x^{r-l} - 1$ is the so-called characteristic polynomial of the recurrence sequence defining $t$. The Abel-Ruffini theorem [8, p. 264] states that there is no general closed-form formula for roots of polynomials with degrees five or higher. Since we are dealing with a particular trinomial, a formula still might exist, however we have not been able to determine one. Throughout the paper we therefore resort to numerical methods to determine $\varphi$ (see Sect. 6.2). The characteristic polynomial for trees with arbitrary degrees has been discussed at length in a SAT setting in [10], where the ratio $\varphi$ is referred to as the $\tau$-value.

Examples of variables and their respective ratios are given in Table 1. In the case where the sum of the variable gains is fixed (Table 1a), choosing $l$ close to $r$ minimizes the ratio of a variable $(l, r)$. However, if the product is fixed (Table 1b), then the opposite is preferable. Figure 3 gives a plot of the ratio as a function of the left and right gains.

Once the value of $\varphi$ is determined, it is then easy to circumvent the $O(\log^2(k))$ complexity of the closed-form Formula (2), by evaluating $t(F)$ for some value $F$ much smaller than $G$, and approximating $t(G)$ using the formula

\[ t(G) \approx \varphi^{G-F} t(F). \]

We now establish some useful properties of $p$.

**Theorem 3** The characteristic polynomial $p : x \to x^r - x^{r-l} - 1$ satisfies the following properties.
- $p$ is monotonically increasing in $[1, \infty)$
- in $[1, \infty)$, $p$ has a single real root $\varphi > 1$

Proof Consider the sign of $p'(x) = rx^{r-1} - (r - l)x^{r-l-1}$ when $x \in R^+$.

$$p'(x) > 0 \iff x^{r-l-1}(rx^l - r + l) > 0$$
$$\iff rx^l - r + l > 0$$
$$\iff x > \sqrt{1 - \frac{l}{r}}$$
$$\iff x \geq 1.$$  

We thus know that $p$ is increasing over $[1, \infty)$. Furthermore, $p(1) = -1$ and $\lim_{x \to \infty} p(x) = \infty$, and $p$ is continuous, therefore there exists a single root $\varphi$ in $(1, \infty)$.  

This analysis is corroborated by the fact that the tree-size increases when incrementing the gap by $l$ or $r$, thus we should have $\varphi > 1$.  

4 The Multiple Variable Branching problem

We consider the Multiple Variable Branching problem (MVB), which naturally extends the SVB problem, defined as follows:

Problem: Multiple Variable Branching
Input: $n$ variables encoded by $(l_i, r_i)$, $i \in \{1, \ldots, n\}$, an integer $G > 0$, an integer $k > 0$.

Question: Is there a Branch & Bound tree with at most $k$ nodes that closes the gap $G$, using each variable as many times as needed?
Figure 4 illustrates the difference between MVB and SVB, given two variables \((2, 5)\) and \((3, 3)\), and \(G = 7\). In the MVB tree, variable \((2, 5)\) is branched on at all (inner) nodes except the bottom left one, where variable \((3, 3)\) is branched on.

There is a recursive equation for MVB, similar to (1) for SVB. For a given gap \(G\), \(t(G)\) is defined as the minimum size of a MVB tree that closes \(G\):

\[
t(G) = \begin{cases} 
1 & \text{if } G \leq 0 \\
1 + \min_{1 \leq i \leq n} (t(G - l_i) + t(G - r_i)) & \text{otherwise}.
\end{cases}
\]  

(4)

The time complexity of this algorithm is \(O(nG)\), which is pseudo-polynomial in the size of the input. Unlike SVB, we do not know whether MVB can be solved in polynomial time. For a given \(G\), we say that variable \(i\) is branched on at the corresponding node if \(i \in \arg \min_{1 \leq i \leq n} (t(G - l_i) + t(G - r_i))\).

### 4.1 Motivating example (continued)

Following the example introduced in Sect. 3.1 for the SVB problem, consider now the plot of the MVB tree-sizes in Fig. 5 (the SVB plots are the same as in Fig. 2).

As anticipated, for a given gap, the MVB tree-size is smaller than each of the SVB tree-sizes. However, the difference in tree-sizes with \((2, 49)\) does not seem to increase drastically. Indeed, for a gap of 1000, the SVB tree-size of \((2, 49)\) is only 1.798 times
larger than the MVB tree-size (and this ratio is approximately constant for larger gaps).
In fact, for this instance, we can verify experimentally that variable (2, 49) is branched on at every node for which the gap to close is at least 31. This phenomenon is the subject of the next section.

4.2 Asymptotic study

Let \( \phi \) (resp. \( \phi_i \)) now denote the ratio associated with the recursive equation (4) of MVB (resp., for SVB, equation (1) for variable \( i \)). Furthermore, for a variable \( i \), let \( p_i \) be its characteristic polynomial, for \( i = 1, \ldots, n \). Formally, we define the ratio for a MVB problem to be

\[
\phi = \lim_{G \to \infty} \sqrt[\frac{t(G+z)}{t(G)}]
\]

where \( z \) is the least common multiple of all \( l_i \) and \( r_i \).

**Theorem 4** \( \phi = \min_i \phi_i \)

**Proof** See “Proof of Theorem 4 in Appendix”.

Essentially, this means that for a large \( G \), the tree-size grows at the growth rate of the best variable for SVB. Therefore we also have

\[
t(G) \approx \varphi^{G-F} t(F)
\]

for large gaps \( F \) and \( G \), with \( G \geq F \). Furthermore, we propose the following conjecture.

**Conjecture 1** For each instance of MVB, there exist a gap \( H \) such that for all gaps greater than \( H \), variable \( i = \arg \min_j \varphi_j \) is always branched on at the root node.
Proving this conjecture (and proving that $H$ is polynomially bounded by the size of the input) may help design a polynomial-time algorithm for MVB. Figure 6 provides the value of $H$ for instances of MVB with two variables, $(10, 10)$ and $(2, r)$, where $r$ is the value on the horizontal axis. The value of $H$ is maximized for $r = 29$, which also minimizes $|\varphi(10, 10) - \varphi(2, r)|$.

5 The General Variable Branching problem

The General Variable Branching problem (GVB) is defined as follows:

**Problem: General Variable Branching**

**Input:** $n$ variables encoded by $(l_i, r_i), i = 1, \ldots, n$, an integer $G > 0$, an integer $k > 0$, and a vector of multiplicities $m \in \mathbb{Z}_{>0}^n$.

**Question:** Is there a Branch & Bound tree with at most $k$ nodes that closes the gap $G$, branching on each variable $i$ at most $m_i$ times on each path from the root to a leaf?

Observe that GVB comprises the input of problem MVB, and additionally stipulates that each variable may not be used more than a given number of times on each path from the root to a leaf. Consequently, GVB corresponds to MVB whenever the multiplicity of each input variable is large enough, therefore GVB generalizes MVB.

When solving a MIP instance using a B&B search, the minimization version of GVB models the problem of choosing a variable to branch on under the following simplifying hypotheses:

- The up and down gains are known for each variable, and are invariant.
- If the problem is feasible, the optimal value is known.

The first hypothesis is not satisfied in practice. However, as the B&B tree grows, pseudo-costs become better approximations for the up and down gains. Likewise, the second hypothesis becomes true at some point during the B&B search. So, while both
Fig. 7 Minimum-size trees for MVB and GVB with input $G = 6$ and variables $(1, 1), (2, 5), (3, 3)$ (in GVB, with multiplicities 1). a Minimum-size MVB tree (size = 7). b Minimum-size GVB tree (size = 11)

hypotheses are not initially satisfied, they become more accurate as the tree grows. As we will see, these hypotheses are not sufficient to render the GVB problem theoretically tractable. A recursion for GVB similar to (4) for MVB is

$$t(G, m) = \begin{cases} 1 & \text{if } G \leq 0 \\ 1 + \min_{1 \leq i \leq n, m_i > 0} (t(G - l_i, m - v_i) + t(G - r_i, m - v_i)) & \text{otherwise} \end{cases}$$

where $m$ and $v_i$, are both vectors of size $n$; $m$ is the vector of multiplicities, and $v_i$ is the indicator vector of the set $\{i\}$ (i.e. the $i$th element of $v_i$ is 1, the $n - 1$ others are 0). This algorithm keeps track of the variables already branched on, and, in the case where all multiplicities are 1, its running time is $O(2^n G)$.

Figure 7 gives the minimum-size B&B tree for a GVB instance and an MVB instance that has the same variables and gap as input.

5.1 Complexity of GVB

For the complexity result presented in this section, it is sufficient to consider the case where all multiplicities are one. We prove the #P-hardness of GVB by reducing it to a variant of counting knapsack solutions:

**Problem: #Knapsack**

**Input:** N items with weights $w_1, \ldots, w_N$ and $W$ the capacity of the knapsack, a given $K$, all integer positive.

**Question:** Are there at least $K$ distinct feasible solutions $x$ to the (covering) knapsack constraint $\sum_{i=1}^{N} w_i x_i \geq W$, with $x_i \in \{0, 1\}, i = 1, \ldots, N$?

Note that in our definition of the #KNAPSACK problem, we ask for at least $K$ solutions. This variant of the counting problem can solve the exact counting problem by e.g. dichotomy in $O(N)$ calls, hence it is #P-hard.

**Theorem 5** GVB is (weakly) #P-hard

**Proof** We build an instance of the GVB problem that embeds an instance of the #KNAPSACK problem. Set $n = N + 1$, and for each $i \in \{1, \ldots, n - 1\}$, create a variable
with gains \((C, C + w_i)\), where \(C = \sum_{j \in \{1, \ldots, N\}} w_j\); furthermore, create the \(n\)th variable \((C, C)\). Set the target gap \(G = (n - 1)C + W\) and the number of allowed nodes \(k = 2^{n+1} - 1 - 2K\). Finally, set all multiplicities to one. Suppose, without loss of generality, that the knapsack instance is feasible, and thus \(C \geq W\). We prove that the answer to this GVB instance is YES if and only if the answer to the instance of \(\#\text{Knapsack}\) is YES.

First, observe that a minimum-size tree is obtained by branching on the variable with the biggest right gains, i.e. corresponding to the biggest items (w.r.t. \(w_i\)), and finally by branching on the dummy variable with gains \((C, C)\). Consider a tree \(T\) that closes the gap \(G\). Observe that each leaf of \(T\) has at least depth \(n - 1\), and at most \(n\). In fact, there is a one-to-one mapping between (not necessarily feasible) solutions of the knapsack instance and nodes at depth \(n - 1\). Furthermore, there is a one-to-one mapping between feasible solutions of the knapsack instance and leaves at depth \(n - 1\). Therefore, since an infeasible knapsack instance would yield a perfect tree, with \(2^{n+1} - 1\) nodes, each feasible solution decreases the number of nodes by 2 exactly. If there are at least \(K\) feasible solutions to \(\#\text{Knapsack}\), there is a tree with at most \(2^{n+1} - 1 - 2K\) nodes for their corresponding GVB instance, and vice-versa.

In this proof we use the fact that the domination relation, formally defined in Sect. 6, is a total order on the set of variables, and thus which variables to branch on to obtain a minimum-size tree is trivial. As a result, we actually prove Theorem 5 for a special case of the GVB problem, where at a given level, a single variable is branched on at every node. Similar to SVB, this special case is therefore a tree measurement problem.

Theorem 5 in particular implies, under the widely believed conjecture that the polynomial hierarchy \(\text{PH} [15]\) is proper to the second level, that GVB is neither in \(\text{NP}\) nor in \(\text{co-NP}\). This is because \(\text{PH} \subseteq \text{P}^{\#P} [16]\).

6 Using the abstract model for scoring branching candidates

6.1 Applicability of our results to rational numbers

The results we present in Sects. 3–5 suppose integer gains and gap. This does not hold in general when solving MIP instances, where numbers are encoded as rationals. It is however sufficient to notice that in SVB, MVB or GVB, if some data is rational, there exists a scaling factor \(q \in \mathbb{Q}_+\), such that, if the gap and all gains are multiplied by \(q\), all data become integer. Indeed, in the abstract setting described in Sect. 2, which holds for SVB, MVB or GVB, suppose a B&B tree closes a gap \(G\), and all gains are multiplied by \(q\), then the same tree closes a gap \(q \times G\), since the gap closed at each node is the sum of all gains along the path from the root to the leaf.

Furthermore, given a variable with rational gains \((l, r) \in \mathbb{Q}_+^2\), and a scaling factor \(q \in \mathbb{Q}_+\), such that the scaled variable has gains \((q \times l, q \times r) \in \mathbb{Z}_+^2\), then the ratio \(\phi\) of the scaled variable can be computed as described in Sect. 3, and we have
\[
\phi^qr - \phi^q(r-l) - 1 = 0 \Leftrightarrow (\phi^q)^r - (\phi^q)^{(r-l)} - 1 = 0,
\]
therefore the ratio of variable \((l, r)\) is given by \(\phi^q\).
Finally, the proof of Theorem 1 (for the closed-form formula) does not require data to be integer, or even rational, and Corollary 1 (for the fixed-point equation) inherently works on rational numbers. We implement both formulas in our code without any scaling.

6.2 The ratio scoring function

Recall from Sect. 3.1 that a scoring function combines the left and right LP gains to score candidate variables. The variable with the highest score is then branched on. The linear and product functions are two scoring functions that were presented in Sect. 3.1. In this section, we introduce the ratio scoring function, based on Definition 5 for the SVB problem. Given two variables $i$ and $j$ and their respective ratios $\varphi_i$ and $\varphi_j$, it selects the variable with the smallest ratio.

To help analyze the behavior of the ratio scoring function, we formalize the concept of domination for two variables.

**Definition 6** Given two variables $(l_1, r_1)$ and $(l_2, r_2)$, we say that $(l_1, r_1)$ dominates $(l_2, r_2)$ if and only if $l_1 \geq l_2$ and $r_1 \geq r_2$, with at least one of the inequalities strict.

**Proposition 3** Suppose variable $(l_1, r_1)$ dominates variable $(l_2, r_2)$. Then,

1. Both the product and the linear function assign a better score to $(l_1, r_1)$ than to $(l_2, r_2)$.
2. In the MVB or GVB problem setting, if both $(l_1, r_1)$ and $(l_2, r_2)$ are branched on on a common path, branching on $(l_1, r_1)$ before branching on $(l_2, r_2)$ yields a tree of size no larger than the converse.

**Proof** Part 1 is obvious. Part 2 can be proven by observing that if in a B&B tree, variable $(l_2, r_2)$ is branched on before $(l_1, r_1)$, then the tree where the variables are switched closes a gap at least as large. \(\square\)

In other words, the linear and the product scoring functions prefer non-dominated variables, as dominated variables necessarily yield larger trees. We prove that the ratio scoring function also exhibits this property:

**Theorem 6** Suppose variable $(l_1, r_1)$ dominates variable $(l_2, r_2)$, then $\varphi_1 < \varphi_2$.

**Proof** Starting with the definition of the characteristic polynomial of variable $j$, we establish:

\[
0 = p_2(\varphi_2) = \varphi_2^r - \varphi_2^l l_2 - 1 = \varphi_2^r (1 - \varphi_2^{-l_1}) - 1 \]

\[
< \varphi_2^r (1 - \varphi_2^{-l_1}) - 1 = p_1(\varphi_2).
\]

Therefore, by Theorem 3, $\varphi_2 > \varphi_1$. \(\square\)

As a consequence, the ratio of a dominated variable does not have to be computed. Moreover, to perform comparisons between ratios, it is not necessary to compute both of them.
Proposition 4  Suppose that the ratio $\varphi_1$ of variable $(l_1, r_1)$ has been computed, then, given a variable $(l_2, r_2)$, $\varphi_2 < \varphi_1$ if and only if $p_2(\varphi_1) > 0$.

Proof  Direct using Theorem 3.

Hence a strategy to reduce the number of ratio computations is to compute $\varphi$ for a variable that is believed to be a good candidate for branching (e.g. the best according to the product function), and use Proposition 4 to test all other candidates, computing the ratio of a variable only if it is proven to be smaller than the current best one.

The scoring rule that we implement uses the strategy described in Algorithm 1. The computation of the ratio at Steps 2 and 5 is done numerically. We have implemented and tested five different methods, namely the fixed-point iteration method described in Corollary 1, a simple bisection method, Newton and Laguerre’s method [14, Chapter 9], and a direct method computing (5) using the closed-form formula described in Theorem 1 for a large enough gap. Iterative methods are initialized with starting point $\sqrt{2}$, which is a proven lower bound (see Proposition 2). Experiments have shown that if $\frac{r}{l} \leq 100$, Laguerre’s method was the most effective, otherwise the fixed-point iteration method should be used. Following this rule, computing the ratio of a variable takes around 20 milliseconds of CPU time on a modern computer.

Algorithm 1  Implementation of the ratio scoring function.

1: Filter out the branching candidates with dominated gains.
2: Compute the ratio $\varphi^*$ of the best variable according to the product function.
3: for all remaining branching candidates $i$ do
4:  if $p_i(\varphi^*) < 0$ then $\triangleright$ Proposition 4
5:    Compute the ratio $\varphi_i$ and set $\varphi^* = \varphi_i$.
6:  end if
7: end for
8: return the variable with ratio $\varphi^*$.

Observe that if the ratio and one of the left (resp. right) gain are fixed, then the right (resp. left) gain can be computed analytically:

Proposition 5  Given a variable $(l, r)$ and its ratio $\varphi$, then

$$r = -\frac{\log(1 - \varphi^{-l})}{\log \varphi} \quad l = -\frac{\log(1 - \varphi^{-r})}{\log \varphi}.$$  

Proof  Direct using the definition of the polynomial $p$.

Figure 8 has been generated using this property. The leftmost point (corresponding to variable $(100, 100)$) serves as a reference, and each curve is such that its points have the same score according to one scoring function: the dashed curve corresponds to the linear function with default $\mu$, dotted to the product, and solid to the ratio. Unsurprisingly, the dashed curve is a line and the dotted curve is a parabola. Note how the product function and the ratio function match very closely when $l$ and $r$ are close to each other. This essentially means that if all variables have roughly equal left
and right gains, the product and ratio functions behave similarly. If this is not the case, then the ratio function prefers variables with unequal left and right gains compared to the product function.

6.3 The svts scoring function

Ranking the variables by ratio only makes sense for large enough gaps. Consider again the example given in Sect. 3.1, and suppose that variables (10, 10) and (2, 49) are the only candidates for branching. The ratio function would score variable (2, 49) higher, independent of the gap. However, in the SVB setting, if the gap is no larger than 40 then (10, 10) is the better variable. We therefore define a more refined scoring function that takes the gap into account. Given the gap at a given node, this scoring rule ranks the variables according to their SVB tree-sizes. We therefore call this scoring function the svts function, for “Single Variable Tree-Size”. Algorithm 2 describes the implementation of the svts function. We introduce a parameter $D$, which determines whether an approximation of the tree-size should be used, depending on the smallest depth of a leaf in the SVB B&B tree, i.e. $\lceil \frac{G}{r_l} \rceil$, given a variable $(l_i, r_i)$ and a gap $G$. Parameter $D$ effectively bounds the number of terms to add in the closed-form formula (2) to compute the SVB tree-size. If $\lceil \frac{G}{r_l} \rceil > D$ (i.e. the gap $G$ is such that the formula (2) would require computing more than $D$ terms), then the exact tree-size is computed for a smaller gap, namely $r_i D$, and an approximation of the tree-size $t(G)$ is computed approximately using the ratio. In our implementation we have set $D = 100$. Deactivating Step 5 (i.e. not filtering out dominated variables) increases the average running time of a MIP solver implementing Algorithm 2 as a scoring function in B&B by 4% (in the setup of Sect. 7.2.1). Deactivating Step 11 (i.e. not restricting the evaluation of formula (2) to “small” trees) increases the average running time by 11%.
Algorithm 2 Implementation of the svts scoring function.

1: Compute the absolute gap $G$ at the current node.
2: if $G = \infty$ then \Comment{e.g. no primal solution is known.}
3: return the variable selected by Algorithm 1. \Comment{Theorem 4}
4: end if
5: Filter out the branching candidates with dominated gains.
6: for all remaining branching candidates $i$ with gains $(l_i, r_i)$ do
7: Let $d = \lceil \frac{G}{r_i} \rceil$. \Comment{The minimum depth of the SVB tree}
8: if $d = \infty$ then
9: Set the tree-size $t_i(G) = \infty$.
10: else
11: Let $\tilde{G} = r_i \ast \min(d, D)$. \Comment{$D$ is a parameter}
12: Compute the SVB tree-size $t_i(\tilde{G})$ using formula (2).
13: if $\tilde{G} < G$ then
14: Compute the ratio $\phi_i$.
15: Compute the SVB tree-size $t_i(G) \approx \phi_i^{G-\tilde{G}} t_i(\tilde{G})$. \Comment{see (3.3)}
16: end if
17: end if
18: end for
19: if all variables $i$ have $t_i(G) = \infty$ then
20: return the variable selected by Algorithm 1.
21: else
22: return the variable $i$ with minimum $t_i(G)$.
23: end if

Note that for a large enough gap, the SVB function is equivalent to the ratio function, i.e. the rankings of variables are equal.

We next evaluate the ratio and SVB scoring function in simulations as well as MIP benchmarks.

7 Experimental results

Numerous MIP solving algorithms are developed and parametrized by a trial-and-error loop of experiments on MIP benchmarks. The major drawback of this procedure is evidently the computational burden it involves. Another significant flaw is that the instances used to develop the algorithm are often the ones used to benchmark it. This may lead to overfitting, which means that the algorithm will perform much better on these instances than on general ones. The magnitude of this problem increases with the number of parameters of a method. In an effort to mitigate these issues, we carry out simulations on a large set of random instances in Sect. 7.1. The results we obtain on these simulated instances confirm the improvements we achieve on standard MIP benchmarks in Sect. 7.2.

7.1 Numerical simulations

In order to evaluate the performance of the linear, product and ratio scoring functions detailed in Sect. 6, we run simulations in which these scoring functions are used to select variables on MVB and GVB instances. The GVB problem models the
B&B algorithm more closely than the MVB problem, but in practice it is only possible to obtain exact tree-sizes for the latter problem. We do not run simulations on the svts function, as it requires computing tree-sizes for all variables and each possible gap.

We evaluate the scoring functions on synthetically generated instances. The variable gains are generated in one of four different ways:

- **Balanced** [B] Both left and right gains are integers uniformly drawn in the interval \([1, 1000]\) (if necessary, gains are switched to ensure that the left gain is less than or equal to the right gain).
- **Unbalanced** [U] The left (resp. right) gain is an integer uniformly drawn in the interval \([1, 500]\) (resp. \([501, 1000]\)).
- **Very Unbalanced** [V] The left (resp. right) gain is an integer uniformly drawn in the interval \([1, 250]\) (resp. \([251, 1000]\)).
- **Extremely Unbalanced** [X] The left (resp. right) gain is an integer uniformly drawn in the interval \([1, 100]\) (resp. \([101, 1000]\)).

We will refer to these different ways of generating variables as different *data types*.

### 7.1.1 Simulation results for the MVB problem

Recall that in the MVB problem, one variable can be used an arbitrary number of times. In this section, two types of methods are evaluated. The first type correspond to *scoring functions* that can be or are actually implemented in MIP solvers to choose which variable to branch on (e.g. the *product* function). The second type are *problem solving algorithms*, which use knowledge of the MVB problem, and cannot directly be implemented in MIP solvers (e.g. an exact recursive algorithm). The methods of the first type are the ones we are really interested in, while the methods of the second type help evaluate the first ones. We compute the tree-sizes that result from the following methods, listed by order of increasing exactness.

- **linear** (with different values of the parameter \(\mu\)), *product* and *ratio* scoring functions. Since the ranking of variables obtained by scoring functions only depends on the left and right gains, the same variable is branched on at every node, which means the tree-size obtained by selecting variables using a scoring function corresponds to the tree-size of the SVB tree for the variable with the best score.
- **Lower Bound** (LB) for scoring functions: it is the minimum tree-size of all SVB instances, for all variables in the input of the MVB instance.
- **The exact recursive function**, as described in Sect. 4: this method produces a minimum-size tree for the instance.

Each branching strategy in this list necessarily produces trees with fewer nodes than those listed higher in the list. Note that the *linear*, *product* and *ratio* scoring functions are listed together, as it is *a priori* not clear whether any of the scoring function dominates the others. These are the scoring functions we are primarily interested in.

The other parameters have been chosen as follows:

- The number of variables for each instance is 100. Increasing the number of variables is possible, however the chances of producing a variable that dominates all others would increase.
Table 2  Simulation results on MVB

| Data | linear(μ) | product | ratio | LB |
|------|-----------|---------|-------|----|
|      | μ = 0 | 6 | 1/5 | 1/2 | 1 |
| B    | 26.2 | 8.23 | 6.23 | 3.10 | 10^{17} | 2.84 | 2.84 | 2.83 |
| wins | 88   | 93   | 99   | 8    | 1000  | 100  | 100  | 100  |
| U    | 10^{08} | 96.79 | 6.07 | 8.43 | 10^{23} | 6.60 | 3.35 | 3.14 |
| wins | 11   | 67   | 93   | 90   | 93    | 98   | 98   | 98   |
| V    | 10^{20} | 57.93 | 24.85 | 282.28 | 10^{23} | 29.82 | 5.75 | 5.75 |
| wins | 19   | 77   | 86   | 65   | 14    | 83   | 99   | 99   |
| X    | 10^{47} | 6.07 | 731.44 | 10^{5} | 10^{31} | 93.86 | 6.07 | 6.07 |
| wins | 7    | 100  | 72   | 50   | 13    | 82   | 100  | 100  |

- The gap is set at $G = 10^5$. Increasing $G$ by an order of magnitude may cause the tree-sizes to exceed the maximum number that can be encoded by a double-precision floating-point representation ($\approx 10^{307}$).
- The number of instances for each of the B, U, V and X type of variables is 100.

Table 2 provides two different performance measures for each type of data and each branching strategy. The first performance measure, “t-s”, is the percentile change in geometric means of the tree-sizes compared to the (exact) minimum tree-size. For instance, the number 26.2 (at the top-left of the table) indicates that this branching strategy yielded trees with, on average, 26.2% more nodes than the minimum-size trees. For the cases where the difference is in orders of magnitude, only the order is indicated. Note that the tree-sizes obtained in the experiments vary from around $10^{30}$ for the balanced data to up to $10^{300}$ for extremely unbalanced data. The second performance measure, “wins”, is the number of times a scoring function produces the smallest tree-size among the linear, product and ratio functions (LB is excluded as it necessarily produces smaller tree-sizes). In case of a tie, multiple scoring functions win. In Table 2 and throughout, numbers in bold font indicate the best results on a given line.

First, observe that the performance (w.r.t. tree-size) of branching strategies generally decreases as the left and right gains of the variables become less balanced. For some values of $\mu$, the linear function performs extremely badly. More interestingly, it appears that the best value of $\mu$ heavily depends on the balance of data. The value $\mu = \frac{1}{2}$ is better for balanced data (B), while for extremely unbalanced data (X), $\mu = \frac{1}{6}$ performs best, in a tie with the ratio function. For this type of data, the tie may be explained by Fig. 8, where the linear and ratio functions tend to pick similar variables for very unbalanced gains. Similarly, the tie between the product function and the ratio function for balanced data may be explained by the fact that they pick similar variables for balanced gains. The ratio function wins on 397 out of 400 instances, and is very close to the theoretical bound LB that these scoring functions can achieve. Note that both the ratio function and LB are within 7% of the actual minimum-tree size, even for the most unbalanced data.

Additional results with smaller gaps are given in “Appendix 2”.

\[ \text{ Springer} \]
7.1.2 Simulation results for the GVB problem

The experiments we carry out in this section are significantly different than those we did on MVB: indeed, in the GVB problem, each variable can only be used a given number of times on a path from the root to a leaf. As a consequence, we are not able to compute the exact minimum tree-sizes, as the state space of the recurrence equation (5) is too large. For simplicity, and to ensure that the results in this section would be different than those from Sect. 7.1, we have set multiplicities to 1. Note that the algorithms that compute the tree-sizes for scoring functions are also harder to solve than in the previous section, and the gaps $G$ used in our instances are therefore smaller. The parameters of these experiments are presented below.

- As in the previous section, the number of variables for each instance is 100, and the number of instances for each of the B, U, V and X type of variables is also 100.
- The gap takes different values for each type of data. Indeed, while a gap of e.g. $G = 5000$ can be closed with a few hundred nodes in the balanced case, it can only be closed in half of the instances for extremely unbalanced data. We therefore adapt the gap to the data. In each case, the minimum gap tested is chosen such that the tree-sizes are at least 100. The maximum gap tested is chosen such that the depth of the trees is close to 100, i.e. the number of variables, but also such that all 100 instances can be solved. The average between the minimum and the maximum gap is also tested.

Table 3 presents the simulation results, in a format similar to Table 2. Since we do not know the minimum tree-sizes in these experiments, the reference is the product function, as it is the state of the art. As a consequence, negative changes may (and do) occur. The observation we made in the previous experiments on MVB, which is that the ratio function benefits from unbalancedness, clearly carries over. Furthermore, for a given type of data, it also profits from an increase in the gap. An approximate 7% decrease in the number of nodes occurs when these two facts are taken together. Moreover, the ratio function has, by far, the largest number of wins. Note that the linear function outperforms the ratio or the product function in some cases.

A table similar to Table 3 for the MVB problem is given in “Appendix 2”.

Since the product function seems to perform slightly better for small gaps, and the ratio improves with larger gaps, the idea of changing the scoring function depending on the gap naturally arises. Table 4 provides the results of this hybrid function for the maximum gaps considered for each data type. The notation is the same as in Table 3 (including the fact that the product function is taken as a reference). The hybrid ratio-product function is parametrized by $h$, the height at which the scoring function used at a node switches from ratio to product (the height of a node is the length of the longest path to reach a leaf from this node). For example, $h = 10$ means that at nodes that would have a height of 10 or less with the ratio function, the product function is used instead. As a consequence, $h = 0$ corresponds to the ratio function, and $h = 100$ to the product function (since there are 100 variables, the depth of a leaf is at most 100).
First, observe that we do obtain a slight reduction in the number of nodes for any value of $h$ (except for data U and $h = 50$). Second, note that the optimal value for the height parameter increases as the data becomes less balanced. Third, observe how for each data type, the relative change compared to the product function is unimodal, which suggests a low variability of the results and backs up our analysis.

### 7.2 Experiments on MIP instances

We have modified the solver SCIP 3.1.1 [2] to use the ratio or the svt function, alternatively to the default product function. Besides the changes necessary for the implementation of these new scoring functions, no other change to SCIP has been made. In particular, the decision to use strong branching or pseudo-costs on a given variable and at a given node is unchanged. Furthermore, note that in SCIP, the selection of the branching variable does not solely rely on the product. There are a number

---

| Data | Gap   | $\mu$ | linear($\mu$) | product | ratio |
|------|-------|-------|---------------|---------|-------|
|      |       | 0     | $\frac{1}{6}$ | $\frac{1}{3}$ | $\frac{1}{2}$ | 1 |
| B    | 5000  | t-s   | 1.94          | 0.55    | $-0.15$ | 0.41 | 235.83 | 0.00 | $-0.03$ |
|      |       |       |               |         |         |       |         |       |         |
|      |       |       |               |         |         |       |         |       |         |
|      |       |       |               |         |         |       |         |       |         |
|      |       |       |               |         |         |       |         |       |         |
| U    | 5000  | t-s   | 50.86         | 10.63   | $-0.02$ | 4.27 | 314.73 | 0.00 | 0.17   |
|      |       |       |               |         |         |       |         |       |         |
|      |       |       |               |         |         |       |         |       |         |
|      |       |       |               |         |         |       |         |       |         |
| V    | 5000  | t-s   | 257.39        | 4.76    | 6.50    | 54.40 | 282.93 | 0.00 | $-0.91$ |
|      |       |       |               |         |         |       |         |       |         |
|      |       |       |               |         |         |       |         |       |         |
|      |       |       |               |         |         |       |         |       |         |
| X    | 2000  | t-s   | 445.42        | 8.29    | 42.80   | 63.76 | 89.95  | 0.00 | 3.15   |
|      |       |       |               |         |         |       |         |       |         |
|      |       |       |               |         |         |       |         |       |         |
|      |       |       |               |         |         |       |         |       |         |

---

| Data | Gap   | $\mu$ | linear($\mu$) | product | ratio |
|------|-------|-------|---------------|---------|-------|
|      |       | 0     | $\frac{1}{6}$ | $\frac{1}{3}$ | $\frac{1}{2}$ | 1 |
|      |       |       |               |         |         |       |         |       |         |
|      |       |       |               |         |         |       |         |       |         |
|      |       |       |               |         |         |       |         |       |         |
|      |       |       |               |         |         |       |         |       |         |
| B    | 15000 | t-s   | 21.34         | 9.37    | 1.18    | 2.36  | 1051.56| 0.00 | 0.06   |
|      |       |       |               |         |         |       |         |       |         |
|      |       |       |               |         |         |       |         |       |         |
|      |       |       |               |         |         |       |         |       |         |
| U    | 12500 | t-s   | 86.38         | 30.30   | 0.75    | 19.99 | 1562.15| 0.00 | $-0.74$|
|      |       |       |               |         |         |       |         |       |         |
|      |       |       |               |         |         |       |         |       |         |
|      |       |       |               |         |         |       |         |       |         |
| V    | 7500  | t-s   | 428.67        | 11.88   | 14.85   | 102.26| 425.88 | 0.00 | $-2.29$|
|      |       |       |               |         |         |       |         |       |         |
|      |       |       |               |         |         |       |         |       |         |
|      |       |       |               |         |         |       |         |       |         |
| X    | 3000  | t-s   | 903.02        | 6.72    | 47.56   | 73.37 | 107.32 | 0.00 | $-1.16$|
|      |       |       |               |         |         |       |         |       |         |
|      |       |       |               |         |         |       |         |       |         |
|      |       |       |               |         |         |       |         |       |         |
| X    | 4000  | t-s   | 1537.76       | 2.47    | 41.26   | 68.27 | 105.88 | 0.00 | $-6.90$|
|      |       |       |               |         |         |       |         |       |         |
|      |       |       |               |         |         |       |         |       |         |
|      |       |       |               |         |         |       |         |       |         |
of “soft” tie-breakers that can come into play if the \textit{product} scores of multiple variables are almost equal. This is referred to as hybrid branching [3]. If at a given node, the variable that the hybrid branching rule selects is not the one that maximizes the \textit{product} function, then we keep that variable and do not use the \textit{ratio} or \textit{svts} function. Finally, for the \textit{ratio}, we use the parameter $h = 10$ as defined in Sect. 7.1.2, following the results of Table 4. This essentially means that whenever a node is believed to be “close” to the leaves, the \textit{product} function is used rather than the \textit{ratio} function.

\subsection*{7.2.1 Benchmark instances}

The benchmark test set comprises the instances from MIPLIB 3.0, MIPLIB 2003 [4], and MIPLIB 2010 Benchmark [9]. These instances are the state of the art in MIP benchmarking. The instances that are at the time of writing classified as “open” have been removed, which leaves 159 instances. For feasible instances, the optimal value is provided as a primal bound, and primal heuristics are disabled (for all instances). Cuts after the root node are also disabled to reduce performance variability [12]. The disconnected components presolver is disabled as it can lead to different transformed problems, and thus possibly increased variability. Furthermore, for each instance, ten different seeds are used to create random permutations of the input (using SCIP’s internal procedure). A time limit of two hours is specified. In this setting, the testbed comprises 1590 instances, and the total running time that this experiment required on a single machine is 4 months.

Three different scoring functions are tested. The first one is the \textit{product} function, which is the default in SCIP. The second one is the \textit{ratio} function, defined in Sect. 6.2. The third one is the \textit{svts} function defined in Sect. 6.3.

Table 5 gives the result for these three scoring functions, with \textit{product} as the reference. Each line corresponds to the 10 permutations of each instance of the test set. There are three columns for each scoring function. The columns of the \textit{product} scoring function gives absolute performance measures, while the two others give...
Table 5  Comparison of scoring function on the benchmark test set

| Instance      | product ratio svts | Time (s) Nodes | Time (s) Nodes | Time (s) Nodes |
|---------------|--------------------|----------------|----------------|----------------|
| 30n20b8       | 10 521 10 +0       | 1.02           | 1.49           | +0 1.01        |
| aflow30a      | 10 36 2.5k         | +0 0.98        | 0.99           | +0 0.96        |
| aflow40b      | 10 3483 210.9k     | +0 0.68        | 0.65           | +0 0.70        |
| air04         | 10 30 7            | +0 0.96        | 1.00           | +0 1.03        |
| app1-2        | 10 845 565         | +0 0.90        | 1.01           | +0 0.89        |
| arki001       | 2 4862 1.3m        | +0 0.73        | 0.74           | +1 0.85        |
| ash608gpi-a-3col | 10 87 5        | +0 1.04        | 1.00           | +0 1.01        |
| bell5         | 10 1 1.1k          | +0 0.99        | 1.00           | +0 0.99        |
| biella1       | 10 217 2.8k        | +0 0.98        | 1.06           | +0 0.99        |
| bienst2       | 10 288 112.6k      | +0 1.05        | 0.98           | +0 1.11        |
| binkar10-1    | 10 431 135.8k      | +0 1.09        | 1.03           | +0 1.01        |
| blend2        | 10 2 222           | +0 0.99        | 1.00           | +0 1.01        |
| cap60000      | 10 8 1.8k          | +0 1.00        | 1.04           | +0 1.01        |
| csched010     | 9 6035 721.2k      | −1 0.86        | 0.86           | +0 0.90        |
| dancoint      | 2 7047 1.2m        | −1 1.00        | 0.96           | −1 1.02        |
| dcmulti       | 10 2 9             | +0 1.00        | 1.00           | +0 1.00        |
| dfn-gwin-UUM  | 10 256 62.9k       | +0 1.03        | 1.07           | +0 1.05        |
| eil33-2       | 10 103 496         | +0 0.97        | 0.99           | +0 1.11        |
| eilB101       | 10 134 4.4k        | +0 0.82        | 0.84           | +0 0.88        |
| enlight13     | 0 − −             | +10 − −        | − −            | +0 − −         |
| fast0507      | 10 129 494         | +0 0.98        | 1.18           | +0 0.99        |
| fiber         | 10 5 12            | +0 1.00        | 1.00           | +0 1.01        |
| fixnet6       | 10 5 9             | +0 1.00        | 1.00           | +0 1.01        |
| gesa2-o       | 10 4 4             | +0 1.01        | 1.00           | +0 1.01        |
| gesa2_o       | 10 4 5             | +0 0.99        | 1.00           | +0 1.00        |
| gesa3         | 10 4 20            | +0 0.99        | 1.00           | +0 0.99        |
| gesa3_o       | 10 4 9             | +0 0.99        | 1.00           | +0 1.00        |
| glass4        | 10 29 11.5k        | +0 1.26        | 1.34           | +0 1.05        |
| iis-1000-cov  | 10 2286 83.6k      | +0 0.97        | 0.99           | +0 0.97        |
| iis-pima-cov  | 10 806 6.7k        | +0 1.00        | 0.96           | +0 1.00        |
| khb05250      | 10 1 7             | +0 0.99        | 1.00           | +0 0.98        |
| l152lav       | 10 2 16            | +0 1.01        | 1.05           | +0 1.01        |
| lseu          | 10 2 1.0k          | +0 0.92        | 0.83           | +0 1.09        |
| map18         | 10 462 260         | +0 1.03        | 0.99           | +0 1.05        |
| map20         | 10 422 300         | +0 1.06        | 1.00           | +0 1.00        |
| mas74         | 10 1810 3.3m       | +0 0.87        | 0.88           | +0 0.92        |
| mas76         | 10 200 484.2k      | +0 0.83        | 0.83           | +0 0.98        |
| mcscsched     | 10 268 18.1k       | +0 0.99        | 1.01           | +0 0.98        |
| mik-250-1-100-1 | 10 1832 829.9k    | +0 1.18        | 1.17           | +0 1.30        |
| Instance | product | ratio | svts |
|----------|---------|-------|------|
|          | #      | Time (s) | Nodes | #      | Time (s) | Nodes | #      | Time (s) | Nodes |
| mine-166-5 | 10 | 30 | 356 | +0 | 1.01 | 0.95 | +0 | 1.05 | 1.13 |
| mine-90-10 | 10 | 1228 | 116.9k | +0 | 0.92 | 0.96 | +0 | 1.13 | 1.18 |
| misc03 | 10 | 2 | 125 | +0 | 0.98 | 1.02 | +0 | 0.98 | 0.97 |
| misc06 | 10 | 1 | 7 | +0 | 0.99 | 1.00 | +0 | 1.01 | 1.00 |
| misc07 | 10 | 45 | 23.8k | +0 | 1.24 | 1.28 | +0 | 1.06 | 0.99 |
| mod008 | 10 | 5 | 12 | +0 | 0.99 | 1.07 | +0 | 0.99 | 1.00 |
| mod011 | 10 | 74 | 886 | +0 | 0.97 | 0.89 | +0 | 0.98 | 0.96 |
| modglob | 10 | 1 | 20 | +0 | 1.00 | 1.04 | +0 | 1.00 | 1.01 |
| momentum2 | 1 | 4417 | 13.6k | +1 | 1.30 | 1.33 | +1 | 1.15 | 1.13 |
| msc98-ip | 10 | 1959 | 6.3k | +0 | 0.90 | 1.00 | +0 | 0.77 | 0.75 |
| mzzv11 | 10 | 732 | 233 | +0 | 0.99 | 0.89 | +0 | 0.97 | 0.89 |
| mzzv42z | 10 | 554 | 11 | +0 | 1.01 | 1.04 | +0 | 0.97 | 1.04 |
| n4-3 | 10 | 1213 | 46.6k | +0 | 1.04 | 0.99 | +0 | 1.03 | 0.95 |
| neos-1109824 | 10 | 395 | 23.2k | +0 | 0.52 | 0.37 | +0 | 0.93 | 0.85 |
| neos-1396125 | 10 | 486 | 69.9k | +0 | 1.00 | 1.01 | +0 | 0.81 | 0.81 |
| neos-476283 | 10 | 431 | 131 | +0 | 0.88 | 0.99 | +0 | 1.00 | 0.94 |
| neos-686190 | 10 | 142 | 1.8k | +0 | 1.13 | 1.22 | +0 | 1.13 | 1.18 |
| neos13 | 10 | 275 | 12 | +0 | 0.95 | 0.98 | +0 | 1.00 | 0.98 |
| neos18 | 10 | 76 | 5.3k | +0 | 1.27 | 1.59 | +0 | 1.12 | 1.16 |
| net12 | 10 | 2011 | 2.9k | +0 | 1.03 | 1.07 | +0 | 1.08 | 1.08 |
| netdiversion | 10 | 1463 | 13 | +0 | 0.97 | 0.92 | +0 | 0.95 | 0.97 |
| newdano | 5 | 5636 | 1.3m | −1 | 1.02 | 0.96 | +0 | 0.95 | 0.97 |
| noswot | 10 | 573 | 679.8k | +0 | 1.54 | 1.57 | +0 | 1.43 | 1.36 |
| ns1208400 | 10 | 414 | 174 | +0 | 1.10 | 1.20 | +0 | 1.13 | 1.33 |
| ns1688347 | 10 | 81 | 92 | +0 | 1.02 | 1.14 | +0 | 1.05 | 0.92 |
| ns1766074 | 10 | 5740 | 925.7k | +0 | 0.98 | 0.98 | +0 | 0.99 | 0.98 |
| ns1830653 | 10 | 492 | 18.6k | +0 | 0.98 | 1.03 | +0 | 0.91 | 0.92 |
| nsrand-ipx | 3 | 5004 | 584.7k | +7 | 0.33 | 0.30 | +2 | 0.50 | 0.49 |
| nw04 | 10 | 2719 | 10 | +0 | 0.84 | 1.00 | +0 | 0.93 | 1.00 |
| opm2-z7-s2 | 10 | 858 | 977 | +0 | 1.03 | 1.00 | +0 | 0.92 | 0.79 |
| p0201 | 10 | 3 | 10 | +0 | 0.99 | 1.02 | +0 | 1.00 | 1.02 |
| p2756 | 10 | 4 | 10 | +0 | 1.00 | 1.00 | +0 | 1.01 | 1.00 |
| pg5_34 | 10 | 1397 | 116.0k | +0 | 0.93 | 0.93 | +0 | 0.96 | 0.94 |
| pk1 | 10 | 123 | 325.8k | +0 | 1.13 | 1.14 | +0 | 1.09 | 1.06 |
| pp08a | 10 | 2 | 237 | +0 | 1.00 | 1.01 | +0 | 0.98 | 1.01 |
| pp08aCUTS | 10 | 2 | 153 | +0 | 0.99 | 1.02 | +0 | 0.99 | 1.05 |
| pw-myciel4 | 0 | – | – | +7 | – | – | +4 | – | – |
| qiu | 10 | 106 | 12.5k | +0 | 1.00 | 1.02 | +0 | 0.98 | 0.98 |
| qnet1 | 10 | 11 | 4 | +0 | 0.98 | 1.00 | +0 | 1.00 | 1.00 |
Table 5 continued

| Instance       | product # | product Time (s) | product Nodes | ratio # | ratio Time (s) | ratio Nodes | svts #  | svts Time (s) | svts Nodes |
|----------------|-----------|------------------|---------------|---------|----------------|-------------|--------|--------------|------------|
| qnet1_o        | 10        | 6                | 4             | +0      | 0.97           | 1.00        | +0     | 0.98         | 1.00       |
| rail507        | 10        | 130              | 507           | +0      | 1.07           | 1.17        | +0     | 1.08         | 1.19       |
| ran16x16       | 10        | 544              | 349.3k        | +0      | 0.82           | 0.82        | +0     | 0.81         | 0.78       |
| rd-rplusc-21   | 0         | -                | -             | +1      | -              | -           | +0     | -            | -          |
| reblock67      | 10        | 250              | 47.9k         | +0      | 0.98           | 1.00        | +0     | 0.93         | 0.94       |
| rmatr100-p10   | 10        | 109              | 799           | +0      | 1.03           | 1.00        | +0     | 1.02         | 0.93       |
| rmatr100-p5    | 10        | 140              | 387           | +0      | 0.96           | 0.99        | +0     | 0.95         | 0.99       |
| rmine6         | 3         | 3540             | 380.6k        | +1      | 1.10           | 1.05        | -2     | 2.00         | 2.39       |
| recII-4-11     | 10        | 419              | 4.7k          | +0      | 2.43           | 5.91        | +0     | 1.27         | 1.80       |
| roccocoC10-001000 | 1     | 4677             | 323.9k        | +1      | 0.83           | 0.81        | +1     | 0.44         | 0.50       |
| roll3000       | 9         | 3781             | 429.3k        | -2      | 1.15           | 1.18        | -3     | 0.84         | 0.83       |
| rout           | 10        | 111              | 44.6k         | +0      | 0.73           | 0.73        | +0     | 0.94         | 0.94       |
| satellites1-25 | 10        | 1590             | 3.4k          | +0      | 1.03           | 0.98        | +0     | 1.02         | 1.01       |
| set1ch         | 10        | 2                | 5             | +0      | 1.00           | 0.92        | +0     | 1.00         | 0.92       |
| sp98ic         | 10        | 4736             | 276.2k        | +0      | 0.49           | 0.42        | +0     | 0.70         | 0.58       |
| sp98ir         | 10        | 71               | 1.6k          | +0      | 0.97           | 1.16        | +0     | 0.98         | 1.15       |
| stein27        | 10        | 2                | 4.2k          | +0      | 0.98           | 0.98        | +0     | 1.00         | 0.99       |
| stein45        | 10        | 34               | 51.5k         | +0      | 0.92           | 0.98        | +0     | 0.97         | 0.99       |
| tanglegram1    | 7         | 1749             | 289           | -1      | 1.13           | 1.06        | -2     | 0.90         | 0.58       |
| tanglegram2    | 10        | 10               | 4             | +0      | 0.98           | 1.00        | +0     | 1.00         | 1.00       |
| timtab1        | 10        | 1184             | 774.5k        | +0      | 1.04           | 1.04        | +0     | 1.01         | 1.02       |
| tr12-30        | 10        | 3644             | 780.2k        | +0      | 1.02           | 1.08        | +0     | 1.00         | 1.02       |
| unitcal_7      | 9         | 3415             | 32.0k         | +1      | 0.76           | 0.75        | +0     | 0.91         | 0.87       |
| vpm2           | 10        | 2                | 180           | +0      | 1.00           | 1.05        | +0     | 0.98         | 1.01       |
| zib54-UUE      | 3         | 4914             | 413.0k        | +0      | 1.20           | 1.18        | +0     | 1.09         | 1.13       |
| Total          | 1225      | 705680           | 121.4m        | +24     | 0.92           | 0.97        | +1     | 0.94         | 0.97       |
| Geo. mean      | 98        | 1.2k             | 0.98          | 1.00    | 0.99           | 0.98        |
| Sh. geo. mean  | 148       | 2.8k             | 0.98          | 1.00    | 0.99           | 0.98        |

measures relative to product. For product, the first column provides the number of permutations solved, and, for the other two functions, the difference from the reference. Note that all 10 permutations of a given instance may not be solved the same number of times depending on the scoring function used. For each instance, we determine for each scoring function the number of permutations solved. Let $N$ be the minimum and note that $N$ can be as large as 6 and we could still have no single permutation solved by all three scoring functions. After having determined $N$ for a given instance, we take the $N$ best results for each scoring function and compute the arithmetic averages. These averages of time (in seconds) and nodes are directly displayed for the product function. In the nodes column, we use the letters $k$ and $m$ as a shorthand for thousands.
and millions. For the \textit{ratio} and \textit{svts} functions, we display the ratio of averages for both time and nodes, compared to the \textit{product}. Considering the $N$ best permutations rather than the intersection of the sets of solved permutations enables more data to be used, and reflects in the averages the fact that some scoring functions solve more permutations, giving them a fair advantage.

The instances for which at least one permutation can be solved at the root node by any setting, or for which no permutation is solved by all settings (i.e. $N = 0$), are not displayed. We similarly exclude instances that can be solved in less than a second by at least one setting and permutation. Note that these instances are accounted in the total number of instances solved, but not in the different average measures.

The total, geometric mean and shifted geometric mean (with shifts 10 and 100 for time and nodes, respectively) [-1, p. 321] are provided at the bottom of the table. These are not computed on the averages given at each line, but on each value that is used to compute the line averages (i.e. an instance solved 10 times by all settings counts more than an instance solved fewer than 10 times by all settings).

Function \textit{ratio} solves marginally more instances than the \textit{product} function. Both functions \textit{ratio} and \textit{svts} slightly outperform \textit{product}, in terms of time and nodes used. The difference in performance is especially apparent when considering the

| Instance   | \textit{product} | \textit{ratio} | \textit{svts} |
|------------|------------------|----------------|---------------|
|            | #     | Time (s) | Nodes | #     | Time (s) | Nodes | #    | Time (s) | Nodes |
| csched007  | 10    | 31000  | 5.7m  | +0   | 0.78   | 0.75  | +0   | 0.82   | 0.76  |
| csched008  | 10    | 1440   | 112.4k| +0   | 1.21   | 1.03  | +0   | 1.14   | 1.06  |
| glass4     | 10    | 28     | 11.5k | +0   | 1.37   | 1.34  | +0   | 1.03   | 1.03  |
| gmu-35-40  | 0     | –      | –     | +2   | –      | –     | +1   | –      | –     |
| k16x240    | 10    | 23835  | 23.2m | +0   | 0.88   | 0.89  | +0   | 0.86   | 0.85  |
| neos-1616732 | 10  | 9182   | 2.5m  | +0   | 0.85   | 0.94  | +0   | 0.82   | 0.84  |
| neos-942830 | 10  | 9184   | 1.5m  | +0   | 0.95   | 0.95  | +0   | 1.00   | 0.94  |
| neos15     | 1     | 34352  | 12.3m | +0   | 0.91   | 0.94  | +0   | 0.87   | 0.91  |
| neos-858960 | 10  | 3997   | 2.8m  | +0   | 0.92   | 0.93  | +0   | 0.92   | 0.93  |
| noswot     | 10    | 556    | 679.8k| +0   | 1.62   | 1.57  | +0   | 1.44   | 1.36  |
| ns1766074  | 10    | 5747   | 925.7k| +0   | 0.95   | 0.98  | +0   | 0.98   | 0.98  |
| pg         | 10    | 24     | 203   | +0   | 1.01   | 0.97  | +0   | 1.01   | 1.10  |
| ran14x18   | 10    | 23039  | 18.4m | +0   | 0.83   | 0.82  | +0   | 0.87   | 0.81  |
| reblock166 | 4     | 16858  | 1.6m  | +2   | 0.98   | 1.00  | +2   | 0.79   | 0.95  |
| timtab1    | 10    | 1142   | 774.5k| +0   | 1.05   | 1.04  | +0   | 1.02   | 1.02  |
| umts       | 10    | 8726   | 1.5m  | +0   | 0.69   | 0.71  | +0   | 0.64   | 0.65  |
| wachplan   | 10    | 15334  | 324.2k| +0   | 1.02   | 1.01  | +0   | 1.01   | 1.01  |
| Total      | 155   | 1434131| 602.1m| +4   | 0.87   | 0.87  | +3   | 0.87   | 0.84  |
| Geo. mean  | 2879  | 618.6k | 0.97  | 0.96  | 0.95   | 0.95  |
| Sh. geo. mean | 3021 | 636.4k | 0.97  | 0.96  | 0.95   | 0.95  |
total resources used, as it mostly reflects the performance on large instances, on which the scoring functions we introduce perform better. In Sect. 7.2.2, we will consider a set of instances that contains instances that experimentally require large B&B trees, and on which this observation is confirmed.

### 7.2.2 Instances with large B&B trees

MIPLIB 2010 also has a so-called tree set with 52 instances, which “contains instances that (empirically) lead to large enumeration trees” [9]. These instances have been selected because the scoring functions \( \text{ratio} \) and \( \text{svts} \) defined in Sect. 6 have specifically been designed to solve instances with large B&B trees. We have tested our scoring functions on these instances in the same setup as the experiments of Sect. 7.2.1, but we gave a time limit of 12 hours to allow for a more significant number of instances to be solved. Including permutations, the testbed comprises 520 instances, and required the equivalent of 19 months of running time on a single machine.

Table 6 provides the result for the tree test set with the same notation as Table 5 in Sect. 7.2.1. Both functions \( \text{ratio} \) and \( \text{svts} \) largely outperform \( \text{product} \), both in terms of time and nodes, and solve slightly more instances. Note that some instances in this test set do not require larger trees than those in the benchmark test set (e.g. glass4, ns894788 and pg). This is probably due to the fact that MIP solvers have made significant progress for these instances since this test set was designed. A more extensive report of the numerical experiments is provided in the online supplement.

### 8 Conclusions

We developed one of the first models for the branching component of the Branch & Bound algorithm, over fifty years after its introduction. We proved that these models are relevant by theoretically establishing new scoring functions that are efficient for MIP solving. Numerous questions naturally arise regarding these models.

One possible line of investigation relates to the computational complexity of the decision problems we have defined. For instance, can MVB be solved in polynomial time (even for two variables)? Is there an approximation algorithm for the minimization version of GVB?

Scoring functions are the only components of B&B that we numerically analyze through the theory we develop in this paper. Are there other elements of B&B that can be studied via the current models? There certainly exist decision problems that can model the B&B algorithm more accurately than GVB (e.g. if the gains of a variable are not fixed). Would the analysis of these models deepen our understanding of B&B, and lead to additional MIP solving improvements?

**Acknowledgements** The authors would like to thank Eduardo Uchoa for pointing out reference [10] and Graham Farr for helping with the discussion at the end of Sect. 5.1. The authors would like to extend their gratitude to the reviewers for the numerous improvements they contributed to through their comments.
research was funded by AFOSR grant FA9550-12-1-0151 of the Air Force Office of Scientific Research and the National Science Foundation Grant CCF-1415460 to the Georgia Institute of Technology.

Appendix 1: Proofs

Proof of Theorem 2 and Corollary 1

Theorem When $G$ tends to infinity, both sequences $\sqrt[t]{\frac{t(G+l)}{t(G)}}$ and $\sqrt[t]{\frac{t(G+r)}{t(G)}}$ converge to $\varphi$, which is the unique root greater than 1 of the equation $p(x) = x^r - x^{r-1} - 1 = 0$.

Proof Proposition 2 proves the case $l = r$, therefore we suppose throughout the proof that $r > l$. First, we define the notation

$$L = \liminf_{G \to \infty} \frac{t(G+l)}{t(G)}, \quad \tilde{L} = \limsup_{G \to \infty} \frac{t(G+l)}{t(G)}$$

$$R = \liminf_{G \to \infty} \frac{t(G+r)}{t(G)}, \quad \tilde{R} = \limsup_{G \to \infty} \frac{t(G+r)}{t(G)}.$$

Using the recursive definition (1) of $t$ between the first and the second lines, we can easily establish:

$$L = \left( \limsup_{G \to \infty} \frac{t(G)}{t(G)} \right)^{-1} = \left( \limsup_{G \to \infty} \frac{t(G)}{t(G)} \right)^{-1}$$

$$= \left( \limsup_{G \to \infty} \frac{t(G) - t(G - r)}{t(G)} \right)^{-1} = \left( 1 - \limsup_{G \to \infty} \frac{t(G - r)}{t(G)} \right)^{-1}$$

$$= \left( 1 - \limsup_{G \to \infty} \frac{t(G)}{t(G)} \right)^{-1} = \left( 1 - \left( \liminf_{G \to \infty} \frac{t(G + r)}{t(G)} \right)^{-1} \right)^{-1}$$

$$= (1 - \tilde{R})^{-1} = 1 + \frac{1}{\tilde{R} - 1}. \quad \text{(note that } \tilde{R} \geq 2.)$$

Similarly, $\tilde{L} = 1 + \frac{1}{\tilde{R} - 1}$. Next, we obtain

$$\liminf_{G \to \infty} \frac{t(G + lr)}{t(G)} = \liminf_{G \to \infty} \frac{t(G + lr)}{t(G + (l-1)r)} \frac{t(G + (l-1)r)}{t(G + (l-2)r)} \cdots \frac{t(G + r)}{t(G)}$$

$$\geq \liminf_{G \to \infty} \frac{t(G + lr)}{t(G + (l-1)r)} \liminf_{G \to \infty} \frac{t(G + (l-1)r)}{t(G + (l-2)r)} \cdots \liminf_{G \to \infty} \frac{t(G + r)}{t(G)}$$

$$= \left( \liminf_{G \to \infty} \frac{t(G + r)}{t(G)} \right)^l = R^l.$$
Also, we have

\[ \limsup_{G \to \infty} \frac{t(G + lr)}{t(G)} = \limsup_{G \to \infty} \frac{t(G + lr)}{t(G + l(r - 1))} \cdots \frac{t(G + l)}{t(G)} \leq \limsup_{G \to \infty} \frac{t(G + lr)}{t(G + l(r - 1))} \cdots \limsup_{G \to \infty} \frac{t(G + l)}{t(G)} \leq \left( \limsup_{G \to \infty} \frac{t(G + l)}{t(G)} \right)^r = \bar{L}^r. \]

Together, these yield

\[ \bar{L}^r \geq \limsup_{G \to \infty} \frac{t(G + lr)}{t(G)} \geq \liminf_{G \to \infty} \frac{t(G + lr)}{t(G)} \geq R^l. \]

Following the same steps, we can show \( \bar{L}^l \geq L^r \). Starting from this inequality, we prove

\[ \bar{L}^l \geq L = 1 + \frac{1}{R - 1} \Rightarrow R \geq 1 + \frac{1}{\bar{L}^l - 1}. \]

Likewise, we can show that the inequality \( \bar{L} \leq 1 + \frac{1}{R - 1} \) holds.

We now introduce two monotonic sequences \( \alpha_n \) and \( \omega_n \) that respectively bound \( \bar{R} \) from below and bound \( \bar{R} \) from above, and we prove that they converge to the same limit. For all non-negative \( n \), let \( \alpha_n \) and \( \omega_n \) be defined as follows:

\[ \alpha_n = \begin{cases} 2 & \text{if } n = 0 \\ f(\omega_{n-1}) & \text{otherwise} \end{cases} \]

\[ \omega_n = f(\alpha_n) \]

where \( f(x) = 1 + \frac{1}{x} \) for all \( x \in (1, \infty) \). We first prove by induction that for all non-negative integer \( n \), \( \alpha_n \) and \( \omega_n \) satisfy \( \alpha_n \leq R \leq \bar{R} \leq \omega_n \). Proposition 2 ensures that \( \alpha_0 = 2 \) is a lower bound on \( \bar{R} \). Suppose that for a given non-negative \( n \), the inequality \( \alpha_n \leq \bar{R} \) holds. We prove that \( \bar{R} \leq \omega_n \) holds too:

\[ \alpha_n \leq \bar{R} \Rightarrow 1 + \frac{1}{\alpha_n^l} \leq 1 + \frac{1}{\bar{R}^l - 1} \Rightarrow f(\alpha_n) \geq \bar{R} \Rightarrow \omega_n \geq \bar{R}. \]

The same reasoning also proves that for all non-negative integers \( n \), \( \omega_n \geq \bar{R} \) implies \( \alpha_{n+1} \leq \bar{R} \). The sequence \( \alpha_n \) thus bounds \( R \) from below, and \( \omega_n \) bounds \( \bar{R} \) from above.
We now prove that $\alpha_n$ is monotonically increasing. Consider the inequality:

$$\alpha_1 > \alpha_0 \iff f(f(2)) > 2 \iff \frac{1}{f(2)} > 1 \iff f(2) < 2^{\frac{1}{r}}$$

$$\iff 1 + \frac{1}{2^{\frac{1}{r}} - 1} < 2^{\frac{1}{r}} \iff 2^{\frac{1}{r}} < 2^{\frac{1}{r}} \iff r > l.$$ 

Suppose now that for a given non-negative $n$, $\alpha_n > \alpha_{n-1}$. This implies that $\omega_n < \omega_{n-1}$:

$$\omega_n - \omega_{n-1} = \frac{1}{\alpha_n^{\frac{1}{r}} - 1} - \frac{1}{\alpha_{n-1}^{\frac{1}{r}} - 1} < 0 \iff \alpha_n^{\frac{1}{r}} > \alpha_{n-1}^{\frac{1}{r}}.$$ 

In turn, for that given $n$, $\omega_n < \omega_{n-1}$ implies $\alpha_{n+1} > \alpha_n$. The sequence $\alpha_n$ and $\omega_n$ are thus increasing and decreasing, respectively. Since they are bounded, each of them converges to one of the solutions of the fixed-point equation $x = f(x)$. We now prove that there is a unique solution to that equation:

$$x = f(x) \iff x = 1 + \frac{1}{x^{\frac{1}{r}} - 1} \iff (x^{\frac{1}{r}} - 1)(x - 1) = 0$$

$$\iff x^{\frac{1+x}{r}} - x^{\frac{1}{r}} - x = 0$$

$$\iff p(X) = X^r - X^{r-1} - 1 = 0$$

where $X = x^{\frac{1}{r}}$. We establish in Theorem 3 that the polynomial $p$ has a unique root in $(1, \infty)$, hence the fixed-point equation $x = f(x)$ also has a unique solution. Consequently, it is necessary that both sequences $\alpha_n$ and $\omega_n$ converge to this unique fixed point, and thus $\bar{R} = \bar{R}$. Furthermore, the sequence $\sqrt{\frac{l(G+R)}{r(G)}}$ converges to the root $\varphi > 1$ of the polynomial $p$.

Since we have established that $\bar{L} = 1 + \frac{1}{\bar{R} - 1}$ and $\bar{L} = 1 + \frac{1}{\bar{R} - 1}$, it follows that $\bar{L} = \bar{L}$. Since $p(\varphi) = 0$, $\varphi$ equivalently satisfies $\varphi^l = 1 + \frac{1}{\varphi^{r-1}}$, therefore $\bar{L} = \bar{L} = \varphi^l$. 

**Corollary** A numerical approximation of $\varphi^r$ is given by the fixed-point iteration

$$f(x) = 1 + \frac{1}{x^{\frac{1}{r}} - 1}.$$ 

with the starting point $x = 2$.

**Proof** Recall the definition of the sequences $\alpha_n$ and $\omega_n$ as given in the proof of Theorem 2 (notice that function $f$ has the same definition). The sequence $f_n$ generated by the fixed-point equation is $(2, f(2), f(f(2)), f(f(f(2))), \ldots)$, which is equal to $(\alpha_0, \omega_0, \alpha_1, \omega_1, \ldots)$. Formally, the sequence $f_n$ generated by the fixed-point equation
An abstract model for branching and its application to...

satisfies

\[ f_n = \begin{cases} 
  \frac{\alpha_n}{z} & \text{if } n \text{ is even} \\
  \frac{\omega_{n+1}}{2} & \text{if } n \text{ is odd}.
\end{cases} \]

In the proof of Theorem 2, we prove that both \( \alpha_n \) and \( \omega_n \) converge to \( \varphi^r \) when \( n \) tends to infinity, therefore \( f_n \) also converges to \( \varphi^r \). \( \square \)

**Proof of Theorem 4**

Recall that \( z \) is the least common multiple of all \( l_i \) and \( r_i \).

**Theorem** \( \varphi = \min_i \varphi_i \)

**Proof** Let \( \alpha \) be such that for all positive integers \( x \leq z \),

\[ \alpha^x \leq \liminf_{G \to \infty} \frac{t(G+x)}{t(G)} \leq \limsup_{G \to \infty} \frac{t(G+x)}{t(G)}. \quad (6) \]

A possible value is \( \alpha = 1 \). Similar to the proof of Theorem 2, we use the notation

\[ Z = \liminf_{G \to \infty} \frac{t(G+z)}{t(G)} \]
\[ \bar{Z} = \limsup_{G \to \infty} \frac{t(G+z)}{t(G)}. \]

For all variables \( i \), we have:

\[ \bar{Z} = \limsup_{G \to \infty} \left( \frac{t(G+z)}{t(G)} t(G+z-l_i) \cdots \frac{t(G+z-(\frac{z}{l_i}-1)l_i)}{t(G)} \right) \]
\[ \leq \left( \limsup_{G \to \infty} \frac{t(G+l_i)}{t(G)} \right)^{\frac{z}{l_i}} \]
\[ = \left( \limsup_{G \to \infty} \frac{1 + \min_j (t(G+l_i-l_j) + t(G+l_i-r_j))}{t(G)} \right)^{\frac{z}{l_i}} \]
\[ \leq \left( 1 \limsup_{G \to \infty} \frac{t(G+l_i-r_i)}{t(G)} \right)^{\frac{z}{l_i}} \]
\[ \leq \left( 1 + \left( \liminf_{G \to \infty} \frac{t(G+l_i-r_i)}{t(G)} \right)^{-1} \right)^{\frac{z}{l_i}} \]
\[ \leq \left( 1 + \left( \liminf_{G \to \infty} \frac{t(G+r_i-l_i)}{t(G)} \right)^{-1} \right)^{\frac{z}{l_i}} \]
\[ \leq \left( 1 + \alpha^{l_i-r_i} \right)^{\frac{z}{l_i}} \]

where the last line follows from the lower bound on \( \alpha^{-x} \) from (6).
Suppose there exists a variable $i$ such that $\alpha > \varphi_i$, then we establish, using Theorem 3:

$$\alpha > \varphi_i \Rightarrow p_i(\alpha) > 0$$
$$\Rightarrow \alpha^{r_i} - \alpha^{r_i-l_i} - 1 > 0$$
$$\Rightarrow 1 + \alpha^{l_i-r_i} < \alpha^{l_i}$$
$$\Rightarrow (1 + \alpha^{l_i-r_i})^{\hat{r}_i} < (\alpha^{l_i})^{\hat{r}_i}$$
$$\Rightarrow \bar{Z} < \alpha^z.$$ 

This contradicts expression (6), hence for all variables $i$, $\alpha \leq \varphi_i$. Suppose that there exists a variable $i$ such that $\varphi_i^z < \bar{Z}$, then

$$\alpha^z \leq \varphi_i^z < \bar{Z} \leq \left(1 + \alpha^{l_i-r_i}\right)^{\hat{r}_i} \Rightarrow \alpha^{l_i} < 1 + \alpha^{l_i-r_i}$$
$$\Rightarrow \alpha^{l_i-r_i} p_i(\alpha) < 0$$
$$\Rightarrow \alpha < \varphi_i,$$

which is a contradiction, hence for all variables $i$, $\bar{Z} \leq \varphi_i^z$. In addition, for each variable $i$,

$$Z \geq \left(\liminf_{G \to \infty} \frac{t(G + r_i)}{t(G)}\right)^{\hat{r}_i}$$
$$\geq \left(\liminf_{G \to \infty} \frac{1 + \min_j (t(G + r_i - l_j) + t(G + r_i - r_j))}{t(G)}\right)^{\hat{r}_i}$$
$$\geq \left(\min_j \liminf_{G \to \infty} \frac{t(G + r_i - l_j) + t(G + r_i - r_j)}{t(G)}\right)^{\hat{r}_i}.$$ 

Hence there must exist a variable $j$ such that

$$Z \geq \left(1 + \liminf_{G \to \infty} \frac{t(G + r_j - l_j)}{t(G)}\right)^{\hat{r}_j}$$
$$\geq \left(1 + \alpha^{r_j-l_j}\right)^{\hat{r}_j}.$$ 

Suppose $\alpha < \varphi_j$, then, using Theorem 3 in the first line:

$$\alpha < \varphi_j \Rightarrow p_j(\alpha) < 0$$
$$\Rightarrow 1 + \alpha^{r_j-l_j} > \alpha^{r_j}$$
$$\Rightarrow (1 + \alpha^{r_j-l_j})^{\hat{r}_j} > (\alpha^{r_j})^{\hat{r}_j}$$
$$\Rightarrow Z > \alpha^z.$$
Since $\varphi_j^z > \alpha^z$ implies $Z > \alpha^z$, then $Z \geq \varphi_j^z$ must be true. This can be shortly proven by writing

$$(\varphi_j^z > \alpha^z \Rightarrow Z > \alpha^z) \Rightarrow \neg(\varphi_j^z > \alpha^z \land Z \leq \alpha^z) \Rightarrow \neg(\varphi_j^z > Z) \Rightarrow \varphi_j^z \leq Z.$$ 

Since $\varphi_i^z \geq \bar{Z}$ holds for all variables $i$, then variable $j$ satisfies $\varphi_j^z \leq Z \leq \bar{Z} \leq \varphi_i^z$ for all variables $i$. We can finally conclude that

$$\lim_{G \to \infty} \sqrt{t(G+z)} = \min_i \varphi_i^z.$$ 

\[\square\]

**Appendix 2: Additional numerical simulations**

We give an additional set of simulations on the MVB model, with the same target gaps as in the simulations on GVB, so that one can compare both experimental setups easily. For the same experiment, we present two different tables, Tables 7 and 8, where the

| Data | Gap   | linear ($\mu$) | product ratio | LB   |
|------|-------|----------------|---------------|------|
|      |       | $\mu$ | 0   | $\frac{1}{5}$ | $\frac{1}{3}$ | $\frac{1}{2}$ | 1   |       |
| B    | 5000  | t-s   | **0.13** | 0.15 | 0.15 | 0.46 | 517.40 | 0.16 | 0.16 | 0.11 |
|      |       | wins  | 99   | 98  | 98  | 96  | 21   | 97   | 97   |
|      | 15000 | t-s   | 6.02 | 3.49 | 3.13 | 2.80 | $10^04$ | **2.76** | **2.76** | 2.62 |
|      |       | wins  | 89   | 94  | 96  | 98  | 8    | **99** | **99** |
|      | 25000 | t-s   | 7.78 | 3.68 | 3.18 | 2.72 | $10^05$ | **2.71** | **2.71** | 2.71 |
|      |       | wins  | 88   | 93  | 95  | 99  | 8    | **100** | **100** |
| U    | 5000  | t-s   | 98.02| 2.47 | 0.96 | 0.90 | 1353.05 | 0.96 | **0.87** | 0.73 |
|      |       | wins  | 16   | 85  | 95  | 89  | 12   | **96** | 93   |
|      | 12500 | t-s   | 544.24| 10.22 | **2.76** | 3.59 | $10^04$ | 2.82 | 2.79 | 2.50 |
|      |       | wins  | 13   | 74  | **99** | 84  | 10   | 97   | 92   |
|      | 20000 | t-s   | 1860.53| 16.64 | 3.58 | 4.15 | $10^06$ | 3.80 | **3.17** | 3.09 |
|      |       | wins  | 13   | 70  | 92  | 87  | 11   | 90   | **95** |
| V    | 5000  | t-s   | 694.14| 7.43 | 7.49 | 15.47 | 1587.97 | 6.52 | **5.80** | 5.49 |
|      |       | wins  | 20   | 82  | 81  | 61  | 13   | 87   | **92** |
|      | 7500  | t-s   | 2176.92| 7.79 | 7.81 | 18.20 | 5596.92 | 6.75 | **5.89** | 5.69 |
|      |       | wins  | 19   | 81  | 83  | 64  | 14   | 86   | **94** |
|      | 10000 | t-s   | 6456.22| 9.48 | 8.17 | 22.52 | $10^04$ | 7.61 | **5.81** | 5.72 |
|      |       | wins  | 19   | 76  | 86  | 66  | 15   | 83   | **97** |
Table 7 continued

| Data | Gap | linear (μ) | product ratio | LB |
|------|-----|------------|---------------|----|
|      |     | 0 | 1/6 | 1/3 | 1/2 | 1 |
| X    | 2000 t-s | 626.55 | 5.64 | 14.14 | 36.89 | 527.77 | 5.08 | 5.64 | 4.79 |
|      | wins | 11 | 85 | 64 | 45 | 12 | 91 | 85 |
|      |     | 3000 t-s | 1943.09 | **6.21** | 16.60 | 48.57 | 1198.56 | 6.76 | **6.21** | 5.67 |
|      | wins | 9 | **88** | 68 | 47 | 12 | **88** | **88** |
|      |     | 4000 t-s | 5797.74 | **6.03** | 18.73 | 60.70 | 2449.16 | 7.30 | **6.03** | 5.68 |
|      | wins | 9 | **93** | 67 | 47 | 12 | 89 | **93** |

Table 8 Simulation results on MVB presented as Table 3

| Data | Gap | linear (μ) | product ratio |
|------|-----|------------|---------------|
|      |     | 0 | 1/6 | 1/3 | 1/2 | 1 |
| B    | 5000 t-s | **−0.03** | −0.02 | −0.02 | 0.30 | 516.40 | 0.00 | 0.00 |
|      | wins | **99** | 98 | 98 | 96 | 21 | 97 | 97 |
|      |     | 15000 t-s | 3.18 | 0.71 | 0.36 | 0.04 | 10^{04} | 0.00 | 0.00 |
|      | wins | 89 | 94 | 96 | 98 | 8 | **99** | **99** |
|      |     | 25000 t-s | 4.94 | 0.94 | 0.46 | 0.01 | 10^{05} | 0.00 | 0.00 |
|      | wins | 88 | 93 | 95 | 99 | 8 | **100** | **100** |
| U    | 5000 t-s | 96.14 | 1.50 | 0.00 | −0.06 | 1339.26 | 0.00 | **−0.09** |
|      | wins | 16 | 85 | 95 | 89 | 12 | **96** | 93 |
|      |     | 12500 t-s | 526.58 | 7.20 | **−0.06** | 0.75 | 10^{04} | 0.00 | −0.03 |
|      | wins | 13 | 74 | **99** | 84 | 10 | 97 | 92 |
|      |     | 20000 t-s | 1788.75 | 12.37 | −0.21 | 0.34 | 10^{06} | 0.00 | **−0.61** |
|      | wins | 13 | 70 | 92 | 87 | 11 | 90 | **95** |
| V    | 5000 t-s | 645.51 | 0.85 | 0.90 | 8.40 | 1484.61 | 0.00 | **−0.68** |
|      | wins | 20 | 82 | 81 | 61 | 13 | 87 | **92** |
|      |     | 7500 t-s | 2032.98 | 0.97 | 0.99 | 10.73 | 5236.79 | 0.00 | **−0.81** |
|      | wins | 19 | 81 | 83 | 64 | 14 | 86 | **94** |
|      |     | 10000 t-s | 5992.84 | 1.74 | 0.52 | 13.86 | 10^{04} | 0.00 | **−1.67** |
|      | wins | 19 | 76 | 86 | 66 | 15 | 83 | **97** |
| X    | 2000 t-s | 591.45 | 0.54 | 8.62 | 30.28 | 497.44 | **0.00** | 0.54 |
|      | wins | 11 | 85 | 64 | 45 | 12 | **91** | 85 |
|      |     | 3000 t-s | 1813.72 | **−0.52** | 9.21 | 39.16 | 1116.33 | 0.00 | **−0.52** |
|      | wins | 9 | **88** | 68 | 47 | 12 | **88** | **88** |
|      |     | 4000 t-s | 5396.50 | **−1.18** | 10.65 | 49.77 | 2275.74 | 0.00 | **−1.18** |
|      | wins | 9 | **93** | 67 | 47 | 12 | 89 | **93** |
results are presented as in Tables 2 and 3, respectively. The two differences between Tables 7 and 8 are thus the presence (or absence) of the last column LB, and the reference used for relative performance (for Table 7 it is the minimum tree-size, and for Table 8 it is the tree-size produced by product).

These results show that ratio performs generally better than product and linear, and that this phenomenon becomes more significant as the gap to close increases. If we compare Table 8 to Table 3, it appears that ratio is even more at an advantage on GVB than on MVB. One reason may be that, for a given instance, the best variable for ratio and product may be the same, or not very different, and in the MVB experiments this best variable would be branched on at every node. However, in GVB, this variable would be branched on only at the root node, and the subsequent best variables chosen by ratio and product are likely to differ. Indeed, observe that there are many more ties in the MVB experiments than in the GVB experiments.

References

1. Achterberg, T.: Constraint Integer Programming. Ph.D. thesis, Technische Universität Berlin (2007)
2. Achterberg, T.: SCIP: solving constraint integer programs. Math. Program. Comput. 1(1), 1–41 (2009)
3. Achterberg, T., Berthold, T.: Hybrid branching. In: Höve, W.J., Hooker, J.N. (eds.) Integration of AI and OR Techniques in Constraint Programming for Combinatorial Optimization Problems, Volume 5547 of Lecture Notes in Computer Science, pp. 309–311. Springer, Berlin (2009)
4. Achterberg, T., Koch, T., Martin, A.: MIPLIB 2003. Oper. Res. Lett. 34(4), 361–372 (2006)
5. Achterberg, T., Wunderling, R.: Mixed Integer Programming: Analyzing 12 Years of Progress. Springer, Berlin (2013)
6. Conforti, M., Cornuéjols, G., Zambelli, G.: Integer Programming, Volume 271 of Graduate Texts in Mathematics. Springer, Berlin (2014)
7. Cormen, T.H., Leiserson, C.E., Rivest, R.L., Stein, C.: Introduction to Algorithms, 3rd edn. MIT Press, Cambridge (2009)
8. Jacobson, N.: Basic Algebra I: Second Edition. Dover, Mineola (2009)
9. Koch, T., Achterberg, T., Andersen, E., Bastert, O., Berthold, T., Bixby, R.E., Danna, E., Gamrath, G., Gleixner, A.M., Heinz, S., Lodi, A., Mittelmann, H., Ralphs, T., Salvagnin, D., Steffy, D.E., Wolter, K.: MIPLIB 2010. Math. Program. Comput. 3(2), 103–163 (2011)
10. Kullmann, O.: Fundaments of branching heuristics. In: Biere, A., Heule, M., van Maaren, H., Walsh, T. (eds.) Handbook of Satisfiability, Volume 185 of Frontiers in Artificial Intelligence and Applications, pp. 205–244. IOS Press, Amsterdam (2009)
11. Land, A.H., Doig, A.G.: An automatic method of solving discrete programming problems. Econometrica 28(3), 497–520 (1960)
12. Lodi, A., Tramontani, A.: Performance variability in mixed-integer programming. Tuter. Oper. Res. Theory Driven Infl. Appl. 2, 1–12 (2013)
13. Nemhauser, G.L., Wolsey, L.A.: Integer and Combinatorial Optimization. Wiley, New York (1988)
14. Press, W.H., Teukolsky, S.A., Vetterling, W.T., Flannery, B.P.: Numerical Recipes 3rd Edition: The Art of Scientific Computing, 3rd edn. Cambridge University Press, New York (2007)
15. Stockmeyer, L.J.: The polynomial-time hierarchy. Theor. Comput. Sci. 3(1), 1–22 (1976)
16. Toda, S.: PP is as hard as the polynomial-time hierarchy. SIAM J. Comput. 20(5), 865–877 (1991)