Abstract Speakman and Lee analytically developed the idea of using volume as a measure for comparing relaxations in the context of spatial branch-and-bound. Specifically, for trilinear monomials, they analytically compared the three possible “double-McCormick relaxations” with the tight convex-hull relaxation. Here, again using volume as a measure, for the convex-hull relaxation of trilinear monomials, we establish simple rules for determining the optimal branching variable and optimal branching point. Additionally, we compare our results with current practice in software.

1 Introduction

The spatial branch-and-bound (sBB) algorithm (see [1], [13], [15] for example) aims to find globally-optimal solutions of factorable mathematical-optimization formulations via a divide-and-conquer approach (see [8]). The algorithm works by introducing auxiliary variables in such a way as to decompose every function of the original formulation as a labeled directed graph (DAG). Leaves correspond to original model variables, and we assume that the domain of each such model variable is a finite interval. The out-degree of each internal node, labeled by a library function \( f \in F \), is typically small (say \( d_f \leq 3 \), for all \( f \in F \)). We assume that we have methods for convexifying each low-dimensional library function \( f \) on an arbitrary box domain in \( R^{d_f} \). From these DAGs, relaxations are composed and refined (see

* This work extends and presents parts of the first author’s doctoral dissertation [19], and it corrects results first announced in the short abstract [16].

E. Speakman
Dept. of Industrial and Operations Engineering, University of Michigan, Ann Arbor.
E-mail: eespeakm@umich.edu

J. Lee
Dept. of Industrial and Operations Engineering, University of Michigan, Ann Arbor.
E-mail: jonxlee@umich.edu
For a given function $f$, the associated DAG can be constructed in more than one way, and therefore sBB has choices to make in this regard. Such choices can have a strong impact on the quality of the convex relaxation obtained from the formulation. Because sBB obtains bounds from these convex relaxations, this choice can have a significant impact on the performance of the algorithm.

There has been substantial research on how to obtain good-quality convex relaxations (many references can be found in [4]), and some consideration has been given to constructing DAGs in a favorable way. In particular, in [17], we obtained analytic results regarding the convexifications obtained from different ways of treating trilinear monomials, $f = x_1 x_2 x_3$, on non-negative box domains $\{(x_1, x_2, x_3) : x_i \in [a_i, b_i], i = 1, 2, 3\}$. We computed both the extreme point and inequality representations of the alternative relaxations (derived from iterating McCormick inequalities) and calculated their 4-dimensional volumes (in the space of $(f, x_i, x_j, x_k) \in \mathbb{R}^4$) as a comparison measure. Using volume as a measure gives a way to analytically compare formulations and corresponds to a uniform distribution of the optimal solution across a relaxation; when concerned with non-linear optimization, such a uniform measure is quite natural. Experimental corroboration for using volume as a measure of the quality of relaxations for trilinear monomials appears in [18] (also see [4], concerning quadrilinear monomials).

Along with utilizing good convex relaxations, other important issues in the effective implementation of sBB are: (i) the choice of branching variable, and (ii) the selection of the branching point. There has been extensive computational research into branching-point selection (e.g., see [3]). It is common practice for solvers to branch on the value of the variable at the current solution, adjusted using some method to ensure that the branching point is not too close to either of the interval endpoints. Often this is done by taking a convex combination of the interval midpoint and the variable at the current solution, and/or restricting the branching choice to a central part of the interval. For example, in [5] (also see [3]), they suggest branching at the current relaxation point when it is in the middle 60% of the interval and failing that, branch at the midpoint. The ooOPS software, see [2], uses the solution of an upper-bounding problem as a reference solution, if such a solution is found; otherwise the solution of the lower-bounding relaxation is used as a reference solution. ooOPS then identifies the non-convex term with the greatest error with respect to its convex relaxation. The branching variable is then chosen as the variable whose value at the reference solution is nearest to the midpoint of its range. But it is not clear how ooOPS then chooses the branching point.

[20] describe a typical way to avoid the interval endpoints by choosing the branching point as

$$\max \left\{ l_i + \beta (a_i - b_i), \min \left\{ b_i - \beta (b_i - a_i), \alpha \tilde{x}_i + (1 - \alpha) (a_i + b_i)/2 \right\} \right\},$$

(1)

where $\tilde{x}_i$ is the value of the branching variable $x_i$ at the current solution, and $b_i$ (resp., $a_i$) is the current upper (lower) bound of variable $x_i$. The constants $\alpha \in [0, 1]$ and $\beta \in [0, 1/2]$ are algorithm parameters. So, the branching point is the closest point in $[l_i + \beta (a_i - b_i), b_i - \beta (b_i - a_i)]$ to the weighted combination $\alpha \tilde{x}_i + (1 - \alpha) (a_i + b_i)/2$ (of the current value and the interval midpoint), thus explicitly ruling out branching in the bottom and top $\beta$
fraction of the interval. Note that if $\beta \leq (1 - \alpha)/2$, then there is no such explicit restriction, because already the weighted combination $\alpha \hat{x}_i + (1 - \alpha)(a_i + b_i)/2$ precludes branching in the bottom and top $(1 - \alpha)/2$ fraction of the interval.

Current available software use a variety of values for the parameters $\alpha$ and $\beta$. The method (mostly) employed by SCIP (see [21] and the open-source code itself) is to select the branching point as the closest point in the middle 60% of the interval to the variable value $\hat{x}_i$. This is equivalent to setting $\alpha = 1$ and $\beta = 0.2$ and gives an explicit restriction via the choice of $\beta$. The current default settings of ANTIGONE ([11] and [12]), BARON ([14]) and COUENNE (see [3] and the open-source code itself) all have $\beta \leq (1 - \alpha)/2$, and so the default branching point is simply the weighted combination $\alpha \hat{x}_i + (1 - \alpha)(a_i + b_i)/2$; see Figure 1.

<table>
<thead>
<tr>
<th>Solver</th>
<th>$\alpha$</th>
<th>$\beta$</th>
<th>Condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>SCIP</td>
<td>1.00</td>
<td>0.20</td>
<td>$\leq (1 - \alpha)/2 = 0.000$</td>
</tr>
<tr>
<td>ANTIGONE</td>
<td>0.75</td>
<td>0.10</td>
<td>$\leq (1 - \alpha)/2 = 0.125$</td>
</tr>
<tr>
<td>BARON</td>
<td>0.70</td>
<td>0.01</td>
<td>$\leq (1 - \alpha)/2 = 0.150$</td>
</tr>
<tr>
<td>COUENNE</td>
<td>0.25</td>
<td>0.20</td>
<td>$\leq (1 - \alpha)/2 = 0.375$</td>
</tr>
</tbody>
</table>

Fig. 1 Default parameter settings

The different choices are based on combinations of intuition and substantial empirical evidence gathered by the software developers. We note that there is considerable variation in the settings of these parameters, across the various software packages. Furthermore, there are other factors (especially in BARON) that sometimes supersede selecting a branching point according to the formula (1); in particular, functional forms involved, the solution of the current relaxation, available incumbent solutions, complementarity considerations, etc. Our work is based solely on analyzing the volumes of child relaxations of graphs of a single trilinear monomial, after branching on a variable in that trilinear monomial, with the goal of helping to inform and in some cases mathematically support the choice of a branching point. Of course variables often appear in multiple functions. So, when deciding on a branching variable or a branching point, we may obtain conflicting guidance. But this is an issue with most branching rules, including those developed empirically, and it is always a challenge to find good ways to combine local information to make algorithmic decisions. We hope that our results can help inform such decisions. For example, taking weighted averages of scores based on our metric would be a reasonable way to proceed.

In our work, we focus on trilinear monomials; that is, functions of the form $f = x_1 x_2 x_3$, where each $x_i$ is a simple variable. This is an important class of functions for sBB, because such monomials may also involve auxiliary variables. This means that whenever a formulation contains the product of three (or more) expressions (possibly complex themselves), our results apply. In addition, the case of non-zero lower bounds is particularly important; even if many model variables have lower bounds of zero, this will not typically be the case for an auxiliary variable. Furthermore, after branching, the lower bound of a variable will always be nonzero for at least one of the two children.
Following [17], for the variables \( x_i \in [a_i, b_i], i = 1, 2, 3 \), we assume the following conditions hold:

\[
0 \leq a_i < b_i \quad \text{for} \quad i = 1, 2, 3, \quad \text{and} \quad a_1 b_2 b_3 + b_1 a_2 a_3 \leq b_1 a_2 b_3 + a_1 b_2 a_3 \leq b_1 b_2 a_3 + a_1 a_2 b_3. \tag{Ω}
\]

To see that the latter two inequalities are without loss of generality, let \( O_i := a_i (b_j b_k) + b_i (a_j a_k), \) for \( i = 1, 2, 3 \). Then we can construct a labeling such that \( O_1 \leq O_2 \leq O_3 \). Note that because we are only considering non-negative bounds, the latter part of this condition is equivalent to:

\[
\frac{a_1}{b_1} \leq \frac{a_2}{b_2} \leq \frac{a_3}{b_3}.
\]

This condition also arises in the complete characterization of the inequality description for the (polyhedral) convex hull of the trilinear monomial \( f = x_1 x_2 x_3 \) (see [10] and [9]). We introduce the following notation for the hull:

\[
\mathcal{P}_h := \text{conv} \left\{ (f, x_1, x_2, x_3) \in \mathbb{R}^4 : f = x_1 x_2 x_3, \; x_i \in [a_i, b_i], \; i = 1 \ldots 3 \right\}.
\]

The extreme points of \( \mathcal{P}_h \) are the eight points that correspond to the \( 2^3 = 8 \) choices of each \( x \)-variable at its upper or lower bound (see [9] for a proof). We label these eight points (all of the form \( [f = x_1 x_2 x_3, x_1, x_2, x_3] \)) as follows:

\[
\begin{align*}
v^1 &:= \begin{bmatrix} b_1 a_2 a_3 \hline b_1 \hline a_2 \hline a_3 \end{bmatrix}, & v^2 &:= \begin{bmatrix} a_1 a_2 a_3 \hline b_1 \hline a_2 \hline a_3 \end{bmatrix}, & v^3 &:= \begin{bmatrix} a_1 a_2 b_3 \hline a_1 \hline a_2 \hline b_3 \end{bmatrix}, & v^4 &:= \begin{bmatrix} a_1 b_2 a_3 \hline a_1 \hline b_2 \hline a_3 \end{bmatrix}, \\
v^5 &:= \begin{bmatrix} a_1 b_2 b_3 \hline a_1 \hline b_2 \hline b_3 \end{bmatrix}, & v^6 &:= \begin{bmatrix} b_1 b_2 a_3 \hline b_1 \hline b_2 \hline a_3 \end{bmatrix}, & v^7 &:= \begin{bmatrix} b_1 b_2 a_3 \hline b_1 \hline b_2 \hline a_3 \end{bmatrix}, & v^8 &:= \begin{bmatrix} b_1 a_2 b_3 \hline b_1 \hline a_2 \hline b_3 \end{bmatrix}.
\end{align*}
\]

The (complicated) inequality description of the convex hull (see [10] and [9]) is directly used by some global-optimization software (e.g., BARON and ANTIGONE). However, other software packages (e.g., COUENNE and SCIP) instead use McCormick inequalities iteratively to obtain a (simpler) convex relaxation for trilinear monomials. These alternative approaches reflect the tradeoff between using a more complicated but stronger convexification and a simpler but weaker one, especially in the context of global optimization (see [7], for example).

From [17], we have a formula for the volume of the convex-hull relaxation (additionally, for the various double-McCormick relaxations), parameterized in terms of the upper and lower variable bounds.

**Theorem 1 (see [17])** Under [9] we have

\[
\text{vol}(\mathcal{P}_h) = (b_1 - a_1)(b_2 - a_2)(b_3 - a_3) \times \\
(b_1(5b_2b_3 - a_2b_3 - b_2a_3 - 3a_2a_3) + a_1(5a_2a_3 - b_2a_3 - a_2b_3 - 3b_2b_3))/24.
\]
In the context of branching within sBB, let \( c_i \in [a_i, b_i] \) be the branching point of variable \( x_i \). We obtain two children. By substituting \( a_i = c_i \) and \( b_i = c_i \) for a given variable \( x_i \), into the appropriate formula, and summing the results, we obtain the total resulting volume of the children, given that we branch on variable \( x_i \) at point \( c_i \). It is important to notice that the role of \( x_1 \) in Theorem 1 is special. That is, interchanging \( x_2 \) and \( x_3 \) does not affect the computation in general, but this is not the case with any other relabeling. So, in applying Theorem 1 to the children after branching, we must carefully relabel according to \( \Omega \).

In \[2\], we present our results analyzing optimal branching-point selection for \( x_1 \). Then, in \[3\], we present the analysis for \( x_2 \) and \( x_3 \). Due to a reason that will later become clear (and was alluded to in the previous paragraph), the analysis (and even the result) is significantly simpler than for \( x_1 \). In \[4\], we analyze branching-variable selection, and in particular, we demonstrate that it is always best to branch on \( x_1 \). In \[5\], we make some concluding remarks. In an appendix, \[6\] we provide proofs of various technical results that we utilize.

### 2 Branching on \( x_1 \)

First, we define the following quantities:

\[
q_1 := \frac{3a_1a_2a_3 + a_1a_2b_3 - a_1b_2a_3 - 3a_1b_2b_3 + 4b_1a_2a_3 - 4b_1b_2b_3}{2(3a_2a_3 + a_2b_3 - 4b_2b_3)};
\]

\[
q_2 := \frac{a_1 + b_1}{2}.
\]

\[
q_3 := \frac{4a_1a_2a_3 - 4a_1b_2b_3 + 3b_1a_2a_3 + b_1a_2b_3 - b_1b_2a_3 - 3b_1b_2b_3}{2(4a_2a_3 - 2a_3 - 3b_2b_3)}.
\]

Next, we refer to Figure 2 which depicts a procedure for choosing a branching point when branching on variable \( x_1 \). Note that \( q_1 \) is not used in the procedure, but it is used in the analysis of the procedure.

**Theorem 2** Given initial bounds \( a_1, b_1, a_2, b_2, a_3, b_3 \) that satisfy (1) and given that we branch on \( x_1 \), the procedure of Figure 2 gives the optimal branching point with respect to minimizing the sum of the volumes of the two child convex-hull relaxations.

**Proof** First, consider what happens when we pick a branching variable \( x_i \), and branch at a given point \( c_i \): we obtain two children, now with different bounds on the branching variable. The upper bound of the branching variable in the left child becomes the value of the branching point, as does the lower bound of the branching variable in the right child. That is, the domain of \( x_i \) for the left child is \([a_i, c_i]\), and the domain of \( x_i \) for the right child is \([c_i, b_i]\). We reconvexify the two children using our chosen method of convexification (i.e., the convex hull), and we can sum the volumes from both children to obtain the total volume when branching at that given point. We are interested in finding the branching point that leads to the least total volume. For an example of this principle in a lower dimension, see the diagram of Figure 3 which illustrates reconvexifying after branching in sBB.
Fig. 2 Procedure 1: output is the optimal branching point when branching on variable $x_1$

In the context of this diagram, we wish to find the branching point that minimizes the sum of the areas of the two green regions. Clearly this depends on the choice of convexification method.

We can compute the volume of the relaxation for each of the children using Theorem 1 (i.e., Theorem 4.1 from [17]). To ensure that we compute the appropriate volumes, we need to check that as the bounds on the branching variable change, we still respect the labeling $\Omega$. To illustrate this, consider the left child obtained by branching on variable $x_1$ at some point $c_1 \in [a_1, b_1]$. For this left child, the lower bound on the branching variable remains the same and the new upper bound is $c_1$. Intuitively, we can see that if $c_1$ is close enough to $b_1$, then $\Omega$ will remain satisfied, however as $c_1$ decreases, there comes a point where the labeling must change. By simple algebra, we calculate that this critical point is at $c_1 = \frac{a_1 b_2}{a_2}$ (assuming for now that $a_2 > 0$). We can consider the right child in the same manner. On the right, the upper bound on the branching variable remains
the same, and the new lower bound is $c_1$. When $c_1$ is close to $a_1$, $Ω$ will remain satisfied; however, as $c_1$ becomes larger, eventually the labeling must change. This critical point for the right child is at $c_1 = \frac{b_1a_2}{b_2}$.

Therefore, it is natural to think about two cases. First when

$$\frac{b_1a_2}{b_2} \leq \frac{a_1b_2}{a_2} \iff \frac{a_2^2}{b_2^2} \leq \frac{a_1}{b_1} \iff b_1a_2^2 \leq a_1b_2^2,$$

and second when

$$\frac{b_1a_2}{b_2} > \frac{a_1b_2}{a_2} \iff \frac{a_2^2}{b_2^2} > \frac{a_1}{b_1} \iff b_1a_2^2 > a_1b_2^2.$$

The case of equality, i.e., $\frac{b_1a_2}{b_2} = \frac{a_1b_2}{a_2}$, is arbitrarily included with Case 1. In fact, when equality holds, the analysis that follows is simplified, and it could be contained in either of the cases.

For an illustration of when the labeling must change on one or both of the intervals to ensure that $Ω$ remains satisfied, see Figure 5. Finally, we note that we must consider separately what happens when $a_2 = 0$ because when this happens, our case analysis involves division by zero.

We note that because of the structure of the volume function of the convex hull, (see Theorem 1), the second and third variables are interchangeable. This means that we do not need to consider what happens when the bounds vary enough for $x_1$ to be relabeled as $x_3$. We complete the analysis by considering the two cases described in the previous section, however, we first briefly deal with the $a_2 = 0$ case.
2.1 Case 0: $a_2 = 0$

From the condition, we know that $a_2 = 0 \Rightarrow a_1 = 0$. In this special case, the labeling for the left child does not change no matter how small the upper bound becomes. Conversely, the labeling for the right child changes as soon as the lower bound becomes positive. We therefore have the picture shown in Figure 4, and so we only have one function to consider over the entire domain, $[a_1, b_1]$. As we will see shortly, this function is a convex quadratic, and therefore it is easy to check that in this special case the minimizer of this function, $q_3$, is the minimizer of the total volume of the two children. Furthermore, when $a_2 = 0$ (and therefore $a_1 = 0$), this minimizer simplifies to $\frac{b_1}{a_2} = \frac{a_1 + b_1}{b_2} = q_2$, the midpoint of the interval.
2.2 Case 1: \( \frac{b_1a_2}{a_2} \leq \frac{a_1b_2}{a_2} \)

We define
\[
V(l_1, u_1, l_2, u_2, l_3, u_3) := (u_1 - l_1)(u_2 - l_2)(u_3 - l_3)
\times (u_1(5u_2u_3 - l_2u_3 - u_2l_3 - 3l_2l_3) + l_1(5l_2l_3 - u_2l_3 - l_2u_3 - 3a_2u_3)) / 24
\]
to be the volume of the convex hull with variable lower bounds \( l_i \) and upper bounds, \( u_i \), for \( i = 1 \ldots 3 \).

Then, for a given problem with initial upper and lower bounds \( (a_1, b_1, a_2, b_2, a_3, b_3) \), the total volume of the two children after branching at point \( c_1 \), is given by the following parameterized function:

\[
TV(c_1) := \begin{cases} 
V_1(c_1), & a_1 \leq c_1 \leq \frac{b_1a_2}{a_2}; \\
V_2(c_1), & \frac{b_1a_2}{a_2} \leq c_1 \leq \frac{a_1b_2}{a_2}; \\
V_3(c_1), & \frac{a_1b_2}{a_2} \leq c_1 \leq b_1,
\end{cases}
\]

(2)

where:

\[
V_1(c_1) := V(a_2, b_2, a_1, c_1, a_3, b_3) + V(c_1, b_1, a_2, b_2, a_3, b_3),
\]

\[
V_2(c_1) := V(a_2, b_2, a_1, c_1, a_3, b_3) + V(a_2, b_2, c_1, b_1, a_3, b_3),
\]

\[
V_3(c_1) := V(a_1, c_1, a_2, b_2, a_3, b_3) + V(a_2, b_2, c_1, b_1, a_3, b_3).
\]

This is a piecewise-quadratic function in \( c_1 \). We can easily observe this by noticing that \( V \) is the product of a pair of multilinear functions in the parameters.

It is straightforward to check that the function is continuous over its domain. Furthermore, by observing that the leading coefficient of each piece is positive for all parameter values satisfying \( m \) we conclude that each piece is strictly convex. We are able to claim strict convexity because we assume \( b_i > a_i \) for all \( i \).

The leading coefficient of \( V_1(c_1) \) is:

\[
\frac{(b_1 - a_3)(b_2 - a_2)(6b_2b_3 - a_2a_3) + 2b_3(b_2 - a_2)}{24} > 0.
\]

The leading coefficient of \( V_2(c_1) \) is:

\[
\frac{(b_3 - a_3)(b_2 - a_2)(4b_2b_3 - a_2a_3) + 2(b_3 + a_3)(b_2 - a_2)}{24} > 0.
\]

The leading coefficient of \( V_3(c_1) \) is:

\[
\frac{(b_1 - a_3)(b_2 - a_2)(6b_2b_3 - a_2a_3) + 2a_3(b_2 - a_2)}{24} > 0.
\]

Figure \( m \) gives some idea of what this function apparently could look like.

Now that we know that \( TV(c_1) \) has this structure, to find the minimizer over the domain \([a_1, b_1]\), we can simply find the minimizer on each of the three pieces and pick the point with the least function value. Because we have convex functions, the minimum of a given piece will either occur at the global minimizer of the function (if this occurs over the appropriate subdomain), or at one of the end points of the subdomain. Therefore, to find the minimizer for a given segment, we first find the minimizer of the function over the entire real line and check if it occurs in the interval; if so, it is the minimizer, if not, we examine the interval.
end points to locate the minimizer. We can then compare the function value of the minimizer of each of the three pieces to find the minimizer of $TV(c_1)$, i.e., the branching point that obtains the least total volume.

We compute the following:

The minimum of $V_1(c_1)$ occurs at:

$$ c_1 = \frac{3a_1a_2a_3 + a_1a_2b_3 - a_1b_2a_3 - 3a_1b_2b_3 + 4b_1a_2a_3 - 4b_1b_2b_3}{2(3a_2a_3 + a_2b_3 - 4b_2b_3)} = q_1. $$

The minimum of $V_2(c_1)$ occurs at:

$$ c_1 = \frac{a_1 + b_1}{2} = q_2. $$

The minimum of $V_3(c_1)$ occurs at:

$$ c_1 = \frac{4a_1a_2a_3 - 4a_1b_2b_3 + 3b_1a_2a_3 + b_1a_2b_3 - b_1b_2a_3 - 3b_1b_2b_3}{2(4a_2a_3 - b_2a_3 - 3b_2b_3)} = q_3. $$

Therefore, the candidate points for the minimizer are $a_1$, $\frac{b_1a_2}{b_2}$, $\frac{a_1b_2}{a_2}$, $b_1$, $q_1$, $q_2$ and $q_3$. We can immediately discard $a_1$ and $b_1$ because these are both equivalent to not branching. By branching and reconvexifying over the two children, we can never do worse with regard to volume. Therefore, we have five points to consider. For a given set of parameters, it is straightforward to evaluate and check which of these five points is the minimizer. However, making use of the following observations, we can further reduce the possibilities.

If $q_1$ were to be the global minimizer, then it must fall in the appropriate subdomain; i.e., it must be that $q_1 \leq \frac{b_1a_2}{b_2}$. However, by Lemma 3 (see §6), in Case 1 we always have $q_1 \geq \frac{b_1a_2}{b_2}$. Therefore, we can discard $q_1$ as a candidate point for the minimizer because for it to be the minimizer, this quantity would have to be exactly equal to $\frac{b_1a_2}{b_2}$, which is already on the list of candidate points.

Now, consider the quantities:

$$ q_1 - \frac{a_1 + b_1}{2} = \frac{(b_1 - a_1)(b_1a_2 - a_1b_2)}{2(b_2b_3 - a_2b_3 - 3a_2a_3)} \geq 0, \quad (3) $$
and
\[ q_3 = \frac{a_1 + b_1}{2} = \frac{(a_3 - b_3)(b_1 a_2 - a_1 b_2)}{2(3b_2 b_3 + b_2 a_3 - 4a_2 a_3)} \leq 0. \tag{4} \]

We therefore have:
\[ q_1 \geq q_2 \geq q_3 = \frac{a_1 + b_1}{2} \geq q_3. \tag{5} \]

From this, we can observe that if \( q_3 \geq \frac{a_1 + b_1}{2} \), then \( q_2 \geq q_3 \geq \frac{a_1 + b_1}{2} \), and therefore \( q_3 \) is the minimizer. This is because neither \( q_1 \) nor \( q_2 \) fall in their key intervals (i.e. in the appropriate subdomain); furthermore, by the definition of \( q_3 \) as the minimizer of \( V_3 \), we must have that \( V_3(q_3) \leq V_3 \left( \frac{a_1 + b_1}{2} \right) \), and by Lemma 1 (see [6]), we have that \( V_3 \left( \frac{a_1 + b_1}{2} \right) \leq V_3 \left( \frac{a_1 b_2}{a_2^2} \right) \).

If this does not occur, i.e. \( q_3 < \frac{a_1 + b_1}{2} \), then if \( \frac{b_1 a_2}{a_2^2} \leq \frac{a_1 + b_1}{2} < \frac{a_1 b_2}{a_2^2} \), the midpoint \( q_2 \) is the minimizer. This is because under these conditions, \( q_2 \) is the only minimizer that occurs in the ‘correct’ function piece, and by definition of \( q_2 \) as the minimizer of \( V_2 \), the function value is not more than at either of the end points.

Otherwise, if none of the above occurs (i.e., none of the intervals contain their function global minimizer), we have that \( \frac{a_1 b_2}{a_2^2} \) is the minimizer by Lemma 2.

As an interesting side point, we note that if it were possible to have \( q_1 \leq \frac{b_1 a_2}{a_2^2} \), then \( q_3 \leq q_2 \leq q_1 \leq \frac{b_1 a_2}{a_2^2} \), and therefore \( q_1 \) would be the minimizer. This is because neither \( q_2 \) nor \( q_3 \) would fall in their key intervals; furthermore, by the definition of \( q_1 \) as the minimizer of \( V_1 \), we have that \( V_1(q_1) \leq V_1 \left( \frac{b_1 a_2}{a_2^2} \right) \), and by Proposition 1 (see [6]) we know that \( V_1(q_1) \leq V_2 \left( \frac{a_1 b_2}{a_2^2} \right) \). However, by Lemma 3 (see [6]) we have already discarded this case.

2.3 Case 2: \( \frac{b_1 a_2}{a_2^2} > \frac{a_1 b_2}{a_2^2} \)

In this second case, for a given problem with initial upper and lower bounds \((a_1, b_1, a_2, b_2, a_3, b_3)\), the total volume of the two children after branching at point \( c_1 \), is given by the following parameterized function (this is similar, but distinct, from the function in Case 1):
\[
\widetilde{TV}(c_1) := \begin{cases} 
V_1(c_1) & a_1 \leq c_1 \leq \frac{b_1 a_2}{a_2^2}; \\
V_4(c_1) & \frac{b_1 a_2}{a_2^2} \leq c_1 \leq \frac{a_1 b_2}{a_2^2}; \\
V_3(c_1) & \frac{a_1 b_2}{a_2^2} \leq c_1 \leq b_1,
\end{cases} \tag{6}
\]

where \( V_1(c_1) \) and \( V_3(c_1) \) are defined as before and:
\[
V_4(c_1) := V(a_1, c_1, a_2, b_2, a_3, b_3) + V(c_1, b_1, a_2, b_2, a_3, b_3).
\]

Again, this is a piecewise-quadratic function in \( c_1 \), and it is simple to check that the function is continuous over its domain. Furthermore, by observing that the leading coefficient of each piece is positive for all parameter values satisfying [6] we know that each piece is (strictly) convex.

The leading coefficient of \( V_4(c_1) \) is:
\[
\frac{8(b_3 - a_3)(b_2 - a_2)(b_2 b_3 - a_2 a_3)}{24} > 0.
\]
Therefore, we can take the same approach as before to find the minimizer: first find the minimizer for each segment. We do this by finding the minimizer for the appropriate function over the whole real line and checking if it occurs in the segment. If it does, we have found the minimizer for that segment, if not, we examine the interval end points. We then compare the minimum in each of the three sections to find the branching point that obtains the least total volume.

From our analysis of Case 1, we know that the minimums of $V_1(c_1)$ and $V_3(c_1)$ occur at $q_1$ and $q_3$ respectively. We compute that the minimum of $V_4(c_1)$ occurs at the midpoint of the whole interval, i.e., at

$$c_1 = \frac{a_1 + b_1}{2} = q_2.$$

As before, the candidate points for the minimizer are $\frac{b_1 a_2}{b_2}$, $\frac{a_1 b_2}{a_2}$, $q_1$, $q_2$ and $q_3$. However, by making the following observations we can further reduce the points we need to examine.

If $q_1$ were to be the global minimizer, then it must fall in the appropriate subdomain, i.e., it must be that $q_1 \leq \frac{a_1 b_2}{a_2}$. However, by Lemma 2 (see §6), in Case 2 we always have $q_1 \geq \frac{a_1 b_2}{a_2}$. Therefore, we can discard $q_1$ as a candidate point for the minimizer because for it to be the minimizer it would have to be exactly equal to $\frac{a_1 b_2}{a_2}$, which is already on the list of candidate points.

If $q_3 \geq \frac{b_1 a_2}{b_2}$, then $q_2 \geq q_3 \geq \frac{a_1 b_3}{b_3}$, and therefore $q_3$ is the minimizer. This is because neither $q_1$ nor $q_2$ fall in their key intervals; furthermore, by definition of $q_3$ as the minimizer of $V_3$, we must have that $V_3(q_3) \leq V_3 \left( \frac{b_1 a_2}{b_2} \right)$, and by Lemma 2 (see §6) we know that $V_3 \left( \frac{b_1 a_2}{b_2} \right) \leq V_1 \left( \frac{a_1 b_2}{a_2} \right)$.

If this does not occur, i.e. $q_3 < \frac{b_1 a_2}{b_2}$, then if $\frac{a_1 b_2}{a_2} \leq q_3 \leq \frac{a_1 b_3}{b_3}$, the midpoint $q_2$ is the minimizer. This is because under these conditions, $q_2$ is the only minimizer that occurs in the ‘correct’ function piece, and by definition of $q_2$ as the minimizer of $V_4$, the function value is no more than at either of the end points.

Otherwise, we have that $\frac{b_1 a_2}{b_2}$ is the minimizer by Lemma 2 (see §6).

As another interesting side point, we also note that if it were possible to have $q_1 \leq \frac{a_1 b_2}{a_2}$, then $q_3 \leq q_2 \leq q_1 \leq \frac{a_1 b_3}{b_3}$, and $q_1$ would be the minimizer. This is because neither $q_2$ nor $q_3$ would fall in their key intervals. Furthermore, by definition of $q_1$ as the minimizer of $V_1$, we must have that $V_1(q_1) \leq V_1 \left( \frac{a_1 b_2}{a_2} \right)$, and by Proposition 2 (see §6) we know that $V_1(q_1) \leq V_4 \left( \frac{b_1 a_2}{b_2} \right)$. However, by Lemma 4 (see §6) we have already discarded this case.

\[\Box\]

2.4 Some examples

We can illustrate these piecewise-quadratic functions for the possible outcomes of Procedure 1. In this illustration, we focus on Case 1, and therefore Figure 7 shows the function $TV(c_1)$ over the domain $[a_1, b_1]$. The red curve illustrates an example where the minimizer of $V_3(c_1)$, (i.e. $q_3$), falls in the relevant interval, and therefore is the minimizer over our whole domain. The blue curve illustrates an example where $q_3$ does not fall in this interval, however the midpoint, $q_2$, falls in between the quantities $\frac{b_1 a_2}{b_2}$ and $\frac{a_1 b_2}{a_2}$ and is therefore the required minimizer. The green
curve illustrates an example where neither of the above happens, and therefore the breakpoint between the function $V_2(c_1)$ and the function $V_3(c_1)$ is the minimizer. In this example we are in Case 1, and therefore this point is $\frac{a_1 b_2}{a_2}$.

![Fig. 7 Picture to illustrate the possible outcomes of Procedure 1 in Case 1](image)

It is important to note that each of the cases in Procedure 1 actually can occur. It is easy to check the following:

- An example of a red curve (minimum occurs at $q_3$) is $(a_1 = 1, b_1 = 35, a_2 = 2, b_2 = 12, a_3 = 12, b_3 = 35)$.
- An example of a blue curve (minimum occurs at $q_2$) is $(a_1 = 1, b_1 = 34, a_2 = 2, b_2 = 36, a_3 = 12, b_3 = 35)$.
- An example of a green curve (minimum occurs at $\frac{a_1 b_2}{a_2}$) is $(a_1 = 1, b_1 = 8, a_2 = 5, b_2 = 22, a_3 = 1, b_3 = 4)$.

Unfortunately, the plots of the actual functions do not display the key details as clearly as our illustration, so we do not include them here.

Furthermore, an example of Case 2, where the minimum occurs at the breakpoint between the function $V_4$ and the function $V_3$, i.e. the point $\frac{a_1 b_2}{a_2}$ is $(a_1 = 1, b_1 = 13, a_2 = 1, b_2 = 2, a_3 = 2, b_3 = 4)$. Finally, a simple example of Case 0, is the special case $(a_1 = 0, b_1 = 1, a_2 = 0, b_2 = 1, a_3 = 0, b_3 = 1)$. In Figure 8 we can see the plot of this function and the minimum, which falls at the midpoint. In Case 0 we always have $q_1 = q_2 = q_3 = \frac{a_1 + b_1}{2} = \frac{b_2}{2}$.

2.5 Global convexity of our piecewise-quadratic function over its domain

We have seen that each piece of $TV(c_1)$ and $\hat{TV}(c_1)$ is a convex quadratic function. However, this does not imply that the functions are convex over the whole domain,
Fig. 8 Plot of the total volume function (when branching on $x_1$), for parameter values: $(a_1 = 0, b_1 = 1, a_2 = 0, b_2 = 1, a_3 = 0, b_3 = 1)$

$[a_1, b_1]$, and in fact, we hinted at the possibility of non-convexity in our sketch of Figure 6. Nevertheless, as we show in the following theorem, with a bit more work, we are able to demonstrate that $TV(c_1)$ and $\hat{TV}(c_1)$ are convex over the domain, $[a_1, b_1]$. Therefore, a more appropriate picture for Figure 6 would be the illustration in Figure 9. It is very useful that these functions are globally convex; if a variable appears in many trilinear terms, it is quite reasonable to combine volumes in a reasonable manner (see [18]). For example, we can take a weighted average (of the sum of the two volumes for each term) as a measure for deciding on a branching point. A weighted average (assuming positive weights) of convex functions is convex, and therefore, the global-convexity property of these functions allows us to find the optimal branching point (defined as the minimum of the weighted-average function) by a simple bisection search.

**Theorem 3** Given that the upper- and lower-bound parameters respect the labeling $\Omega$, the functions $TV(c_1)$ and $\hat{TV}(c_1)$ are globally-convex functions in the branching point $c_1$ over the domain $[a_1, b_1]$.

**Proof** To demonstrate the global convexity of a continuous piecewise-convex quadratic, we must look at each breakpoint separately. If the first derivative of the left quadratic at the breakpoint is less than or equal to the first derivative of the right quadratic at the breakpoint, and moreover, this is true for all breakpoints, then we have global convexity on the domain (i.e., the second derivative remains non-negative).

The functions $TV(c_1)$ and $\hat{TV}(c_1)$ are both continuous on $[a_1, b_1]$. Furthermore, they each have two breakpoints: one at $\frac{b_1 a_2}{b_2}$, and the other at $\frac{a_1 b_2}{a_2}$. Therefore, to demonstrate global convexity over $[a_1, b_1]$ for each function, we have two breakpoints to consider in each case.
Global convexity of $TV(c_1)$: First, we compare the first derivatives of $V_1$ and $V_2$ at the breakpoint $\frac{b_1a_2}{b_2}$:

$$\frac{dV_2}{dc_1} \left( \frac{b_1a_2}{b_2} \right) - \frac{dV_1}{dc_1} \left( \frac{b_1a_2}{b_2} \right) = \frac{1}{12} b_1(b_3 - a_3)^2(b_2 - a_2)^2 \geq 0.$$

Secondly, we compare the first derivatives of $V_2$ and $V_3$ at breakpoint $\frac{a_1b_2}{a_2}$:

$$\frac{dV_3}{dc_1} \left( \frac{a_1b_2}{a_2} \right) - \frac{dV_2}{dc_1} \left( \frac{a_1b_2}{a_2} \right) = \frac{1}{12} a_1(b_3 - a_3)^2(b_2 - a_2)^2 \geq 0.$$

These quantities are both non-negative; therefore, we observe that $TV(c_1)$ is globally convex over the domain $[a_1, b_1]$.

Global convexity of $\hat{TV}(c_1)$: First, we compare the first derivatives of $V_1$ and $V_4$ at the breakpoint $\frac{a_1b_2}{a_2}$:

$$\frac{dV_4}{dc_1} \left( \frac{a_1b_2}{a_2} \right) - \frac{dV_1}{dc_1} \left( \frac{a_1b_2}{a_2} \right) = \frac{1}{12} a_1(b_3 - a_3)^2(b_2 - a_2)^2 \geq 0.$$

Secondly, we compare the first derivatives of $V_4$ and $V_3$ at the breakpoint $\frac{b_1a_2}{b_2}$:

$$\frac{dV_3}{dc_1} \left( \frac{b_1a_2}{b_2} \right) - \frac{dV_4}{dc_1} \left( \frac{b_1a_2}{b_2} \right) = \frac{1}{12} b_1(b_3 - a_3)^2(b_2 - a_2)^2 \geq 0.$$

These quantities are both non-negative; therefore, we observe that $\hat{TV}(c_1)$ is also globally convex over the domain $[a_1, b_1]$. □
2.6 Bounds on where the optimal branching point can occur

We have seen in \[ \square \] that software employ methods to avoid selecting a branching point that falls too close to either endpoint of the interval. Therefore, a natural issue to consider is whether our minimizer can fall close to either of the endpoints. We want to know how likely it is that solvers are routinely precluding our optimal branching point. The following theorems give some insight on this issue and show that, in fact, software is unlikely to be cutting off our optimal branching point.

**Theorem 4** The branching point for variable \( x_1 \) that obtains the least total volume, never occurs at a point in the interval greater than the midpoint.

*Proof* If \( a_2 = 0 \), then we are in Case 0, and the minimizer is at the midpoint, which is clearly no greater than the midpoint.

If \( \frac{a_2}{a_1} b_2 \geq \frac{b_2}{a_1} a_2 \), then we are in Case 1. If \( q_3 \geq \frac{a_2}{a_1} b_2 \), then \( q_3 \) is the minimizer, but we know that \( q_3 \leq \frac{a_2 + b_1}{2} \) (see [4]). If \( q_2 = \frac{a_1 + b_1}{2} \) falls in the interval \( \left[ \frac{b_1 a_2}{b_2}, \frac{a_1 b_2}{a_2} \right] \), then the midpoint is the minimizer. If it does not, then (i) \( \frac{a_2}{a_1} b_2 \) is the minimizer, and (ii) it must be that either that \( \frac{a_2}{a_1} b_2 > \frac{b_1}{b_2} b_2 \), in which case our claim is valid, or \( \frac{a_2}{a_1} b_2 \leq \frac{b_2}{b_2} a_2 \). We will show by contradiction that this cannot be the case.

Toward this end, assume that:

\[
\frac{a_1 + b_1}{2} < \frac{b_1 a_2}{b_2} \quad \text{and} \quad \frac{a_1 + b_1}{2} < \frac{a_1 b_2}{a_2}.
\]

This implies:

\[
2b_1 a_2 - b_1 b_2 - a_1 b_2 = b_1 (a_2 - b_2) + (b_1 a_2 - a_1 b_2) > 0, \quad \text{and} \quad 2a_1 b_2 - a_1 a_2 - b_1 a_2 = a_1 (b_2 - a_2) + (a_1 b_2 - b_1 a_2) > 0.
\]

Now let \( X := b_2 - a_2 \) and \( Y := b_1 a_2 - a_1 b_2 \) (note that both \( X \) and \( Y \) are non-negative: Lemma [5]). Therefore we can write our assumption as:

\[
b_1 (-X) + Y > 0 \quad \text{and} \quad a_1 (X) + (-Y) > 0,
\]

which implies

\[
Y > b_1 X \quad \text{and} \quad Y < a_1 X,
\]

a contradiction. Therefore, in Case 1 the minimizer must be no larger than the midpoint.

We make a similar argument for Case 2. Here \( \frac{a_1 b_2}{a_2} < \frac{b_1 a_2}{b_2} \). If \( q_3 \geq \frac{b_1 a_2}{b_2} \), then \( q_3 \) is the minimizer, but we know that \( q_3 \leq \frac{a_2 + b_1}{2} \) (see [4]). If \( q_2 = \frac{a_1 + b_1}{2} \) falls in the interval \( \left[ \frac{a_1 b_2}{a_2}, \frac{b_1 a_2}{b_2} \right] \), then the midpoint is the minimizer. If it does not, then (i) \( \frac{b_1 a_2}{b_2} \) is the minimizer, and (ii) it must be that either that \( \frac{a_2}{a_1} b_2 > \frac{b_1}{b_2} b_2 \), in which case our claim is valid, or \( \frac{a_2}{a_1} b_2 < \frac{b_2}{b_2} a_2 \). However, we have just shown by contradiction that this cannot be the case. Therefore, in Case 2 the minimizer must be no larger than the midpoint. \( \square \)
This theorem gives an upper bound on the fraction through the interval the minimizer can fall (namely $\frac{1}{2}$). Furthermore, this bound is sharp, given that we know examples when the minimizer is exactly at the midpoint. It would be nice to also obtain a sharp lower bound on this fraction. By demonstrating that the minimizer cannot fall too close to the end points of the interval, we are providing mathematical evidence to justify the current choices of branching point in software, as discussed in §1. The following theorem gives a lower bound on this fraction when $a_2 \neq 0$, (when $a_2 = 0$, we know that the minimizer will be exactly at the midpoint).

**Theorem 5** Given upper- and lower-bound parameters $(a_1, b_1, a_2, b_2, a_3, b_3)$ satisfying (2) and $a_2 \neq 0$. The branching point for variable $x_1$ that obtains the least total volume, never occurs at a point in the interval less than

$$\min \left\{ \max \left\{ \frac{a_1(b_2 - a_2)}{a_2(b_1 - a_1)}, \frac{b_1a_2 - a_1b_2}{b_1b_2 - a_1b_2} \right\}, \frac{1}{2} \right\}.$$

of the way through the interval.

**Proof** There are four candidate points where the minimizer can occur. Namely, $q_2 = \frac{a_1 + b_1}{2}$, $q_3$, $\frac{a_1b_2}{a_2}$, and $\frac{b_1a_2}{b_2}$. Therefore

$$\min \left\{ a_1 + \frac{b_1}{2}, q_3, \frac{a_1b_2}{a_2}, \frac{b_1a_2}{b_2} \right\},$$

is a trivial lower bound on this minimizer.

We know that if $q_3$ is the minimizer, then we must have $q_3 \geq \frac{a_1b_2}{a_2}$ (Case 1), or $q_3 \geq \frac{b_1a_2}{b_2}$ (Case 2), so we can discard this point.

Additionally, we know that if $\frac{a_1b_2}{a_2}$ is the minimizer, then we have $\frac{a_1b_2}{a_2} \geq \frac{b_1a_2}{b_2}$ (Case 1), and if $\frac{b_1a_2}{b_2}$ is the minimizer, then we have $\frac{b_1a_2}{b_2} > \frac{a_1b_2}{a_2}$ (Case 2).

Therefore we have that a lower bound on the minimizer is:

$$\min \left\{ \max \left\{ \frac{a_1b_2}{a_2}, \frac{b_1a_2}{b_2} \right\}, \frac{a_1 + b_1}{2} \right\}.$$

Moreover, a lower bound for the fraction of the interval where this point can fall is:

$$\min \left\{ \max \left\{ \frac{a_1b_2}{a_2} - a_1, \frac{b_1a_2}{b_2} - a_1 \right\}, \frac{a_1 + b_1}{2} - a_1 \right\} = \min \left\{ \max \left\{ \frac{a_1(b_2 - a_2)}{a_2(b_1 - a_1)}, \frac{b_1a_2 - a_1b_2}{b_1b_2 - a_1b_2} \right\}, \frac{1}{2} \right\}.$$

We note that this lower bound is unlikely to be sharp. Consider the case where $a_1 = 0$, $a_2 = \epsilon > 0$ and $b_2 = 1$. This bound becomes $\epsilon$, and is therefore not particularly informative, given that we can make $\epsilon$ as close to zero as we wish. However, we have computationally checked many examples, and we have yet to find an example where the minimizer occurs less than $\sim 0.45$ of the way through the interval. It would be nice to sharpen this bound, and our computations indicate that this should be possible.
3 Branching on $x_2$ and $x_3$

We noted in §2 that because of the structure of the volume function of the convex hull, the second and third variables are interchangeable. Therefore, the branching-point analyses for these variables will be equivalent. To see how the results in this case are less complex than in the $x_1$ case, recall the condition $\Omega$, which due to our non-negativity assumption can be written as

$$\frac{a_1}{b_1} \leq \frac{a_2}{b_2} \leq \frac{a_3}{b_3}.$$ 

Now consider what happens to the quantity $\frac{a_2}{b_2}$ when we branch on $x_2$. In the left interval, $a_2$ remains constant, and $b_2$ becomes the branching point, $c_2 < b_2$. Therefore, $\frac{a_2}{b_2}$ cannot decrease further. In the right interval $b_2$ remains constant and $a_2$ becomes the branching point, $c_2 > a_2$. Therefore, again, $\frac{a_2}{b_2}$ cannot decrease further. Because of this, the labeling for $x_1$ and $x_2$ will not have to be switched to ensure $\Omega$ remains satisfied. Furthermore, $x_2$ and $x_3$ are interchangeable in the formula, so we do not need to consider what happens when the ratios change such that $\frac{a_2}{b_2} > \frac{a_3}{b_3}$.

The case of $x_2$ and $x_3$ therefore both require the analysis of only one convex quadratic function. This is formalized in the following theorem.

**Theorem 6** Let $c_i \in [a_i, b_i]$ be the branching point for $x_i$, $i = 2, 3$. With the convex-hull relaxation, the least total volume after branching is obtained when $c_i = (a_i + b_i)/2$, i.e., branching at the midpoint is optimal.

**Proof** We first consider branching on $x_2$. Consider the sum of the two resulting volumes, given by the following function:

$$TV_2(c_2) = V(a_1, b_1, c_2, b_2, a_3, b_3) + V(a_1, b_1, a_2, c_2, a_3, b_3),$$

which is quadratic in $c_2$. The leading coefficient (i.e. second derivative) is

$$TV_2(c_2) = \frac{1}{12}(b_1 - a_1)(b_3 - a_3)(3(b_1 b_3 - a_1 a_3) + (b_1 a_3 - a_1 b_3)),$$

which is greater than or equal to zero for all parameters satisfying $\Omega$ and hence all $c_2 \in [a_2, b_2]$ (Lemma 3). Therefore this function is convex. Setting the first derivative equal to zero and solving for $c_2$, we obtain that the minimum occurs at $c_2 = (a_2 + b_2)/2$. Similar analysis can be completed for $i = 3$ to obtain the result.

4 The optimal branching variable

Now that we have established the optimal branching point for each variable in all cases, it is interesting to compare the total volumes obtained when branching at the optimal point for each variable. In this section we establish the optimal branching variable.
Theorem 7 Given that the upper- and lower-bound parameters respect the labeling \( \Omega \), if we assume optimal branching-point selection, then branching on \( x_1 \) obtains the least total volume, and branching on \( x_3 \) obtains the greatest total volume. Additionally, even if we branch at the midpoint for \( x_1 \) (which may not be optimal), this is at least as good as doing optimal branching-point selection (i.e., midpoint branching) on either \( x_2 \) or \( x_3 \).

Proof First, we establish that branching optimally (at the midpoint) on variable \( x_2 \) obtains a lower total volume than branching optimally (at the midpoint) on variable \( x_3 \).

The optimal total volume when branching on variable \( x_3 \) is:

\[
\frac{(b_3 - a_3)(b_2 - a_2)(b_1 - a_1)}{48} \times (7a_1a_2a_3 + a_1a_2b_3 - 3a_1a_3b_2 - 5a_1b_2b_3 - 5a_2a_3b_1 - 3a_2b_1b_3 + a_3b_1b_2 + 7b_1b_2b_3).
\]

The optimal total volume when branching on variable \( x_2 \) is:

\[
\frac{(b_3 - a_3)(b_2 - a_2)(b_1 - a_1)}{48} \times (7a_1a_2a_3 - 3a_1a_2b_3 + a_1a_3b_2 - 5a_1b_2b_3 - 5a_2a_3b_1 + a_2b_1b_3 - 3a_3b_1b_2 + 7b_1b_2b_3).
\]

Therefore, the difference in total volume from branching on \( x_3 \) compared with \( x_2 \) is:

\[
\frac{(b_3 - a_3)(b_2 - a_2)(b_1 - a_1)^2(b_2a_3 - a_2b_3)}{12},
\]

which is greater than or equal to zero by Lemma 5. Therefore, if we assume optimal branching, branching on \( x_3 \) always results in a greater volume than branching on \( x_2 \).

Now let us consider the optimal total volume when branching on \( x_1 \), this quantity must always be less than or equal to the total volume when branching at the midpoint of the interval (it will be equal exactly when the midpoint is the optimal branching point). Therefore, if we can establish that branching on variable \( x_1 \) at the midpoint always obtains a lesser total volume than branching on variable \( x_2 \) at the midpoint, we will have shown our claim.

Recall Figures 4 and 5. We know from the proof of Theorem 3 that the midpoint can never be less than: \( \min\left(\frac{a_1b_2}{a_1}, \frac{b_1a_2}{a_2}\right) \). Therefore, in every case, the midpoint must fall in a subdomain where: (i) the labeling for left interval stays the same, and the labeling for the right changes; (ii) the labeling changes for both intervals; or, (iii) the labeling remains the same for both intervals. This means that we are interested in the function value (total volume) at the midpoint for the functions \( V_2(c_1), V_3(c_1) \) and \( V_4(c_1) \).

The total volume of branching (on variable \( x_1 \)) at the midpoint if it occurs in the subdomain corresponding to \( V_2 \) is:

\[
\frac{(b_3 - a_3)(b_2 - a_2)(b_1 - a_1)}{48} \times (7a_1a_2a_3 - 3a_1a_2b_3 - 5a_1a_3b_2 + a_1b_2b_3 + a_2a_3b_1 - 5a_2b_1b_3 - 3a_3b_1b_2 + 7b_1b_2b_3).
\]
Therefore, the difference in total volume from branching on $x_2$ compared with this quantity is:

$$\frac{(b_3 - a_3)^2(b_2 - a_2)(b_1 - a_1)(b_1a_2 - a_1b_2)}{8},$$

which is greater than or equal to zero by Lemma 5.

The total volume of branching (on variable $x_1$) at the midpoint if it occurs in the subdomain corresponding to $V_3$ is:

$$\frac{(b_3 - a_3)(b_2 - a_2)(b_1 - a_1)}{48} \times (6a_1a_2a_3 - 2a_1a_2b_3 - 3a_1a_3b_2 - a_1b_2b_3 - 4a_2b_1b_3 - 3a_3b_1b_2 + 7b_1b_2b_3).$$

Therefore, the difference in total volume from branching on $x_2$ compared with this quantity is:

$$\frac{(b_3 - a_3)^2(b_2 - a_2)(b_1 - a_1)(4(b_1a_2 - a_1b_2) + a_2(b_1 - a_1))}{48},$$

which is greater than or equal to zero by Lemma 5.

The total volume of branching (on variable $x_1$) at the midpoint if it occurs in the subdomain corresponding to $V_4$ is:

$$\frac{(b_3 - a_3)(b_2 - a_2)(b_1 - a_1)}{24} \times (3a_1a_2a_3 - a_1a_2b_3 - a_1a_3b_2 - a_1b_2b_3 - a_2a_3b_1 - a_2b_1b_3 - a_3b_1b_2 + 3b_1b_2b_3).$$

Therefore, the difference in total volume from branching on $x_2$ compared with this quantity is:

$$\frac{(b_3 - a_3)^2(b_2 - a_2)(b_1 - a_1)(b_1b_2 - a_1a_2 + 3(b_1a_2 - a_1b_2))}{48},$$

which is greater than or equal to zero by Lemma 5.

Therefore, for each one of these possible scenarios, optimally branching on $x_2$ results in a greater volume than branching on $x_1$ at the midpoint. And so we can conclude that given optimal branching, branching on $x_1$ obtains the least total volume, and branching on $x_3$ obtains the greatest total volume.

\[\square\]

5 Concluding remarks and future work

We have presented some analytic results on branching variable and branching-point selection in the context of sBB applied to models having functions involving the multiplication of three or more terms. In particular, for trilinear monomials $f = x_1x_2x_3$ on a box domain satisfying (2) we have shown that when the convex-hull relaxation is used, and the branching variable is $x_2$ or $x_3$, branching at the commonly-used midpoint results in the least total volume.

We have presented a simple procedure for obtaining the optimal branching point when using the convex-hull relaxation and branching on variable $x_1$. We have provided a sharp upper bound on where in the interval the minimizer can
occur, and we have also obtained a lower bound for this fraction. We have computational evidence to suggest that this lower bound can be sharpened, thus providing analysis that backs up software’s current choice of branching point. Furthermore, we have shown that the piecewise-quadratic functions we have been considering are globally convex over their entire domain.

Given that we branch at an optimal branching point, we have also compared the choice of branching variable. We demonstrate that branching on \( x_1 \) gives the least total volume.

We are in the process of carrying out a similar analysis to what we have done here, but for the best of the double-McCormick convexifications rather than for the convex-hull relaxation. However, due to the structure of the volume formula for the best double-McCormick convexification (see [17]), our task is significantly more complex.

Finally, we hope that our mathematical results can be used as some guidance toward justifying, developing and refining practical branching rules. We believe that our work is just a first step in this direction. In this regard, we hope to further extend our mathematical analysis to directly deal with variables appearing in multiple non-convex terms.

6 Appendix: technical propositions and lemmas

In this section, we provide the technical propositions and lemmas used for our analysis.

**Proposition 1** Given that the upper- and lower-bound parameters respect the labeling \( \Omega \) and \( \frac{b_1 a_2}{b_2} \leq \frac{a_1 b_2}{a_2} \),

\[
V_1(q_1) \leq V_2\left( \frac{a_1 b_2}{a_2} \right) = V_3\left( \frac{a_1 b_2}{a_2} \right).
\]

**Proof** It is easy to check that \( V_2\left( \frac{a_1 b_2}{a_2} \right) = V_3\left( \frac{a_1 b_2}{a_2} \right) \).

\[
V_2\left( \frac{a_1 b_2}{a_2} \right) - V_1(q_1) = \frac{(b_3 - a_3)(b_2 - a_2)}{48(4b_2 b_3 - a_2 b_3 - 3a_2 a_3)a_2^2} \times \left( p a_1^2 + qa_1 + r \right),
\]

where

\[
p = \left( -3a_2 a_3 - a_2 b_3 + b_2 a_3 + 3b_2 b_3 \right) \times \\
\left( -3a_2^2 a_3 - a_2^2 b_3 + 13a_2^2 b_2 a_3 + 7a_2^2 b_2 b_3 - 12a_2 b_2^2 a_3 - 20a_2 b_2^2 b_3 + 16b_2^3 b_3 \right) \\
= \left( 3(b_2 b_3 - a_2 a_3) + b_2 a_3 - a_2 b_3 \right) \times \\
\left( -3a_2^3 + 13a_2^2 b_2 - 12a_2 b_2^2 a_3 + (-a_2^3 + 7a_2^2 b_2 - 20a_2 b_2^2 + 16b_2^2 b_3) b_3 \right).
\]

\[
q = 4a_2 b_1(2a_2^2 a_3 - 3a_2 b_2 a_3 - 3a_2 b_2 b_3 + 4b_2^2 b_3) \times (3a_2 a_3 + a_2 b_3 - b_2 a_3 - 3b_2 b_3),
\]

\[
r = 4a_2^2 b_1(2a_2 a_3 + a_2 b_3 - 2b_2 b_3)^2.
\]
To show that $V_2 \left( \frac{a_1 b_2}{a_2^2} \right) - V_1 (q_1)$ is non-negative for all parameters satisfying $\Omega$, we will show that $p a_1 + qa_1 + r \geq 0$ for all parameters satisfying $\Omega$.

We observe:

$$\left(-a_2^3 + 7a_2^2b_2 - 20a_2b_2^2 + 16b_2^3 \right) b_3 + \left(-3a_2^3 + 13a_2^2b_2 - 12a_2b_2^2a_3 \right) =: b_3 Y + a_3 Z,$$

where

$$Y + Z = 4(b_2 - a_2)(2b_2 - a_2)^2 \geq 0,$$

and

$$Y = \left(b_2 - a_2 \right) \left( 4b_2(b_2 - a_2) + 12b_2^2 + a_2^2 \right) + 2a_2^2b_2 \geq 0.$$

Therefore, by Lemma 6 we have that $b_3 Y + a_3 Z$ is non-negative and so $p$ is non-negative (Lemma 5). From this we know that $p a_1 + qa_1 + r$ is a convex function in $a_1$ and we can find the minimizer by setting the derivative to zero and solving for $a_1$. The minimum occurs at

$$a_1 = \frac{2b_1a_2(2a_2^2a_3 - 3a_2b_2a_3 - 3a_2b_2b_3 + 4b_2^3b_3)}{(-3a_2^3a_3 - a_2^3b_3 + 13a_2^2b_2a_3 + 7a_2^2b_2b_3 - 12a_2b_2^2a_3 - 20a_2b_2^2b_3 + 16b_2^2b_3)}.$$

Substituting this in to $p a_1 + qa_1 + r$, we obtain that the minimum value of this quadratic is:

$$\frac{4a_2^2b_2^2(b_3 - a_3)(b_2 - a_2)^3(3a_2a_3 + a_2b_2 - 4b_2^3)^2}{(-3a_2^3a_3 - a_2^3b_3 + 13a_2^2b_2a_3 + 7a_2^2b_2b_3 - 12a_2b_2^2a_3 - 20a_2b_2^2b_3 + 16b_2^2b_3)}.$$

In demonstrating the non-negativity of $p$, we have already shown that the denominator is non-negative, and it is easy to see that the numerator is non-negative for all values of the parameters satisfying $\Omega$. Therefore $p a_1 + qa_1 + r \geq 0$, and consequently, $V_2 \left( \frac{a_1 b_2}{a_2^2} \right) - V_1 (q_1) \geq 0$ as required.

□

**Lemma 1** Given that the upper- and lower-bound parameters respect the labeling $\Omega$ and $\frac{b_1 a_2}{b_2} \leq \frac{a_1 b_2}{a_2}$,

$$V_1 \left( \frac{b_1 a_2}{b_2} \right) = V_2 \left( \frac{b_1 a_2}{b_2} \right) \geq V_2 \left( \frac{a_1 b_2}{a_2} \right) = V_3 \left( \frac{a_1 b_2}{a_2} \right).$$

**Proof** It is easy to check that $V_1 \left( \frac{b_1 a_2}{b_2} \right) = V_2 \left( \frac{b_1 a_2}{b_2} \right)$ and $V_2 \left( \frac{a_1 b_2}{a_2} \right) = V_3 \left( \frac{a_1 b_2}{a_2} \right)$.

Furthermore,

$$V_2 \left( \frac{b_1 a_2}{b_2} \right) - V_2 \left( \frac{a_1 b_2}{a_2} \right) = \frac{(b_3 - a_3)(b_2 - a_2)^2(b_3 - a_2b_2)(a_1 b_2^2 - a_2b_2)(3b_2b_3 - a_2a_3) + b_2a_3 - a_2b_3}{12a_2^2b_2^2} \geq 0,$$

as required. □
Lemma 2 Given that the upper- and lower-bound parameters respect the labeling $\Omega$ and $\frac{b_1a_2}{a_2} > \frac{a_1b_2}{b_2}$,

$$V_1(q_1) \leq V_4\left(\frac{b_1a_2}{b_2}\right) = V_3\left(\frac{b_1a_2}{b_2}\right).$$

Proof It is easy to check that $V_4\left(\frac{b_1a_2}{b_2}\right) = V_3\left(\frac{b_1a_2}{b_2}\right)$.

$$V_4\left(\frac{b_1a_2}{b_2}\right) - V_1(q_1) = \frac{(b_3 - a_3)(b_2 - a_2)}{48(4b_2b_3 - a_2b_3 - 3a_2a_3)b_2^2} \times \left(pa_1^2 + qa_1 + r\right),$$

where

$$p = b_2^2(5b_2b_3 - b_2a_3 - a_2b_3 - 3a_2a_3)^2,$$

$$q = 8b_1b_2(6a_2^2a_3 + 2a_2^2b_3 - 3a_2b_2a_3 - 9a_2b_2b_3 + b_2^2a_3 + 3b_2^2b_3)(b_2b_3 - a_2a_3),$$

$$r = 16b_2^2(-3a_2^2a_3 - a_2^2b_3 + 3a_2b_2a_3 + 5a_2^2b_2b_3 - a_2b_2a_3 - 4a_2b_2b_3 + b_2^2b_3)$$

$$\times (b_2b_3 - a_2a_3).$$

To show this is non-negative for all parameters satisfying $\Omega$ we will show $pa_1^2 + qa_1 + r \geq 0$ for all parameters satisfying $\Omega$.

Firstly, we observe that

$$p = b_2^2(5b_2b_3 - b_2a_3 - a_2b_3 - 3a_2a_3)^2 \geq 0.$$  

From this we know that $pa_1^2 + qa_1 + r$ is a convex function in $a_1$, and we can find the minimizer by setting the derivative to zero and solving for $a_1$. The minimum occurs at

$$a_1 = \frac{4b_1(6a_2^2a_3 + 2a_2^2b_3 - 3a_2b_2a_3 - 9a_2b_2b_3 + b_2^2a_3 + 3b_2^2b_3)(a_2a_3 - b_2b_3)}{b_2(3a_2a_3 + a_2b_3 + b_2a_3 - 5b_2b_3)^2}.$$  

Substituting this in to $pa_1^2 + qa_1 + r$, we obtain that the minimum value of this quadratic is:

$$\frac{16b_2^2(b_3 - a_3)(b_2 - a_2)^3(b_2b_3 - a_2a_3)(3a_2a_3 + a_2b_3 - 4b_2b_3)^2}{(3a_2a_3 + a_2b_3 + b_2a_3 - 5b_2b_3)^2},$$

which is non-negative for all parameters satisfying $\Omega$. Therefore $pa_1^2 + qa_1 + r \geq 0$, and consequently, $V_4\left(\frac{b_1a_2}{b_2}\right) - V_1(q_1) \geq 0$, as required.

$\square$

Lemma 2 Given that the upper- and lower-bound parameters respect the labeling $\Omega$ and $\frac{b_1a_2}{b_2} > \frac{a_1b_2}{a_2}$,

$$V_1\left(\frac{a_1b_2}{a_2}\right) = V_4\left(\frac{a_1b_2}{a_2}\right) \geq V_4\left(\frac{b_1a_2}{b_2}\right) = V_3\left(\frac{b_1a_2}{b_2}\right).$$
Proof It is easy to check that $V_1\left(\frac{a_1 b_2}{a_2}\right) = V_4\left(\frac{a_1 b_2}{a_2}\right)$ and $V_4\left(\frac{b_1 a_2}{b_2}\right) = V_3\left(\frac{b_1 a_2}{b_2}\right)$.

Furthermore,

$$V_4\left(\frac{a_1 b_2}{a_2}\right) - V_4\left(\frac{b_1 a_2}{b_2}\right) = \frac{(b_3 - a_3)(b_2 - a_2)^2(b_1 a_2^2 - a_1 b_2^2)(b_1 a_2 - a_1 b_2)(b_2 b_3 - a_2 a_3)}{3a_1^2 b_2^2} \geq 0,$$

as required.

\[\square\]

**Lemma 3** Given that the parameters satisfy the conditions $\Omega$ and furthermore, $\frac{b_1 a_3}{b_2} \leq \frac{a_1 b_3}{a_2}$, we have

$$q_1 \geq \frac{b_1 a_2}{b_2}.$$

Proof From the proof of Theorem 4, we know that the midpoint, $q_2$, cannot be less than both $\frac{a_1 b_2}{b_1}$ and $\frac{a_1 b_3}{b_2}$. Therefore we have:

$$q_2 \geq \min\left\{\frac{a_1 b_2}{b_1}, \frac{b_1 a_3}{b_2}\right\},$$

and because we saw in 5 that $q_1 \geq q_2$ we also have

$$q_1 \geq \min\left\{\frac{a_1 b_2}{b_1}, \frac{b_1 a_2}{b_2}\right\}.$$

Therefore, under the conditions of the lemma, $q_1 \geq \frac{b_1 a_2}{b_2}$ as required.

\[\square\]

**Lemma 4** Given that the parameters satisfy the conditions $\Omega$ and furthermore, $\frac{b_1 a_2}{b_2} \geq \frac{a_1 b_2}{a_2}$, we have

$$q_1 \geq \frac{a_1 b_2}{a_2}.$$

Proof We saw in the proof of Lemma 3 that

$$q_1 \geq \min\left\{\frac{a_1 b_2}{b_1}, \frac{b_1 a_2}{b_2}\right\}.$$

Therefore, under the conditions of the lemma, $q_1 \geq \frac{a_1 b_2}{a_2}$ as required.

\[\square\]

For completeness, we state and give proofs of two very simple lemmas (from [17]) which we used several times.

**Lemma 5 (Lemma 10.1 in [17])** For all choices of parameters $0 \leq a_i < b_i$ satisfying $\Omega$, we have: $b_1 a_2 - a_1 b_2 \geq 0$, $b_1 a_3 - a_1 b_3 \geq 0$ and $b_2 a_3 - a_2 b_3 \geq 0$.

Proof $(b_3 - a_3)(b_1 a_2 - a_1 b_2) = b_1 a_2 b_3 + a_1 b_2 a_3 - a_1 b_2 b_3 - b_1 a_2 a_3 \geq 0$ by [2]. This implies $b_1 a_2 - a_1 b_2 \geq 0$, because $b_3 - a_3 > 0$. $b_1 a_3 - a_1 b_3 \geq 0$ and $b_2 a_3 - a_2 b_3 \geq 0$ follow from [17] in a similar way.
Lemma 6 (Lemma 10.4 in [17]) Let $A, B, C, D \in \mathbb{R}$ with $A \geq B \geq 0$, $C + D \geq 0$, $C \geq 0$. Then $AC + BD \geq 0$.

Proof $AC + BD \geq B(C + D) \geq 0$. $\square$

Acknowledgements This work was supported in part by ONR grants N00014-14-1-0315 and N00014-17-1-2296. The authors gratefully acknowledge conversations with Ruth Misener and Nick Sahinidis concerning how branching points are selected in ANTIGONE and BARON.

References