Volumetric Guidance for Handling Triple Products in Spatial Branch-and-Bound

by

Emily E. Speakman

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Industrial and Operations Engineering)
in The University of Michigan
2017

Doctoral Committee:

Professor Jon Lee, Chair
Associate Professor Kevin J. Compton
Associate Professor Marina A. Epelman
Assistant Professor Siqian M. Shen

Emily E. Speakman

ORCID iD: 0000-0002-7352-1355

ACKNOWLEDGEMENTS

First, I would like to thank my advisor, Professor Jon Lee. His guidance and direction have been invaluable throughout this journey, and I truly can't imagine having undertaken this process without his mentorship. Any amount of success I may have, either now or in the future, also belongs to him.

Secondly, thank you to Mr. Harris, who made me want to pursue mathematics, and to all my teachers, advisors, and professors after that, who prepared me for my studies at Michigan. In particular, I owe much gratitude to Jill Hardin Wilson. She believed I could do this well before it even occurred to me to believe in myself. I wouldn't be here without her.

Thank you to my committee members, Professor Marina Epelman, Professor Siqian Shen, and Professor Kevin Compton, not only for participating in my defense, but for your courses and input throughout my time here. Between the three of you, I have learned an enormous amount. Thank you as well to all the faculty, staff, and students in the IOE department, for being there throughout the past five years. This experience would not have been all that it was without you.

I would certainly never have made it to this point without my friends and family. I will avoid using names in the fear that I may miss someone out, but to those of you far away: thank you. You have supported me, cheered for me and visited me despite the distance. It is impossible for me to overstate my appreciation for you. To my friends in Ann Arbor: you have kept me encouraged even during the low points. Thank you for making this place a second home.

To mum and dad, with more love and gratitude than I can ever express, no one could ask for more supportive parents than the two of you.

Finally, I give thanks to God, who holds all things together.

In addition, I gratefully acknowledge partial support from NSF grant CMMI1160915, ONR grant N00014-14-1-0315, a Rackham Summer Award, and the Dwight F. Benton Fellowship.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS ii

LIST OF FIGURES v

LIST OF TABLES vii

ABSTRACT viii

CHAPTER

1. Introduction 1

1.1 Thesis overview 7

2. Preliminaries 8

2.1 Double McCormick 8
2.1.1 Convexification 9
2.1.2 Hull 10
2.2 Alternatives 12

3. Volume Formulae 14

3.1 Introduction 14
3.2 Theorems 14
3.3 Proof of Thm. 3.1 17
3.3.1 An interesting note about $\mathcal{P}_h$ 24
3.4 Proof of Thm. 3.4 25
3.4.1 Keeping track of facets 32
3.5 Proof of Thm. 3.2 37
3.5.1 Keeping track of facets 46
3.6 Proof of Thm. 3.3 54
3.7 Concluding remarks and future work 54
3.8 Technical lemmas 55
3.8.1 Useful lemmas 56

3.8.2 Proving non-negativity 56

4. Experimental Justification of Volume 64

4.1 Introduction 64
4.2 Measuring relaxations via volume in mathematical optimization 64
4.3 Box cubic programming problems 65
4.4 From volume to objective function gap 66
4.5 Computational experiments 67
4.5.1 Box cubic programming problems and four relaxations 67
4.5.2 Three scenarios for the hypergraph H 68
4.5.3 Quality of relaxations 69
4.5.4 Validating the relationship between volume and objective gap 70
4.5.5 A worst case 71
4.6 Concluding remarks and future work 81

5. Using Volume to Guide Branching-Point Selection 82

5.1 Introduction 82
5.2 Current branching practice 82
5.3 Results 84
5.3.1 The convex-hull convexification 85
5.3.2 The best double-McCormick convexification 99
5.4 Concluding remarks and future work 102
5.5 Technical propositions, lemmas, and theorems 103
5.5.1 Convex-hull convexification 103
5.5.2 Double-McCormick convexification 107

BIBLIOGRAPHY 114

LIST OF FIGURES

1.1 An example of a DAG for the function $f = \frac{x_1^3 e^{x_2}}{\sin(x_2 x_3)}$ 3

1.2 Convexifying and reconvexifying after branching in sBB 4

3.1 Difference in volume between $\mathcal{P}_3$ and $\mathcal{P}_1$ $\left(\frac{3a_3(b_3-a_3)^2}{24b_3}\right)$ vs. parameters a_3 and b_3 ($a_1 = a_2 = 0$ and $b_1 = b_2 = 1$) 18

3.2 Visual representation of the convex-hull extreme points 19

3.3 Visual representation of adding point v^8 to simplex $\mathcal{S}$ 20

3.4 Visual representation of adding points v^3 and v^7 to polytope $\mathcal{Q}$ . . . 22

3.5 The line segment between v^3 and v^7 intersects polytope $\mathcal{Q}$ 23

3.6 Visual representation of the convex-hull polytope 25

3.7 Visual representation of the convex-hull polytope (blue) and the four 'extra' extreme points of $\mathcal{P}_3$ 26

3.8 For Table 3.1 27

3.9 For Table 3.2 38

4.1 Quasi-mean-width differences 73

4.2 Quasi-mean-width performance profiles 74

4.3 Idealized radius predicting quasi mean width 76

4.4 Idealized radial distance predicting quasi mean width difference . . 78

4.5 Worst-case analysis for a_3 ($a_1 = a_2 = 0$, $b_1 = b_2 = 1$) 80

5.1 Variable labeling as the branching point varies in Case 0 86

5.2 Variable labeling as the branching point varies in Case 1 and Case 2 86

5.3 Illustration of a continuous piecewise-quadratic function 88

5.4 Picture to illustrate the possible outcomes of Algorithm 1 in Case 1 93

5.5 Plot of the total volume function for parameter values: $(a_1 = 0, b_1 = 1, a_2 = 0, b_2 = 1, a_3 = 0, b_3 = 1)$ 94

5.6 Illustration of a globally convex piecewise-quadratic function 95

5.7 Case analysis for $\mathcal{P}_3$ 100

LIST OF TABLES

3.1 Summary of midpoint substitutions for Thm. 3.4 27

3.2 Summary of midpoint substitutions for Thm. 3.2 38

5.1 Default parameter settings 83

ABSTRACT

Volumetric Guidance for Handling Triple Products in Spatial Branch-and-Bound
by
Emily E. Speakman

Chair: Jon Lee

Spatial branch-and-bound (sBB) is the workhorse algorithmic framework used to globally solve mathematical mixed-integer non-linear optimization (MINLO) problems. Formulating a problem using this paradigm allows both the non-linearities of a system and any discrete design choices to be modeled effectively. Because of the generality of this approach, MINLO is used in a wide variety of applications, from chemical engineering problems and network design, to medical applications and problems in the airline industry.

Due in part to their generality (and therefore wide applicability), MINLO problems are very difficult in general, and consequently, the best ways to implement many details of sBB are not wholly understood. In this work, we provide analytic results guiding the implementation of sBB for a simple but frequently occurring function building block. As opposed to computationally demonstrating that our techniques work only for a particular set of test problems, we analytically establish results that hold for all problems of the given form. In this way, we also demonstrate that analytic results are indeed obtainable for certain sBB implementation decisions.

In particular, we use volume as a geometric measure to compare different convex relaxations for functions involving trilinear monomials (or any three quantities

multiplied together). We consider different choices for convexifying the graph of a triple product (i.e. $f = x_1x_2x_3$), and obtain formulae for the volume (in terms of the variable upper and lower bounds) for each of these convexifications. We are then able to order the convexifications with regard to their volume. We also provide computational evidence to support our choice of volume as an effective comparison measure, and show that in the context of triple products, volume is an excellent predictor of the objective function gap. Finally, we use the volume measure to provide guidance regarding branching-point selection in the implementation of sBB.

CHAPTER 1

Introduction

Mathematical optimization is a commonly used and powerful paradigm for modeling and solving a variety of important applied problems. The inherent non-linearities and discrete decisions in many aspects of the real world mean that numerous important problems can be solved by globally optimizing a mixed-integer non-linear optimization (MINLO) problem. Therefore, it is important that we understand how to simultaneously deal with *both* the discrete *and* the non-linear aspects of these problems to design efficient algorithms.

To quote R. T. Rockafellar ([43]) "...the great watershed in optimization isn't between linearity and non-linearity, but convexity and non-convexity." This is what makes MINLO problems especially hard — they can contain non-convexity in the form of integer variables, and also in the structure of the objective and constraint functions themselves. Seeking to handle broad classes of non-convex functions implies that state-of-the-art software can only hope to routinely succeed on relatively small problem instances, and that research has great potential to achieve significant improvements on current performance. A well-studied example where these non-convexities occur is the pooling problem, an important application arising from chemical engineering (see [38] for a survey), but non-convex functions feature in many other formulations including the network design of gas ([28]), energy ([21]), and transportation ([16]) networks. For a survey of important applications in non-convex MINLO, see §2 of [9]. Many of these applications require us to find good solutions quickly, and to react to new data as we obtain it. By harnessing both growing computational power and theoretical insights to improve solution methods, we can hope for a great impact on the tractability of a host of applied problems.

A MINLO problem has the form:

$$\min_{x\in\mathbb{Z}^n,\, y\in\mathbb{R}^m} \Big\{ f(x,y) \;:\; (x,y)\in\mathcal{F} \Big\},$$

where $f : \mathbb{Z}^n \times \mathbb{R}^m \to \mathbb{R}$ and $\mathcal{F} \subseteq \mathbb{Z}^n \times \mathbb{R}^m$. The only assumptions we make on the function f and the functions that describe $\mathcal{F}$ are that they are *factorable* (see [32]). A function is factorable with respect to a library of low-dimensional functions, e.g., $\sin(x), \ln(x)$, $\arctan(x)$, xy, xyz, x^y, ..., if it can be composed from these functions in a *finite* number of steps by introducing an appropriate number of auxiliary variables. For example:

$$f = \frac{x_1^3 e^{x_2}}{\sin(x_2 x_3)} \;\to f = y_1 y_2 y_5,\; y_5 = \frac{1}{y_4},\; y_4 = \sin(y_3),\; y_3 = x_2 x_3,\; y_2 = e^{x_2},\; y_1 = x_1^3.$$

The assumption that our functions are factorable is quite unrestrictive, and many functions of interest meet this requirement. We note that the determinant of a matrix is a factorable function, but the factorization is unreasonable. 'Finite' could clearly be very large, so in practice we require that our functions can be computed in a 'reasonable' number of steps; how we define reasonable can depend on the problem, but could mean, for example, logarithmic in the number of model variables. Additionally, we assume that we have convexification methods for the graphs of the functions in our library (often linearizations), and that we can combine them to build convexifications of the graphs of the functions in the optimization problem.

Spatial branch-and-bound (sBB) (see [1, 45, 53]) is the workhorse general-purpose algorithm in the area of global optimization. It works by using additional variables to reformulate every function of the formulation as a (labeled) directed acyclic graph (DAG). Root nodes can be very complicated functions, and leaves are variables that appear in the input formulation, each labeled with its interval domain. Intermediate nodes are labeled with auxiliary variables together with operators from a small dictionary of basic functions of few (often one, two, or three) variables. See Figure 1.1 for an example of a DAG for the factorable function we considered above, $f = \frac{x_1^3 e^{x_2}}{\sin(x_2 x_3)}$. As noted, we assume that we have a method for convexifying the graph of each dictionary function. sBB algorithms work by composing convex relaxations of the dictionary functions, according to the DAG, to get relaxations of the root functions. Bounds on the leaves propagate to other nodes and conversely. Branching (subdividing the domain interval of a variable) creates subproblems, which are treated recursively. Objective bounds for subproblems are appropriately combined to

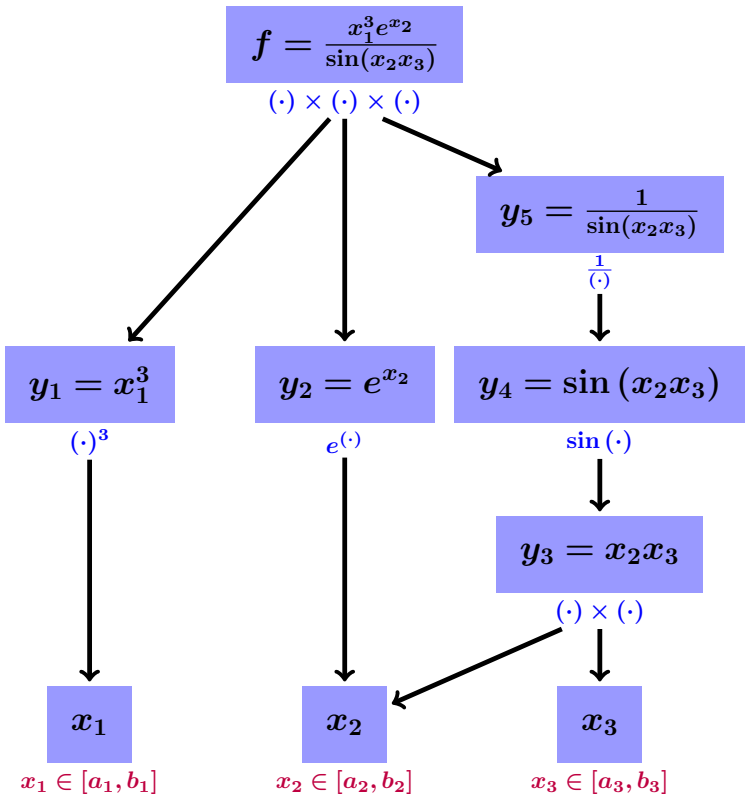


Figure 1.1: An example of a DAG for the function $f = \frac{x_1^3 e^{x_2}}{\sin(x_2 x_3)}$

achieve a global-optimization algorithm.

Figure 1.2 illustrates an example of a univariate function. Here, the sBB algorithm obtains a lower bound for the non-convex blue function by convexifying its graph. The red region is a convexification of the graph of the blue function over the whole domain shown, and the minimum over this set is a lower bound on the value of the blue function over the domain. However, by branching, reconvexifying, and obtaining the two green convex regions, we obtain a tighter lower bound on the blue function (the minimum of the respective minimums over the two green convex regions).

Much of the research on sBB has focused on developing tight convexifications for basic functions of few variables (many references can be found in [11]). Other research has focused on how bounds can be efficiently propagated and how branching can be judiciously be carried out (see [6], for example). From the viewpoint of good convexifications, much less attention has been paid to how the DAGs are created, but this can have a strong impact on the quality of the resulting convex relaxation of the input formulation; see [29, 30, 50, 62] for some key papers with other viewpoints concerning constructing DAGs.

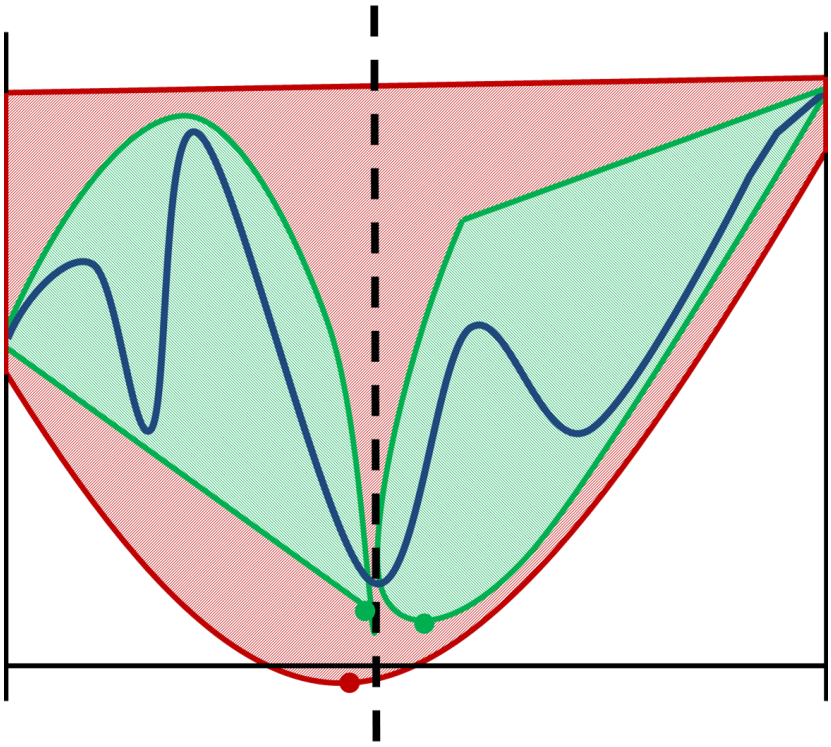


Figure 1.2: Convexifying and reconvexifying after branching in sBB

For basic multilinear monomials $f(x_1, \ldots, x_n) := x_1 \cdots x_n$, with $x_i \in [a_i, b_i]$, there is already a lot of flexibility which can have a significant impact on the overall convexification of the graph of $f(x_1, \ldots, x_n) := x_1 \cdots x_n$ on the box domain $[a_1, b_1] \times \cdots \times [a_n, b_n]$. For $n = 2$, we have the classic McCormick inequalities, see [32], which simply describe the tetrahedron that is the convex hull of the four points

$$(f, x_1, x_2) := (a_1a_2, a_1, a_2), (a_1b_2, a_1, b_2), (b_1a_2, b_1, a_2), (b_1b_2, b_1, b_2),$$

see [3]. The inequalities can be derived from the four inequalities

$$\begin{aligned} (x_1 - a_1)(x_2 - a_2) \geq 0, \quad (x_1 - a_1)(b_2 - x_2) \geq 0, \\ (b_1 - x_1)(x_2 - a_2) \geq 0, \quad (b_1 - x_1)(b_2 - x_2) \geq 0, \end{aligned}$$

by multiplying out and then replacing all occurrences of x_1x_2 by the variable f. This

gives the following linear inequalities:

$$\begin{aligned} f - a_2x_1 - a_1x_2 + a_1a_2 &\geq 0, \\ -f + b_2x_1 + a_1x_2 - a_1b_2 &\geq 0, \\ -f + a_2x_1 + b_1x_2 - b_1a_2 &\geq 0, \\ f - b_2x_1 - b_1x_2 + b_1b_2 &\geq 0. \end{aligned}$$

For general n, there are 2^n points to consider (i.e., all choices of each variable at a bound), and the inequality descriptions in the space of $(f, x_1, \ldots, x_n) \in \mathbb{R}^{n+1}$ get rather complicated. This is true even for $n = 3$, where the exact inequality description for the convex hull is known (see [34, 33]). It is frequent practice, both in modeling and software, to repeatedly use the McCormick inequalities when $n > 2$. Already the trilinear case, $n = 3$, is an interesting one for analysis. Here, we have three choices, which can be thought of as $f = (x_1x_2)x_3$, $f = (x_1x_3)x_2$ and $f = (x_2x_3)x_1$. Because the domain of each variable is its own interval $[a_i, b_i]$, the grouping can affect the quality of the convexification. In Chapter 3, we analytically quantify the quality of these different convexification possibilities, in addition to the trilinear convex hull itself.

There are many implementations of sBB, both commercial and open-source. For example, `BARON` [49], `Couenne` [6], `SCIP` [61], `ANTIGONE` [37], and α`BB` [1]. Both `BARON` and `ANTIGONE` use the complete linear-inequality description of the trilinear convex hull, while `Couenne`, `SCIP` and α`BB` use an *arbitrary* double-McCormick relaxation. Our results in Chapter 3 indicate that there are situations where the choice of `BARON` and `ANTIGONE` may be too heavy, and certainly even restricting to double-McCormick relaxations, `Couenne`, `SCIP` and α`BB` do not systematically choose the best one.

Trilinear monomials appear in many important models, for example, varied stochastic optimization problems such as: probabilistic facility location with random demand, pooling system problems where the quality of reservoirs is uncertain, and probabilistic response models for the propagation of epidemics ([27]). They also arise in photolithography models ([44]). However, *our results are not just relevant to trilinear monomials in formulations.* With the sBB approach for factorable formulations, our results are relevant whenever three quantities are multiplied. That is, as an expression DAG is created and auxiliary variables are introduced, a trilinear monomial will arise whenever three quantities (which can be complicated functions themselves) are multiplied.

In this thesis, we use $(n + 1)$-dimensional volume to compare different natural

convexifications of graphs of functions of n variables on the box domain $[a_1, b_1] \times \cdots \times [a_n, b_n]$. We present a complete analytic analysis of the case of $n = 3$, for all choices of $0 \leq a_i < b_i$. It is perhaps surprising that this can be carried out, and probably less surprising that the analysis is quite complicated.

Computing the volume of a polytope is well known to be strongly #P-hard (see [8]). But in fixed dimension, or, in celebrated work, by seeking an approximation via a randomized algorithm (see [15]), positive results are available. Our work though is motivated not by algorithms for volume calculation, but rather in certain situations where analytic formulae can be derived.

There have been a few papers on analytic formulae for volumes of polytopes that naturally arise in mathematical optimization; see [23], [59], [10], [4], [58], [26]. But none of these works has attempted to apply their ideas to the low-dimensional polytopes that naturally arise in sBB. One notable exception is [11], which is a mostly-computational precursor to our work, focusing on quadrilinear functions (i.e., $f = x_1x_2x_3x_4$).

Motivated by two well-studied applications (the *Molecular Distance Geometry Problem* and the *Hartree-Fock Problem*), [11] first proposed volume in the context of sBB and monomials, but they leapfrogged to the case of $n = 4$ and took a mostly experimental approach. They demonstrated that there can be a significant difference in performance depending on grouping, and they offered some guidance based on computational experiments. However, there was no firm theoretical result grounding the choice of repeated-McCormick relaxation and at the time of that work, it appeared that developing precise formulae for volumes relevant to repeated McCormick was not tractable. In contrast, we establish firm theoretical grounding (in Chapter 3), and we go a step further to see that the theory can be used to rather accurately predict the quality (as measured by objective gap) of an aggregate relaxation built from the different relaxations of individual trilinear monomials (in Chapter 4). With our present work on $n = 3$, it now seems possible that the case of $n = 4$ could be carried out.

Volume as a measure for comparing relaxations was first proposed in [25]. In fact, the *practical* use of volume as a measure for comparing relaxations in the context of non-linear mixed-integer optimization, foreshadowed by [25], was later validated computationally for a non-linear version of the uncapactitated facility-location problem (see [24]). Specifically, using volume calculations, a main mathematical result of [25] is that weak formulations of facility-location problems are very close to strong formulations when the number of facilities is small compared to the number of customers.

Then [24] showed that in this scenario, with a convex objective function, the weak formulation computationally out performs the strong formulation in the context of branch-and-bound.

The emphasis in [25, 23, 59] was not on sBB nor on low-dimensional functions. Because those results pertained to varying dimension and related asymptotics, exactly how volumes are compared and scaled was important (in particular, see [25] which defines the "idealized radial distance"). Because we now focus on low-dimensional polytopes, the exact manner of comparison and scaling is much less relevant. Using volume as a measure corresponds to a uniform distribution of the optimal solution across a relaxation. This is justified in the context of *non-linear* optimization if we want a measure that is robust across all formulations. One can well find situations where the volume measure is misleading. It would not make sense for evaluating polyhedral relaxations of the integer points in a polytope, if we were only concerned with *linear* objectives — in such a case, solutions are concentrated on the boundary and there are better measures available (see [25]). But if we are interested in a mathematically-tractable measure that robustly makes sense in the context of global optimization, volume is quite natural.

There has been considerable research on multilinear monomials and generalizations in the context of global optimization, notably [42, 31, 5, 46, 22, 35]. Most relevant to our work are: the polyhedral nature of the convexification of the graphs of multilinear functions on box domains (see [42]); the McCormick inequalities describing giving the complete linear-inequality description for bilinear functions on a box domain (see [32]); the complete linear-inequality description of the trilinear convex hull (see [34] and [33]). Our work adds to this literature.

1.1 Thesis overview

In Chapter 2, we discuss the possible convexification methods for trilinear monomials in more detail, and introduce some notation that will be helpful throughout this thesis. In Chapter 3, we compute volume formulae for these alternative convexification methods, and draw some important conclusions regarding the choice of convexification in the implementation of sBB. In Chapter 4, we present experimental work justifying the use of volume as a comparison measure in this context. The computational work for Chapter 4 was completed with the assistance of Han Yu, a University of Michigan masters student. In Chapter 5, we use our knowledge of volume from Chapter 3 to analyze the optimal choice of branching point.

CHAPTER 2

Preliminaries

Throughout this thesis, much of our analysis involves considering the three possible "double-McCormick" convexifications for trilinear monomials, alongside the convex-hull convexification. In this section, we formally define the mathematics behind these objects and set our notation.

2.1 Double McCormick

When using the double-McCormick technique to convexify trilinear monomials, a modeling/algorithmic choice is involved: we must choose which pair of variables we will apply the first iteration of McCormick. Assume that we have the variables $x_i \in [a_i, b_i]$, $i = 1, 2, 3$, and that the following conditions hold:

$$\begin{aligned} &0 \leq a_i < b_i \text{ for } i = 1, 2, 3, \quad \text{and} \\ &a_1b_2b_3 + b_1a_2a_3 \;\leq\; b_1a_2b_3 + a_1b_2a_3 \;\leq\; b_1b_2a_3 + a_1a_2b_3. \end{aligned} \tag{Ω}$$

To see this is without loss of generality, let $\mathcal{O}_i := a_i(b_jb_k) + b_i(a_ja_k)$. Then we can label the variables such that $\mathcal{O}_1 \leq \mathcal{O}_2 \leq \mathcal{O}_3$. Note that because we are only considering non-negative bounds, the latter part of this condition is equivalent to:

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

Given the trilinear monomial $f := x_1x_2x_3$, there are three choices of convexifications depending on the bilinear submonomial we convexify first. We could first group x_1 and x_2 and convexify $w = x_1x_2$; after this, we are left with the monomial $f = wx_3$, which we can also convexify using McCormick. Alternatively, we could first group variables x_1 and x_3, or variables x_2 and x_3.

2.1.1 Convexification

To see how to perform these convexifications in general, we show the double-McCormick convexification that first groups the variables x_i and x_j. Therefore, we have $f = x_i x_j x_k$, and we let $w_{ij} = x_i x_j$, so $f = w_{ij} x_k$.

Convexifying $w_{ij} = x_i x_j$, we obtain the inequalities:

$$\begin{aligned} w_{ij} - a_j x_i - a_i x_j + a_i a_j &\geq 0, \\ -w_{ij} + b_j x_i + a_i x_j - a_i b_j &\geq 0, \\ -w_{ij} + a_j x_i + b_i x_j - b_i a_j &\geq 0, \\ w_{ij} - b_j x_i - b_i x_j + b_i b_j &\geq 0. \end{aligned}$$

Convexifying $f = w_{ij} x_k$, we obtain the inequalities:

$$\begin{aligned} f - a_k w_{ij} - a_i a_j x_k + a_i a_j a_k &\geq 0, \\ -f + b_k w_{ij} + a_i a_j x_k - a_i a_j b_k &\geq 0, \\ -f + a_k w_{ij} + b_i b_j x_k - b_i b_j a_k &\geq 0, \\ f - b_k w_{ij} - b_i b_j x_k + b_i b_j b_k &\geq 0. \end{aligned}$$

Using Fourier-Motzkin elimination, we then eliminate the variable w_{ij} to obtain the following system in our original variables f, x_i, x_j and x_k. We are able to eliminate w_{ij} without any case analysis because we assume that our interval bounds are non-negative, and therefore we know the signs of all the variable coefficients. If we wanted to relax the assumption of non-negative bounds, we could perform a similar analysis, but we would have to carefully check each case because Fourier-Motzkin elimination relies on knowing the sign of every coefficient of the variable we are eliminating.

$$\begin{aligned} x_i - a_i &\geq 0, && (2.1) \\ x_j - a_j &\geq 0, && (2.2) \\ f - a_j a_k x_i - a_i a_k x_j - a_i a_j x_k + 2a_i a_j a_k &\geq 0, && (2.3) \\ f - a_j b_k x_i - a_i b_k x_j - b_i b_j x_k + a_i a_j b_k + b_i b_j b_k &\geq 0, && (2.4) \\ - x_j + b_j &\geq 0, && (2.5) \\ - x_i + b_i &\geq 0, && (2.6) \\ f - b_j a_k x_i - b_i a_k x_j - a_i a_j x_k + a_i a_j a_k + b_i b_j a_k &\geq 0, && (2.7) \end{aligned}$$

$$f - b_j b_k x_i - b_i b_k x_j - b_i b_j x_k + 2b_i b_j b_k \geq 0, \tag{2.8}$$
$$-f + b_j b_k x_i + a_i b_k x_j + a_i a_j x_k - a_i a_j b_k - a_i b_j b_k \geq 0, \tag{2.9}$$
$$-f + a_j b_k x_i + b_i b_k x_j + a_i a_j x_k - a_i a_j b_k - b_i a_j b_k \geq 0, \tag{2.10}$$
$$-x_k + b_k \geq 0, \tag{2.11}$$
$$-f + b_j a_k x_i + a_i a_k x_j + b_i b_j x_k - a_i b_j a_k - b_i b_j a_k \geq 0, \tag{2.12}$$
$$-f + a_j a_k x_i + b_i a_k x_j + b_i b_j x_k - b_i a_j a_k - b_i b_j a_k \geq 0, \tag{2.13}$$
$$x_k - a_k \geq 0, \tag{2.14}$$
$$f - a_i a_j x_k \geq 0, \tag{2.15}$$
$$-f + b_i b_j x_k \geq 0. \tag{2.16}$$

It is easy to see that the inequalities 2.15 and 2.16 are redundant: 2.15 is $a_j a_k$(2.1)+ $a_i a_k$(2.2) + (2.3), and 2.16 is $b_j a_k$(2.6) + $a_i a_k$(2.5) + (2.12).

We use the following notation in what follows. For $i = 1, 2, 3$, *system* S_i is defined to be the system of inequalities obtained by first grouping the pair of variables x_j and x_k, with j and k different from i. $\mathcal{P}_i$ is defined to be the solution set of this system.

2.1.2 Hull

As we noted earlier, a convex-hull representation for trilinear monomials is known. From [34], for any labeling that satisfies Ω (or even just: $\mathcal{O}_1 \leq \mathcal{O}_2$ and $\mathcal{O}_1 \leq \mathcal{O}_3$), this inequality system which we refer to as system S_h is:

$$f - a_2 a_3 x_1 - a_1 a_3 x_2 - a_1 a_2 x_3 + 2a_1 a_2 a_3 \geq 0, \tag{2.17}$$
$$f - b_2 b_3 x_1 - b_1 b_3 x_2 - b_1 b_2 x_3 + 2b_1 b_2 b_3 \geq 0, \tag{2.18}$$
$$f - a_2 b_3 x_1 - a_1 b_3 x_2 - b_1 a_2 x_3 + a_1 a_2 b_3 + b_1 a_2 b_3 \geq 0, \tag{2.19}$$
$$f - b_2 a_3 x_1 - b_1 a_3 x_2 - a_1 b_2 x_3 + b_1 b_2 a_3 + a_1 b_2 a_3 \geq 0, \tag{2.20}$$
$$f - \tfrac{\eta_1}{b_1 - a_1} x_1 - b_1 a_3 x_2 - b_1 a_2 x_3 + \left(\tfrac{\eta_1 a_1}{b_1 - a_1} + b_1 b_2 a_3 + b_1 a_2 b_3 - a_1 b_2 b_3 \right) \geq 0, \tag{2.21}$$
$$f - \tfrac{\eta_2}{a_1 - b_1} x_1 - a_1 b_3 x_2 - a_1 b_2 x_3 + \left(\tfrac{\eta_2 b_1}{a_1 - b_1} + a_1 a_2 b_3 + a_1 b_2 a_3 - b_1 a_2 a_3 \right) \geq 0, \tag{2.22}$$
$$-f + a_2 a_3 x_1 + b_1 a_3 x_2 + b_1 b_2 x_3 - b_1 b_2 a_3 - b_1 a_2 a_3 \geq 0, \tag{2.23}$$
$$-f + b_2 a_3 x_1 + a_1 a_3 x_2 + b_1 b_2 x_3 - b_1 b_2 a_3 - a_1 b_2 a_3 \geq 0, \tag{2.24}$$
$$-f + a_2 a_3 x_1 + b_1 b_3 x_2 + b_1 a_2 x_3 - b_1 a_2 b_3 - b_1 a_2 a_3 \geq 0, \tag{2.25}$$
$$-f + b_2 b_3 x_1 + a_1 a_3 x_2 + a_1 b_2 x_3 - a_1 b_2 b_3 - a_1 b_2 a_3 \geq 0, \tag{2.26}$$
$$-f + a_2 b_3 x_1 + b_1 b_3 x_2 + a_1 a_2 x_3 - b_1 a_2 b_3 - a_1 a_2 b_3 \geq 0, \tag{2.27}$$

$$-f + b_2b_3x_1 + a_1b_3x_2 + a_1a_2x_3 - a_1b_2b_3 - a_1a_2b_3 \geq 0, \quad (2.28)$$

$$x_1 - a_1 \geq 0, \quad (2.29)$$

$$-x_1 + b_1 \geq 0, \quad (2.30)$$

$$x_2 - a_2 \geq 0, \quad (2.31)$$

$$-x_2 + b_2 \geq 0, \quad (2.32)$$

$$x_3 - a_3 \geq 0, \quad (2.33)$$

$$-x_3 + b_3 \geq 0, \quad (2.34)$$

where $\eta_1 = b_1b_2a_3 - a_1b_2b_3 - b_1a_2a_3 + b_1a_2b_3$ and $\eta_2 = a_1a_2b_3 - b_1a_2a_3 - a_1b_2b_3 + a_1b_2a_3$.

We refer to the polytope defined as the feasible set of system S_h as $\mathcal{P}_h$. The extreme points of $\mathcal{P}_h$ are the 8 points that correspond to the $2^3 = 8$ choices of each x-variable at its upper or lower bound (see [33] for a proof). We label these 8 points (all of the form $[f = x_1x_2x_3, x_1, x_2, x_3]$) as follows:

$$v^1 := \begin{bmatrix} b_1a_2a_3 \\ b_1 \\ a_2 \\ a_3 \end{bmatrix}, \; v^2 := \begin{bmatrix} a_1a_2a_3 \\ a_1 \\ a_2 \\ a_3 \end{bmatrix}, \; v^3 := \begin{bmatrix} a_1a_2b_3 \\ a_1 \\ a_2 \\ b_3 \end{bmatrix}, \; v^4 := \begin{bmatrix} a_1b_2a_3 \\ a_1 \\ b_2 \\ a_3 \end{bmatrix},$$

$$v^5 := \begin{bmatrix} a_1b_2b_3 \\ a_1 \\ b_2 \\ b_3 \end{bmatrix}, \; v^6 := \begin{bmatrix} b_1b_2b_3 \\ b_1 \\ b_2 \\ b_3 \end{bmatrix}, \; v^7 := \begin{bmatrix} b_1b_2a_3 \\ b_1 \\ b_2 \\ a_3 \end{bmatrix}, \; v^8 := \begin{bmatrix} b_1a_2b_3 \\ b_1 \\ a_2 \\ b_3 \end{bmatrix}.$$

Each alternative double-McCormick polyhedral convexification leads to a different system of inequalities (system S_i, $i = 1, 2, 3$) and therefore a different polytope ($\mathcal{P}_i$, $i = 1, 2, 3$) in $\mathbb{R}^4$ — all three contain the convex hull of the solution set of our original trilinear monomial (on the box domain), i.e. $\mathcal{P}_h$.

To establish if one of these three convexifications is better than another, we need to be able to compare these polytopes in a quantifiable manner. We take the (4-dimensional) volume as our measure, with the idea that a smaller volume corresponds to a tighter convexification. See Chapter 4 for our computational validation of using volume in this context.

For trilinear monomials with domain being a box (in the non-negative orthant), we derive exact expressions for the (4-dimensional) volume for the convex hull of the set

of solutions, and also for each of the three possible double-McCormick convexifications (see Chapter 3). These volumes are in terms of six parameters (the upper and lower bounds on each of the three variables), and are rather complicated. By comparing the volume expressions, we are able to draw conclusions regarding the optimal way to perform double McCormick for trilinear monomials, and to measure the difference between the best double-McCormick convexification and the convex hull. In Chapter 5, we go on to use our volume results to approach the problem of calculating the optimal branching point.

2.2 Alternatives

In practice, there are many possibilities for handling each product of three terms encountered in a formulation. A good choice, which may well be different for different triple products in the same formulation, ultimately depends on trading off the tightness of a relaxation with the overhead in working with it. For clarity, in the remainder of this section, we focus on different possible treatments of $f = x_1x_2x_3$.

One possibility is to use the full trilinear hull $\mathcal{P}_h$. This representation has the benefit of using no auxiliary variables. Another possibility to use the *convex-hull representation* (see [12], for example), writing $f = \sum_{j=1}^{8} \lambda_j v^j$, with $\sum_{j=1}^{8} \lambda_j = 1$, $\lambda_j \geq 0$, for $j = 1, 2, \ldots, 8$. This formulation has the drawback of utilizing *eight auxiliary variables*. But noticing that there are 5 linear equations, we can really reduce to *three auxiliary variables*. In fact, there is a very structured way to do this, where none of the λ_j variables are employed at all, and rather we introduce *three auxiliary variables* w_{12}, w_{13} and w_{23}, which represent the products x_1x_2, x_1x_3 and x_2x_3, respectively. A strong advantage of this last approach is when terms x_1x_2, x_1x_3 and x_2x_3 are also in the model under consideration. We wish to emphasize that projecting any of these convex-hull representations (reduced or not) down to the space of (f, x_1, x_2, x_3) yields again $\mathcal{P}_h$, and so all of these representations have the same bounding power.

We are advocating the *consideration* of double-McCormick relaxations as an alternative when warranted. We have identified the best among the double McCormicks and quantified the error in using it in preference to $\mathcal{P}_h$ (and, ipso facto, with any convex-hull or reduced convex hull representation). A double-McCormick relaxation involves only *one auxiliary variable* (and 8 inequalities). This can be particularly attractive when this particular auxiliary variable already appears in the model under consideration. Alternatively, especially when this particular auxiliary variable does

not appear in the formulation, we can use the formulation with *zero auxiliary variables* (2.1-2.14). We have computationally validated such an approach in the context of "box-cubic programming" or "boxcup" problems (see Chapter 4)

$$\min_{x \in \mathbb{R}^n} \left\{ \sum_{\{i,j,k\}} q_{ijk}\, x_i\, x_j\, x_k \ : \ x_i \in [a_i, b_i],\ i = 1, 2, \ldots, n \right\}.$$

In this type of problem, we can apply (2.1-2.14) *independently* for each trinomial, with no auxiliary variables at all, choosing the best double-McCormick for each trinomial, whenever the associated volume is close to the volume for $\mathcal{P}_h$. We have documented that this can happen quite a lot, and so it is a viable approach. It is important to emphasize that some of the negative experience with double McCormick is related to choosing the wrong one. Indeed, our mathematical and computational results indicate that there are many situations where: (i) the worst double McCormick is quite bad compared to the best one, and (ii) the best one is only slightly worse than $\mathcal{P}_h$ (and its convex-hull representations).

Besides any prescriptive use of double-McCormick relaxations, our results can simply be seen as quantifying the bounding advantage given by $\mathcal{P}_h$ and the various convex-hull representations (reduced or not) as compared to each of the possible double-McCormick relaxations.

In some global-optimization software (e.g., `BARON` and `ANTIGONE`) the complicated inequality description of the trilinear hull is explicitly used. In other global-optimization software (e.g., `Couenne` and `SCIP`) and as a technique at the formulation level, repeated McCormick is used for the trilinear case. It is by no means clear that either approach should be followed all of the time (though this currently seems to be the case), because of the solution-time tradeoff in using more complicated but stronger convexifications. This effect can be especially pronounced in the case of non-linear optimization where solutions may not be on the boundary (see [24], for example). By quantifying the quality of different convexifications, we offer (i) firm and actionable means for deciding between them at run time and, (ii) some explanation for differing behavior of sBB software under different scenarios.

Finally, we note that the double-McCormick approach is often applied at the modeling level (see [27] and [40], for example). In particular, our results are highly relevant to modelers who simply use global-optimization software, often through a modeling language. An uninformed modeler can defeat clever software and therefore, it is very useful for the user to know which double McCormick to employ.

CHAPTER 3

Volume Formulae

3.1 Introduction

In this chapter, we present the volume results upon which the remainder of the dissertation will build (for the paper presenting these results see [56]). We formally present the volume formulae and the technical proofs that establish their correctness. We discuss the important corollaries that are implied by our results, and consider what impact these results may have on algorithm design. In §3.2, we present our main results and their consequences and in §§3.3–3.6, we present the proofs. In §3.7, we make brief concluding remarks and describe some future directions for investigation. The final section, §3.8, contains technical lemmas and calculations which we refer to throughout the proof sections.

3.2 Theorems

Theorem 3.1. Given that the upper and lower bound parameters respect the labeling Ω, we have:

$$\begin{aligned}\mathrm{Vol}_{\mathcal{P}_h} = &(b_1 - a_1)(b_2 - a_2)(b_3 - a_3) \times \\ &\left(b_1(5b_2b_3 - a_2b_3 - b_2a_3 - 3a_2a_3) + a_1(5a_2a_3 - b_2a_3 - a_2b_3 - 3b_2b_3)\right)/24.\end{aligned}$$

Before stating the remaining theorems, we define the following twelve points in $\mathbb{R}^4$, where $j := i + 1 \pmod 3$ and $k := i + 2 \pmod 3$:

$$v_1^9 := \begin{bmatrix} \theta_1^1 \\ \theta_1^2 \\ a_2 \\ b_3 \end{bmatrix}, v_1^{10} := \begin{bmatrix} \theta_1^3 \\ \theta_1^4 \\ b_2 \\ a_3 \end{bmatrix}, v_1^{11} := \begin{bmatrix} \theta_1^5 \\ \theta_1^6 \\ b_2 \\ a_3 \end{bmatrix}, v_1^{12} := \begin{bmatrix} \theta_1^7 \\ \theta_1^8 \\ a_2 \\ b_3 \end{bmatrix},$$

$$v_2^9 := \begin{bmatrix} \theta_2^1 \\ b_1 \\ \theta_2^2 \\ a_3 \end{bmatrix}, v_2^{10} := \begin{bmatrix} \theta_2^3 \\ a_1 \\ \theta_2^4 \\ b_3 \end{bmatrix}, v_2^{11} := \begin{bmatrix} \theta_2^5 \\ a_1 \\ \theta_2^6 \\ b_3 \end{bmatrix}, v_2^{12} := \begin{bmatrix} \theta_2^7 \\ b_1 \\ \theta_2^8 \\ a_3 \end{bmatrix},$$

$$v_3^9 := \begin{bmatrix} \theta_3^3 \\ b_1 \\ a_2 \\ \theta_3^4 \end{bmatrix}, v_3^{10} := \begin{bmatrix} \theta_3^1 \\ a_1 \\ b_2 \\ \theta_3^2 \end{bmatrix}, v_3^{11} := \begin{bmatrix} \theta_3^7 \\ a_1 \\ b_2 \\ \theta_3^8 \end{bmatrix}, v_3^{12} := \begin{bmatrix} \theta_3^5 \\ b_1 \\ a_2 \\ \theta_3^6 \end{bmatrix},$$

where:

$$\theta_i^1 = a_i a_j a_k + \frac{a_j(b_k - a_k)(b_i b_j b_k - a_i a_j a_k)}{b_j b_k - a_j a_k}, \qquad \theta_i^2 = a_i + \frac{a_j(b_i - a_i)(b_k - a_k)}{b_j b_k - a_j a_k},$$
$$\theta_i^3 = a_i a_j a_k + \frac{a_k(b_j - a_j)(b_i b_j b_k - a_i a_j a_k)}{b_j b_k - a_j a_k}, \qquad \theta_i^4 = a_i + \frac{a_k(b_j - a_j)(b_i - a_i)}{b_j b_k - a_j a_k},$$
$$\theta_i^5 = \frac{b_j a_k(a_i b_j b_k - a_i a_j b_k - b_i a_j a_k + b_i a_j b_k)}{b_j b_k - a_j a_k}, \qquad \theta_i^6 = a_i + \frac{b_j(b_i - a_i)(b_k - a_k)}{b_j b_k - a_j a_k},$$
$$\theta_i^7 = \frac{a_j b_k(b_i b_j a_k - b_i a_j a_k - a_i b_j a_k + a_i b_j b_k)}{b_j b_k - a_j a_k}, \qquad \theta_i^8 = a_i + \frac{b_k(b_j - a_j)(b_i - a_i)}{b_j b_k - a_j a_k}.$$

Theorem 3.2. Given that the upper and lower bound parameters respect the labeling Ω, we have that the set of extreme points of $\mathcal{P}_1$ is $\{v^1, \ldots, v^8\} \cup \{v_1^9, \ldots, v_1^{12}\}$. Moreover,

$$\text{Vol}_{\mathcal{P}_1} = \text{Vol}_{\mathcal{P}_h} + (b_1 - a_1)(b_2 - a_2)^2(b_3 - a_3)^2 \times \frac{3(b_1 b_2 a_3 - a_1 b_2 a_3 + b_1 a_2 b_3 - a_1 a_2 b_3) + 2(a_1 b_2 b_3 - b_1 a_2 a_3)}{24(b_2 b_3 - a_2 a_3)}.$$

Theorem 3.3. Given that the upper and lower bound parameters respect the labeling Ω, we have that the set of extreme points of $\mathcal{P}_2$ is $\{v^1, \ldots, v^8\} \cup \{v_2^9, \ldots, v_2^{12}\}$. Moreover,

$$\text{Vol}_{\mathcal{P}_2} = \text{Vol}_{\mathcal{P}_h} + \frac{(b_1 - a_1)(b_2 - a_2)^2(b_3 - a_3)^2\left(5(a_1 b_1 b_3 - a_1 b_1 a_3) + 3(b_1^2 a_3 - a_1^2 b_3)\right)}{24(b_1 b_3 - a_1 a_3)}.$$

Theorem 3.4. Given that the upper and lower bound parameters respect the labeling Ω, we have that the set of extreme points of $\mathcal{P}_3$ is $\{v^1, \ldots, v^8\} \cup \{v_3^9, \ldots, v_3^{12}\}$.

Moreover,

$$\mathrm{Vol}_{\mathcal{P}_3} = \mathrm{Vol}_{\mathcal{P}_h} + \frac{(b_1 - a_1)(b_2 - a_2)^2(b_3 - a_3)^2\,(5(a_1 b_1 b_2 - a_1 b_1 a_2) + 3(b_1^2 a_2 - a_1^2 b_2))}{24(b_1 b_2 - a_1 a_2)}.$$

From the formulae it is easy to see that the volume of $\mathcal{P}_3$ and the volume of $\mathcal{P}_2$ are essentially the same, once we take into account a relabeling of variables x_2 and x_3. However, the volume of $\mathcal{P}_1$ has another form, showing us that something different really is going on when we leave x_1 out of the first round of McCormick. This is perhaps not all that surprising, looking back at the formula for the volume of $\mathcal{P}_h$ we see that variables x_2 and x_3 are interchangeable. Furthermore, in the complete characterization of the facets of the convex hull in [33], we see that variables x_2 and x_3 are interchangeable again.

Our proofs in §§3.3–3.6 all assume that $a_1, a_2, a_3 > 0$. Next, we briefly explain why the theorems hold even when any of the a_i are zero. Taking the convex hull of a compact set is continuous (even 1-Lipschitz) in the Hausdorff metric (see [51, p. 51]). The volume functional is continuous (with respect to the Hausdorff metric) on the set K^n of convex bodies in $\mathbb{R}^n$ (see [52, Theorem 1.8.20; p. 68]). If two sets of m points in $\mathbb{R}^n$ are close as vectors in $\mathbb{R}^{mn}$, then they are also close in the Hausdorff metric. Therefore, the volume of the convex hull of a set of m points in $\mathbb{R}^n$ is a continuous function of the coordinates of the points. Also, the coordinates of the extreme points of our polytopes are all continuous functions (of the six parameters) at $a_i = 0$. Finally, we note that the volume formulae that we derive are continuous functions (of the six parameters) at $a_i = 0$. Therefore, those formulae are also correct when some $a_i = 0$. We do note that we can also modify our constructions to handle these cases where some of the a_i are zero, but our continuity argument is much shorter.

Corollary 3.5. For all values of the parameters a_1, b_1, a_2, b_2, a_3, b_3, meeting the conditions (Ω), we have: $\mathrm{Vol}_{\mathcal{P}_h} \leq \mathrm{Vol}_{\mathcal{P}_3} \leq \mathrm{Vol}_{\mathcal{P}_2} \leq \mathrm{Vol}_{\mathcal{P}_1}$.

From this we can see that with the variables ordered according to their upper and lower bounds per (Ω), the least (double-McCormick) volume will always be obtained by using system S_3 (i.e., first grouping variables x_1 and x_2). In addition, for different values of the upper and lower bounds, we can precisely quantify the difference in volume of the alternative convexifications.

Moreover, by substituting $a_1 = a_2 = 0$ and $b_1 = b_2 = 1$ into the conditions (Ω), we can easily see the following corollary relevant to mixed-integer non-linear optimization.

Corollary 3.6. In the special case where $a_i = a_j = 0$ and $b_i = b_j = 1$, first grouping the two $[0, 1]$-variables gives the convexification with the least volume.

In this special case, we only have two parameters a_3 and b_3 and the volume formulae simplify considerably. In particular, for this special case, $\mathcal{P}_3$ is equal to $\mathcal{P}_h$, and $\mathcal{P}_1$ and $\mathcal{P}_2$ are equivalent (by this we mean that $\text{Vol}_{\mathcal{P}_1} = \text{Vol}_{\mathcal{P}_2}$, but $\mathcal{P}_1 \neq \mathcal{P}_2$). We compute the difference in volume between the two distinct choices of convexification and, in Figure 3.1, plot this expression as the parameters vary (satisfying $0 \leq a_3 < b_3$). The following is easy to establish.

Corollary 3.7. When $a_1 = a_2 = 0$ and $b_1 = b_2 = 1$, as a_3 and b_3 increase, the difference in volumes of $\mathcal{P}_3$ and $\mathcal{P}_1$ (or $\mathcal{P}_2$) becomes arbitrarily large. Additionally, for a fixed b_3, the greatest difference in volume occurs when $a_3 = b_3/3$.

Finally, we note that in the special case in which $a_1 = a_2 = a_3 = 0$, each convexification reduces to the convex hull, which is a result of [46]. So in this case, an *arbitrary* double-McCormick convexification has the power of the more-complicated inequality description of the convex hull. In fact, viewed this way, our results provide a quantified generalization of this result of [46]. We do wish to emphasize that because our results do not just apply to trilinear monomials on the formulation variables, but may well involve auxiliary variables, *the case of non-zero lower bounds is very relevant.*

3.3 Proof of Thm. 3.1

We compute the volume of $\mathcal{P}_h$ by constructing a triangulation, and we will repeatedly use the fact that the volume of an n-simplex in $\mathbb{R}^n$ with vertices $(z^0, \ldots, z^n)$ is:

$$|\det(z^1 - z^0 \;\; z^2 - z^0 \;\; \ldots \;\; z^n - z^0)|/n!\,.$$

See Figure 3.2 for a diagram of the 8 extreme points of $\mathcal{P}_h$. Note that v^2, which has all of the variables at their lower bounds, is at the bottom of the "inner cube",

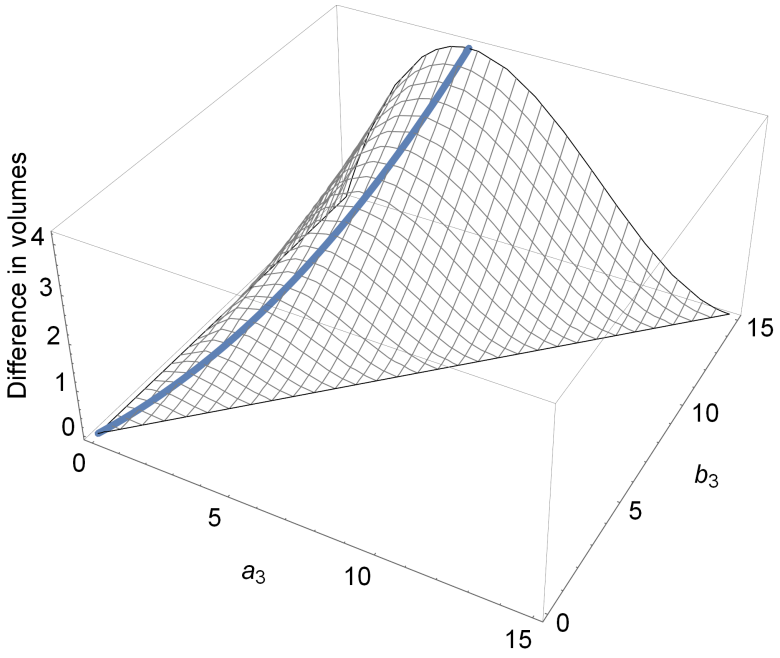


Figure 3.1: Difference in volume between $\mathcal{P}_3$ and $\mathcal{P}_1$ $\left(\frac{3a_3(b_3-a_3)^2}{24b_3}\right)$ vs. parameters a_3 and b_3 ($a_1 = a_2 = 0$ and $b_1 = b_2 = 1$)

and v^6, which has all of the variables at their upper bounds, is at the top of the "outer cube".

We will begin with five of the extreme points (a simplex) and 'add' the remaining points, keeping track of the total volume at each step, until we have computed the total volume of $\mathcal{P}_h$. We begin with the 4-simplex with extreme points v^1, v^2, v^4, v^5 and v^6, which we define as $\mathcal{S} := \text{conv}\{v^1, v^2, v^4, v^5, v^6\}$.

The volume of the 4-simplex, $\mathcal{S}$, is

$$(b_1 - a_1)^2(b_2 - a_2)(b_3 - a_3)(b_2b_3 - a_2a_3)/24.$$

A 4-simplex has 5 facets, each of which is a 3-simplex and is described by the hyperplane through a choice of 4 extreme points. To determine the facet-describing inequalities, we compute each hyperplane and then check the final point to obtain the direction of the inequality. The 5 facets of $\mathcal{S}$ are described as follows:

F^1 (hyperplane through points v^1, v^2, v^4, v^6):

$$-f + a_2a_3x_1 + a_1a_3x_2 + \frac{(a_1a_2a_3 - a_1b_2a_3 - b_1a_2a_3 + b_1b_2b_3)}{(b_3 - a_3)}x_3$$
$$- \frac{(a_1a_2a_3b_3 - a_1b_2a_3^2 - b_1a_2a_3^2 + b_1b_2a_3b_3)}{(b_3 - a_3)} \geq 0$$

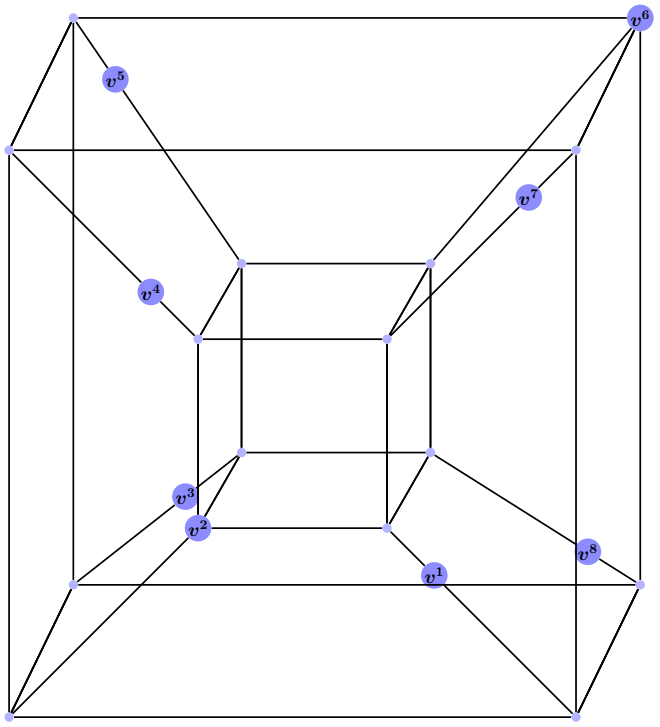


Figure 3.2: Visual representation of the convex-hull extreme points

F^2 (hyperplane through points v^1, v^2, v^4, v^5):

$$f - a_2a_3x_1 - a_1a_3x_2 - a_1b_2x_3 + a_1a_2a_3 + a_1b_2a_3 \geq 0$$

F^3 (hyperplane through points v^1, v^2, v^5, v^6):

$$(b_3 - a_3)x_2 - (b_2 - a_2)x_3 + b_2a_3 - a_2b_3 \geq 0$$

F^4 (hyperplane through points v^1, v^4, v^5, v^6):

$$f - b_2b_3x_1 - \frac{(a_1b_2a_3 - a_1b_2b_3 - b_1a_2a_3 + b_1b_2b_3)}{(b_2 - a_2)}x_2 - a_1b_2x_3 + \frac{(-a_1a_2b_2b_3 + a_1b_2^2a_3 - b_1a_2b_2a_3 + b_1b_2^2b_3)}{(b_2 - a_2)} \geq 0$$

F^5 (hyperplane through points v^2, v^4, v^5, v^6):

$$-f + b_2b_3x_1 + a_1a_3x_2 + a_1b_2x_3 - a_1b_2a_3 - a_1b_2b_3 \geq 0$$

If a hyperplane H intersects a polytope P on a facet F, then H^+ (resp., H^-) denotes the half-space determined by H that contains (does not contain) P. If a point w is not in H but in H^+ (resp., H^-), then w is *beneath* (*beyond*) F (see [17, p. 78]).

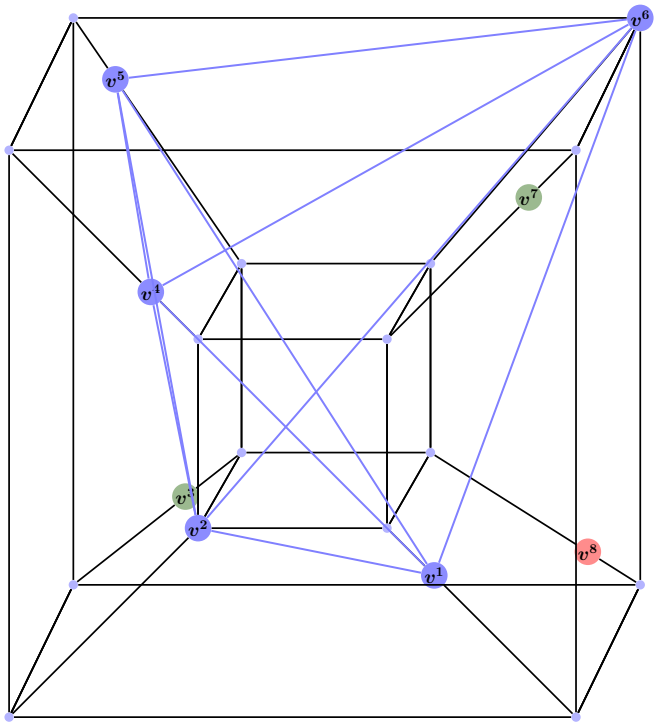


Figure 3.3: Visual representation of adding point v^8 to simplex $\mathcal{S}$

We now compute the volume of $\text{conv}(\mathcal{S} \cup \{v^8\})$. See Figure 3.3 for a visual representation of this. To obtain the additional volume of this polytope compared with $\mathcal{S}$, we sum the volume of $\text{conv}(\{v^8\} \cup F)$ for each facet, F, of $\mathcal{S}$ such that v^8 is beyond that facet. To do this, we first check each of the 5 facets to determine if v^8 is beneath or beyond that facet. To do this, we substitute v^8 into the relevant inequality, and if the result is negative then v^8 lies beyond that facet.

It is easy to check that v^8 satisfies F^1 and F^5 and violates F^3. Using Lemma 3.9, we also check that v^8 satisfies both F^2 and F^4. From this, we have that v^8 is beyond one facet, F^3. Therefore, we need to calculate the volume of $\text{conv}(F^3 \cup \{v^8\}) = \text{conv}\{v^1, v^2, v^5, v^6, v^8\}$, this a a 4-simplex with volume:

$$(b_1 - a_1)^2(b_2 - a_2)(b_3 - a_3)(b_2 b_3 - a_2 a_3)/24.$$

We now have a new polytope which is $\text{conv}\{v^1, v^2, v^4, v^5, v^6, v^8\} = \text{conv}(\mathcal{S} \cup \{v^8\})$. We refer to this polytope as $\mathcal{Q}$. The volume of $\mathcal{Q}$ is given by the sum of the volumes of the two simplices we have computed thus far. The facets of $\mathcal{Q}$ are the facets of the original simplex without F^3, along with the facets of the 4-simplex: $\text{conv}(F^3 \cup \{v^8\})$ (again not including F^3 itself). A facet of $\text{conv}(F^3 \cup \{v^8\})$ is supported by a hyperplane through a choice of 4 of the 5 extreme points (points v^1, v^2, v^5, v^6 and v^8). As before, to determine these facet inequalities, we compute each hyperplane and then check the

final point to obtain the direction of the inequality (note that we exclude the choice v^1, v^2, v^5, v^6 because this corresponds to F^3). The 4 facets are described below:

F^6 (plane through points v^1, v^2, v^5, v^8):

$$f - a_2a_3x_1 - \frac{(-a_1a_2a_3 + a_1b_2b_3 + b_1a_2a_3 - b_1a_2b_3)}{(b_2 - a_2)}x_2 - b_1a_2x_3 + \frac{(-a_1a_2^2a_3 + a_1a_2b_2b_3 - b_1a_2^2b_3 + b_1a_2b_2a_3)}{(b_2 - a_2)} \geq 0$$

F^7 (plane through points v^1, v^5, v^6, v^8):

$$f - b_2b_3x_1 - b_1b_3x_2 - b_1a_2x_3 + b_1a_2b_3 + b_1b_2b_3 \geq 0$$

F^8 (plane through points v^2, v^5, v^6, v^8):

$$-f + b_2b_3x_1 + b_1b_3x_2 + \frac{(-a_1a_2a_3 + a_1b_2b_3 + b_1a_2b_3 - b_1b_2b_3)}{(b_3 - a_3)}x_3 - \frac{(-a_1a_2a_3b_3 + a_1b_2b_3^2 + b_1a_2b_3^2 - b_1b_2a_3b_3)}{(b_3 - a_3)} \geq 0$$

F^9 (plane through points v^1, v^2, v^6, v^8):

$$-f + a_2a_3x_1 + b_1b_3x_2 + b_1a_2x_3 - b_1a_2a_3 - b_1a_2b_3 \geq 0$$

The facets of $\mathcal{Q} = \text{conv}\{v^1, v^2, v^4, v^5, v^6, v^8\}$ are therefore F^1, F^2, F^4, F^5, F^6, F^7, F^8 and F^9.

To obtain the entire volume of $\mathcal{P}_h$, we need to consider two further extreme points: v^3 and v^7 (see Figure 3.4). It would be convenient to add these points separately; i.e., compute the additional volume each produces when added to $\mathcal{Q}$, and sum the results. As the following lemma shows, this will give the correct volume if the intersection of the line segment between these points and $\mathcal{Q}$ is not empty.

Lemma 3.8. Let P be a convex polytope and let w_1 and w_2 be points not in P. Let $L(w_1, w_2)$ be the line segment between w_1 and w_2. If $L(w_1, w_2) \cap P \neq \emptyset$, then $conv(P, w_1) \cup conv(P, w_2)$ is convex. Moreover, in this case, $conv(P, w_1, w_2) = conv(P, w_1) \cup conv(P, w_2)$.

Proof. First, we show that $conv(P, w_1) \cup conv(P, w_2)$ is convex. If we show that $L(w_1, w_2)$ is completely contained in $conv(P, w_1) \cup conv(P, w_2)$, then we will be done. Choose $z \in L(w_1, w_2) \cap P$. Now consider $L(w_1, z)$. Because $z \in P$, this whole line segment must be in $conv(P, w_1)$. Similarly consider $L(z, w_2)$; this whole line segment

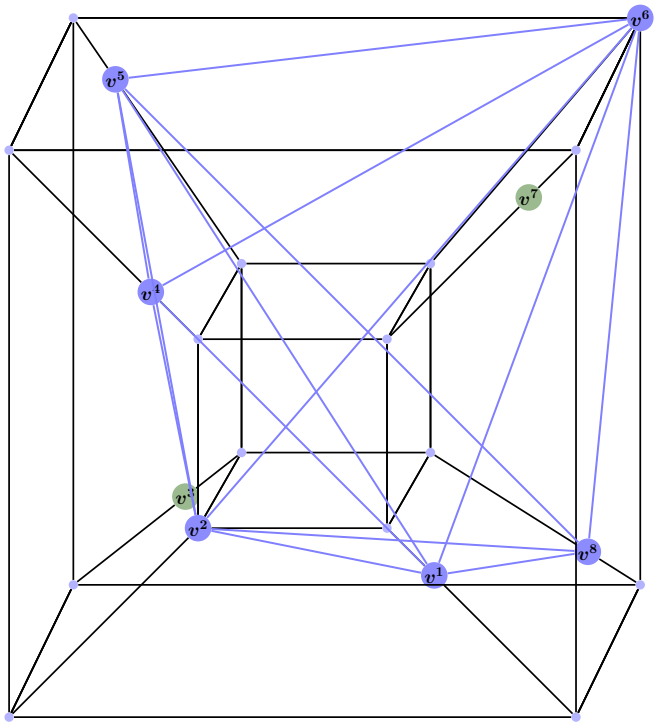


Figure 3.4: Visual representation of adding points v^3 and v^7 to polytope $\mathcal{Q}$

must be contained in $conv(P, w_2)$. Therefore the whole line segment $L(w_1, w_2)$ must be contained in $conv(P, w_1) \cup conv(P, w_2)$ and therefore this set is convex.

Next, we demonstrate that $conv(P, w_1, w_2) = conv(P, w_1) \cup conv(P, w_2)$. First, choose $y \in conv(P, w_1) \cup conv(P, w_2)$; therefore $y \in conv(P, w_1)$ or $y \in conv(P, w_2)$ (or both); in either case it is clear that $y \in conv(P, w_1, w_2)$. In the other direction, choose $y \in conv(P, w_1, w_2)$; therefore y can be written as a convex combination of the extreme points of P and w_1 and w_2. Because $conv(P, w_1) \cup conv(P, w_2)$ is convex, this means $y \in conv(P, w_1) \cup conv(P, w_2)$. Therefore the sets are equal as required. □

We refer to the midpoint of the line between w_1 and w_2 as $M(w_1, w_2)$. To show that the intersection of $L(v^3, v^7)$ and $\mathcal{Q}$ is non-empty, consider the midpoint

$$M(v^3, v^7) = \left[\begin{array}{cccc} \frac{a_1a_2b_3+b_1b_2a_3}{2} & \frac{a_1+b_1}{2} & \frac{a_2+b_2}{2} & \frac{b_3+a_3}{2} \end{array} \right].$$

We show that this point satisfies each of the inequalities of $\mathcal{Q}$ by substituting into each inequality and checking the result. By showing that each resulting quantity is non-negative, we conclude that the midpoint intersects $\mathcal{Q}$. It is easy to see that the midpoint $M(v^3, v^7)$ satisfies F^1, F^5, F^8 and F^9. Using Lemma 3.9, we also check that

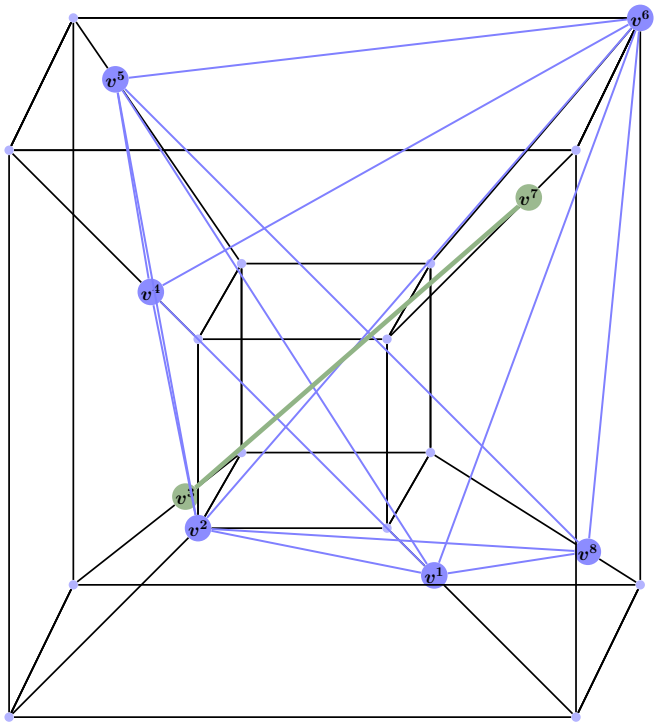


Figure 3.5: The line segment between v^3 and v^7 intersects polytope $\mathcal{Q}$

$M(v^3, v^7)$ satisfies F^2, F^4, F^6 and F^7. Therefore $\text{conv}(\mathcal{Q} \cup \{v^3\}) \cup \text{conv}(\mathcal{Q} \cup \{v^7\}) = \text{conv}(\mathcal{Q} \cup \{v^3\} \cup \{v^7\}) = \mathcal{P}_h$.

Intuitively, we can think of this as v^3 and v^7 being on opposite "sides" of the polytope, this is illustrated in Figure 3.5.

Computing the (additional) volume of $\text{conv}(\mathcal{Q} \cup \{v^3\})$. We now compute the additional volume of $\text{conv}(\mathcal{Q} \cup \{v^3\})$ compared to the volume of $\mathcal{Q}$. To obtain this, we sum the volumes of $\text{conv}(\{v^3\} \cup F)$ for each facet, F, of $\mathcal{Q}$ such that v^3 is beyond that facet. We substitute v^3 into each relevant inequality, and if the result is negative then v^3 lies beyond that facet. It is easy to see that v^3 satisfies F^5, F^7 and F^9 and violates F^2, F^6 and F^8. It can then be checked that v^3 satisfies F^1 using Lemma 3.12 (with $A = b_2, B = a_2, C = (b_1b_3 - a_1a_3), D = (2a_1a_3 - a_1b_3 - b_1a_3)$). We also check that v^3 satisfies F^4 using Lemma 3.12 (with $A = b_2, B = a_2, C = (a_1a_3 - 2a_1b_3 + b_1b_3), D = (a_1b_3 - b_1a_3)$) and Lemma 3.9.

From this, we know that v^3 is beyond F^2, F^6 and F^8; therefore, we need to compute the volume of the convex hulls of v^3 with each of these facets.

The polytope $\text{conv}(F^2 \cup \{v^3\}) = \text{conv}\{v^1, v^2, v^4, v^5, v^3\}$ is a 4-simplex with volume:

$$a_1(b_1 - a_1)(b_2 - a_2)^2(b_3 - a_3)^2/24.$$

The polytope $\text{conv}(F^6 \cup \{v^3\}) = \text{conv}\{v^1, v^2, v^5, v^8, v^3\}$ is a 4-simplex with volume:

$$a_2(b_1 - a_1)^2(b_2 - a_2)(b_3 - a_3)^2/24.$$

The polytope $\text{conv}(F^8 \cup \{v^3\}) = \text{conv}\{v^2, v^5, v^6, v^8, v^3\}$ is a 4-simplex with volume:

$$b_3(b_1 - a_1)^2(b_2 - a_2)^2(b_3 - a_3)/24.$$

Computing the (additional) volume of $\text{conv}(\mathcal{Q} \cup \{v^7\})$. We now compute the additional volume of $\text{conv}(\mathcal{Q} \cup \{v^7\})$ compared to the volume of $\mathcal{Q}$. To obtain this, we sum the volumes of $\text{conv}(\{v^7\} \cup F)$ for each facet, F, of $\mathcal{Q}$ such that v^7 is beyond that facet. We substitute v^7 into each relevant inequality, and if the result is negative then v^7 lies beyond that facet. It is easy to see that v^7 satisfies F^2, F^5 and F^9 and violates F^1, F^4 and F^7. It can then be checked that v^7 satisfies F^6 using Lemma 3.12 (with $A = b_2, B = a_2, C = (b_1a_3 - a_1b_3), D = (a_1a_3 - 2b_1a_3 + b_1b_3)$) and Lemma 3.9. We also check that v^7 satisfies F^8 using Lemma 3.12 (with $A = b_2, B = a_2, C = (2b_1b_3 - a_1b_3 - b_1a_3), D = (a_1a_3 - b_1b_3)$).

From this, we know that v^7 is beyond F^1, F^4 and F^7, therefore we need to compute the volume of the convex hulls of v^7 with each of these facets.

The polytope $\text{conv}(F^1 \cup \{v^7\}) = \text{conv}\{v^1, v^2, v^4, v^6, v^7\}$ is a 4-simplex with volume:

$$a_3(b_1 - a_1)^2(b_2 - a_2)^2(b_3 - a_3)/24.$$

The polytope $\text{conv}(F^4 \cup \{v^7\}) = \text{conv}\{v^1, v^4, v^5, v^6, v^7\}$ is a 4-simplex with volume:

$$b_2(b_1 - a_1)^2(b_2 - a_2)(b_3 - a_3)^2/24.$$

The polytope $\text{conv}(F^7 \cup \{v^7\}) = \text{conv}\{v^1, v^5, v^6, v^8, v^7\}$ is a 4-simplex with volume:

$$b_1(b_1 - a_1)(b_2 - a_2)^2(b_3 - a_3)^2/24.$$

To compute the volume of $\mathcal{P}_h$, we sum the volume of the appropriate eight simplices, and we obtain the volume of $\mathcal{P}_h$ as stated in Theorem 3.1.

Figure 3.6 gives a visual representation of the convex-hull polytope triangulated to compute its volume.

□

3.3.1 An interesting note about $\mathcal{P}_h$

In [18], the authors enumerate the number of different combinatorial types of 4-dimensional simplicial polytopes with 8 vertices (there are 37 distinct classes). A simplicial polytope is a polytope such that each of its facets is a simplex. When

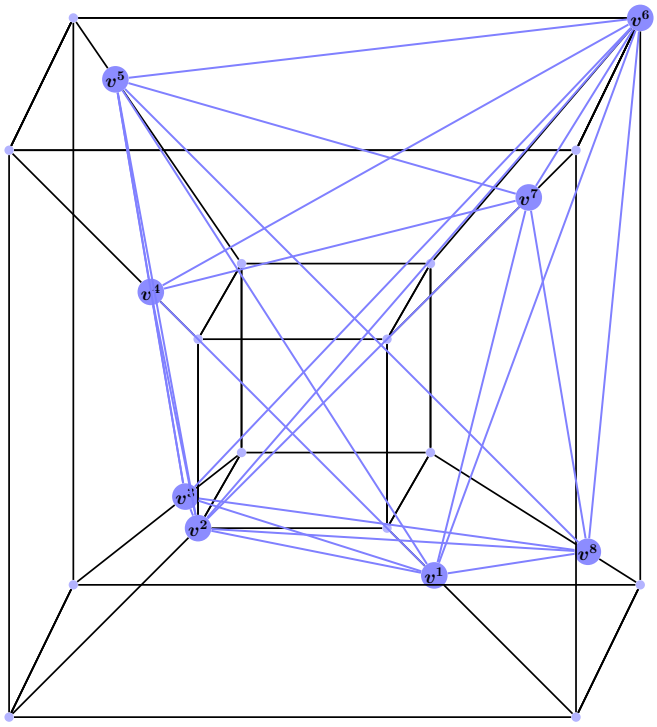


Figure 3.6: Visual representation of the convex-hull polytope

$a_i > 0$ for all i, the ploytope $\mathcal{P}_h$ is a 4-dimensional simplicial polytope with 8 vertices. Therefore, we are able to establish its combinatorial type. We know that $\mathcal{P}_3$ has 18 facets, therefore, in the notation of [18], we know that it must be one of: P^8_{23}, P^8_{24}, P^8_{25}, P^8_{27}, P^8_{28}, P^8_{29}. With some calculation, we establish that it is type P^8_{29}.

3.4 Proof of Thm. 3.4

To compute the volume of polytope $\mathcal{P}_3$, we compute the volume of the convex hull of the 12 extreme points that we claim are exactly the extreme points of system $\mathcal{P}_3$. In computing the volume of this polytope, we also prove that these are the correct extreme points and therefore that the volume we have computed is indeed the volume of $\mathcal{P}_3$.

The relevant points are the eight extreme points of $\mathcal{P}_h$, plus an additional four points. Because we have already computed the volume of $\mathcal{P}_h$, to compute the volume of $\mathcal{P}_3$, we need to compute the additional volume, compared with $\mathcal{P}_h$, added by these four extra extreme points. To show that this is indeed the volume of $\mathcal{P}_3$, we keep track of which facets need to be deleted and added to the system of inequalities as we go. In §3.4.1 we provide more details concerning this part of the proof. When it

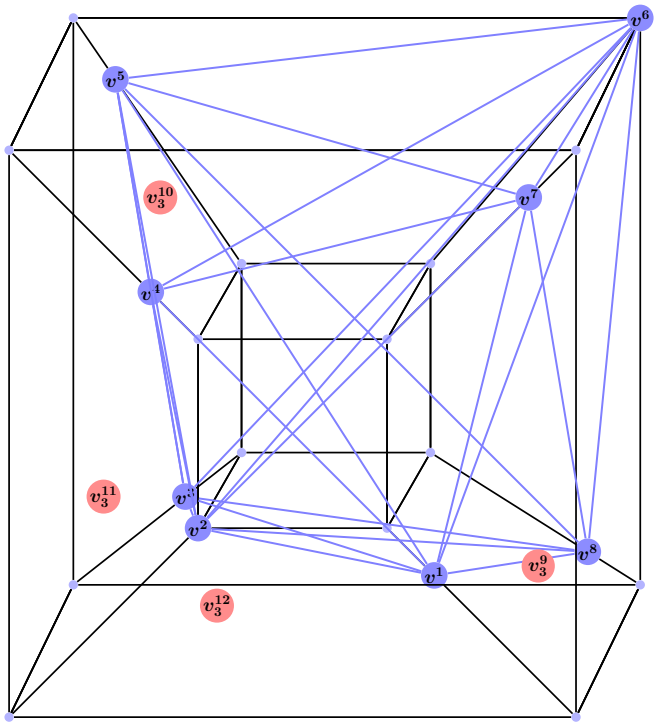


Figure 3.7: Visual representation of the convex-hull polytope (blue) and the four 'extra' extreme points of $\mathcal{P}_3$

is complete, we have exactly system S_3, and therefore we must also have the correct extreme points. See Figure 3.7 for a visual representation of the extreme points of $\mathcal{P}_3$.

We begin this proof with system S_h from §2.1.2. As discussed in §3.3, it would be convenient to add the four new extreme points to $\mathcal{P}_h$ separately; i.e., compute the additional volume each produces when added to $\mathcal{P}_h$, and sum the results. To show that we can add two points separately and obtain the correct volume, we show that the intersection of the line segment between these points and $\mathcal{P}_h$ is non-empty (Lemma 3.8).

We show that we can add v_3^9 separately, v_3^{10} separately, and then v_3^{11} and v_3^{12} together by considering the midpoints of the line segments between the relevant points. We consider $L(v_3^9, v_3^{10})$, $L(v_3^9, v_3^{11})$, $L(v_3^9, v_3^{12})$, $L(v_3^{10}, v_3^{11})$ and $L(v_3^{10}, v_3^{12})$. We show that the midpoint of each line segment satisfies each of the inequalities of $\mathcal{P}_h$ by substituting this point into each inequality and checking the result. See Table 3.1 for a summary of the resulting substitutions. The table notes whether non-negativity of the resulting quantity follows immediately (after factoring), or by use of a technical lemma (after further explanation in the §3.8.2), or after being rewritten in the way

referenced in Figure 3.8. Because we have shown that each resulting quantity is non-negative, we conclude that each of the midpoints intersect $\mathcal{P}_h$, and therefore we can add v_3^9 separately, v_3^{10} separately, and then v_3^{11} and v_3^{12} together.

Ineq	$M(v_3^9, v_3^{10})$	$M(v_3^9, v_3^{11})$	$M(v_3^9, v_3^{12})$	$M(v_3^{10}, v_3^{11})$	$M(v_3^{10}, v_3^{12})$
2.17	immediate	immediate	immediate	immediate	immediate
2.18	immediate	immediate	immediate	immediate	immediate
2.19	immediate	by Lemma 3.9	immediate	by Lemma 3.9	by Lemma 3.9
2.20	immediate	by Lemma 3.9	by Lemma 3.9	immediate	by Lemma 3.9
2.21	immediate	by Lemma 3.9	immediate	by Lemma 3.9	by Lemma 3.9
2.22	immediate	by Lemma 3.9	by Lemma 3.9	immediate	by Lemma 3.9
2.23	immediate	immediate	immediate	immediate	immediate
2.24	immediate	immediate	immediate	immediate	immediate
2.25	see 3.1	see §3.8.2.1	immediate	see 3.2	see 3.3
2.26	see 3.4	see 3.5	see 3.2	immediate	See §3.8.2.2
2.27	immediate	immediate	immediate	immediate	immediate
2.28	immediate	immediate	immediate	immediate	immediate
2.29	immediate	immediate	immediate	immediate	immediate
2.30	immediate	immediate	immediate	immediate	immediate
2.31	immediate	immediate	immediate	immediate	immediate
2.32	immediate	immediate	immediate	immediate	immediate
2.33	immediate	immediate	immediate	immediate	immediate
2.34	immediate	immediate	immediate	immediate	immediate

Table 3.1: Summary of midpoint substitutions for Thm. 3.4

$$\frac{(b_2-a_2)(b_1-a_1)\,(b_1b_3(b_2-a_2)+a_2a_3(b_1-a_1))}{2(b_1b_2-a_1a_2)} \tag{3.1}$$

$$\frac{(b_2b_3-a_2a_3)(b_1-a_1)+(b_1b_3-a_1a_3)(b_2-a_2)}{2} \tag{3.2}$$

$$\frac{(b_2-a_2)(b_1-a_1)\,(b_1(b_2b_3-a_2a_3)+a_2(b_1b_3-a_1a_3))}{2(b_1b_2-a_1a_2)} \tag{3.3}$$

$$\frac{(b_2-a_2)(b_1-a_1)(b_2b_3(b_1-a_1)+a_1a_3(b_2-a_2))}{2(b_1b_2-a_1a_2)} \tag{3.4}$$

$$\frac{(b_2-a_2)(b_1-a_1)\,(b_2(b_1b_3-a_1a_3)+a_1(b_2b_3-a_2a_3))}{2(b_1b_2-a_1a_2)} \tag{3.5}$$

Figure 3.8: For Table 3.1

Computing the (additional) volume of conv($\mathcal{P}_h \cup \{v_3^9\}$). We now compute the additional volume of conv($\mathcal{P}_h \cup \{v_3^9\}$) compared to the volume of $\mathcal{P}_h$. To do this, we sum the volumes of conv($\{v_3^9\} \cup F$) for each facet, F, of $\mathcal{P}_h$ such that v_3^9 is beyond that facet. We substitute v_3^9 into each inequality of system S_h, and we immediately see that it satisfies every inequality except 2.25.

From this, we know that v_3^9 is beyond only one facet. The extreme points that lie on this facet are points v^1, v^2, v^6 and v^8. The polytope conv$\{v^1, v^2, v^6, v^8, v_3^9\}$ is a 4-simplex with volume:

$$b_1 a_2 (b_1 - a_1)^2 (b_2 - a_2)^2 (b_3 - a_3)^2 / \left(24(b_1 b_2 - a_1 a_2)\right).$$

The facets of conv($\mathcal{P}_h \cup \{v_3^9\}$) are the facets of $\mathcal{P}_h$ except inequality 2.25. We see this by computing the four additional facets that come from adding v_3^9 and noting they are already contained in system S_h:

- The facet through points v^1, v^2, v^6 and v_3^9 is 2.23.
- The facet through points v^1, v^2, v^8 and v_3^9 is 2.31.
- The facet through points v^1, v^6, v^8 and v_3^9 is 2.30.
- The facet through points v^2, v^6, v^8 and v_3^9 is 2.27.

Computing the (additional) volume of conv($\mathcal{P}_h \cup \{v_3^{10}\}$). We now compute the additional volume of conv($\mathcal{P}_h \cup \{v_3^{10}\}$) compared to the volume of $\mathcal{P}_h$. To do this, we sum the volumes of conv($\{v_3^{10}\} \cup F$) for each facet, F, of $\mathcal{P}_h$ such that v_3^{10} is beyond that facet. We substitute v_3^{10} into each inequality of system S_h, and we immediately see that every inequality is satisfied except 2.26.

From this, we know that v_3^{10} is beyond only one facet. The extreme points that lie on this facet are points v^2, v^4, v^5 and v^6. The polytope conv$\{v^2, v^4, v^5, v^6, v_3^{10}\}$ is a 4-simplex with volume:

$$a_1 b_2 (b_1 - a_1)^2 (b_2 - a_2)^2 (b_3 - a_3)^2 / \left(24(b_1 b_2 - a_1 a_2)\right).$$

The facets of conv($\mathcal{P}_h \cup \{v_3^{10}\}$) are the facets of $\mathcal{P}_h$ except inequality 2.26. We see this by computing the four additional facets that come from adding v_3^{10} and noting that they are already contained in system S_h:

- The facet through points v^2, v^4, v^5 and v_3^{10} is 2.29.
- The facet through points v^2, v^4, v^6 and v_3^{10} is 2.24.

- The facet through points v^2, v^5, v^6 and v_3^{10} is 2.28.
- The facet through points v^4, v^5, v^6 and v_3^{10} is 2.32.

Computing the (additional) volume of conv($\mathcal{P}_h \cup \{v_3^{11}\} \cup \{v_3^{12}\}$). We now compute the additional volume of conv($\mathcal{P}_h \cup \{v_3^{11}\} \cup \{v_3^{12}\}$) compared to the volume of $\mathcal{P}_h$. Because $L(v_3^{11}, v_3^{12})$ lies entirely outside of $\mathcal{P}_3$, we need to add them sequentially.

We first compute the additional volume of conv($\mathcal{P}_h \cup \{v_3^{11}\}$) compared to the volume of $\mathcal{P}_h$. As we have done previously, we sum the volumes of conv($\{v_3^{11}\} \cup F$) for each facet, F, of $\mathcal{P}_h$ such that v_3^{11} is beyond that facet. We substitute v_3^{11} into each relevant inequality, and if the result is negative then v_3^{11} lies beyond that facet. It is immediate that v_3^{11} violates inequalities 2.19–2.22 and satisfies inequalities 2.17-2.18, 2.23-2.24 and 2.26-2.34. To see that inequality 2.25 is also satisfied see §3.8.2.3.

Therefore, we have that v_3^{11} is beyond four facets, and we need to compute the volume of the convex hulls of v_3^{11} with each of these facets.

The extreme points that lie on the first facet are points v^1, v^3, v^5 and v^8. The polytope conv$\{v^1, v^3, v^5, v^8, v_3^{11}\}$ is a 4-simplex with volume:

$$a_1 b_1 (b_1 - a_1)(b_2 - a_2)^3 (b_3 - a_3)^2 / \left(24(b_1 b_2 - a_1 a_2)\right).$$

The extreme points that lie on the second facet are points v^1, v^4, v^5 and v^7. The polytope conv$\{v^1, v^4, v^5, v^7, v_3^{11}\}$ is a 4-simplex with volume:

$$a_1 b_2 (b_1 - a_1)^2 (b_2 - a_2)^2 (b_3 - a_3)^2 / \left(24(b_1 b_2 - a_1 a_2)\right).$$

The extreme points that lie on the third facet are points v^1, v^5, v^7 and v^8. The polytope conv$\{v^1, v^5, v^7, v^8, v_3^{11}\}$ is a 4-simplex with volume:

$$a_1 b_1 (b_1 - a_1)(b_2 - a_2)^3 (b_3 - a_3)^2 / \left(24(b_1 b_2 - a_1 a_2)\right).$$

The extreme points that lie on the fourth facet are points v^1, v^3, v^4 and v^5. The polytope conv$\{v^1, v^3, v^4, v^5, v_3^{11}\}$ is a 4-simplex with volume:

$$a_1 b_2 (b_1 - a_1)^2 (b_2 - a_2)^2 (b_3 - a_3)^2 / \left(24(b_1 b_2 - a_1 a_2)\right).$$

We now have a new polytope which is conv($\mathcal{P}_h \cup \{v_3^{11}\}$). We refer to this polytope as $\mathcal{T}_3$, and we compute the facets of $\mathcal{T}_3$.

We begin with the facets of $\mathcal{P}_h$ and delete the four facets that v_3^{11} violated (2.19-2.22). Let us call this system $\mathcal{T}_3^-$. Now consider the four simplices we dealt with when computing the additional volume produced with v_3^{11}. Each of these simplices

has 5 facets; one of which corresponds to a deleted facet of $\mathcal{P}_h$.

The remaining 4 facets of the first simplex are described by the planes through the following sets of points: $\{v^1, v^3, v^5, v_3^{11}\}$, $\{v^1, v^3, v^8, v_3^{11}\}$, $\{v^1, v^5, v^8, v_3^{11}\}$ and $\{v^3, v^5, v^8, v_3^{11}\}$.

The remaining 4 facets of the second simplex are described by the planes through the following sets of points: $\{v^1, v^4, v^5, v_3^{11}\}$, $\{v^1, v^4, v^7, v_3^{11}\}$, $\{v^1, v^5, v^7, v_3^{11}\}$ and $\{v^4, v^5, v^7, v_3^{11}\}$.

The remaining 4 facets of the third simplex are described by the planes through the following sets of points: $\{v^1, v^5, v^7, v_3^{11}\}$, $\{v^1, v^5, v^8, v_3^{11}\}$, $\{v^1, v^7, v^8, v_3^{11}\}$ and $\{v^5, v^7, v^8, v_3^{11}\}$.

The remaining 4 facets of the fourth simplex are described by the planes through the following sets of points: $\{v^1, v^3, v^4, v_3^{11}\}$, $\{v^1, v^3, v^5, v_3^{11}\}$, $\{v^1, v^4, v^5, v_3^{11}\}$ and $\{v^3, v^4, v^5, v_3^{11}\}$.

Consider these sixteen facets and exclude the facets that are shared by more than one simplex. This leaves eight facets.

We can compute these eight facets to obtain the following:

- The facet through points v^1, v^3, v^8 and v_3^{11} is

$$\begin{aligned}\frac{1}{b_1b_2 - a_1a_2}\big(&-a_1^2a_2^2b_3 + a_1^2a_2b_3x_2 - a_1b_1a_2^2a_3 + a_1b_1a_2^2x_3\\ &+a_1a_2^2b_3x_1 + a_1b_1a_2b_2a_3 + a_1b_1a_2a_3x_2 - a_1b_1a_2b_3x_2 - a_1b_1b_2a_3x_2\\ &+ b_1^2a_2b_2b_3 - b_1^2a_2b_2x_3 - b_1a_2b_2b_3x_1 - a_1a_2f + b_1b_2f\big) \geq 0.\end{aligned} \tag{3.6}$$

- The facet through points v^3, v^5, v^8 and v_3^{11} is

$$f - a_2b_3x_1 - a_1b_3x_2 - b_1b_2x_3 + a_1a_2b_3 + b_1b_2b_3 \geq 0. \tag{3.7}$$

- The facet through points v^1, v^4, v^7 and v_3^{11} is

$$f - b_2a_3x_1 - b_1a_3x_2 - a_1a_2x_3 + a_1a_2a_3 + b_1b_2a_3 \geq 0. \tag{3.8}$$

- The facet through points v^4, v^5, v^7 and v_3^{11} is 2.32.

- The facet through points v^1, v^7, v^8 and v_3^{11} is

$$\begin{aligned}\frac{1}{b_1b_2 - a_1a_2}\big(&-a_1b_1a_2^2a_3 + a_1b_1a_2^2x_3 + a_1b_1a_2a_3x_2 - a_1b_1a_2b_2b_3 +\\ &a_1a_2b_2b_3x_1 + b_1a_2b_2a_3x_1 + b_1^2a_2b_2b_3 - b_1^2a_2b_2x_3 - b_1a_2b_2b_3x_1 +\\ &b_1^2b_2^2a_3 - b_1^2b_2a_3x_2 - b_1b_2^2a_3x_1 - a_1a_2f + b_1b_2f\big) \geq 0.\end{aligned} \tag{3.9}$$

- The facet through points v^5, v^7, v^8 and v_3^{11} is 2.18.
- The facet through points v^1, v^3, v^4 and v_3^{11} is 2.17.
- The facet through points v^3, v^4, v^5 and v_3^{11} is 2.29.

There are four inequalities that are not already contained in system $\mathcal{T}_3^-$, we add these and in doing so obtain the system of inequalities that describes $\mathcal{T}_3 = \text{conv}(\mathcal{P}_h \cup \{v_3^{11}\})$.

We now compute the additional volume of $\text{conv}(\mathcal{T}_3 \cup \{v_3^{12}\})$ compared to the volume of $\mathcal{T}_3$. As we have done previously, we sum the volumes of $\text{conv}(\{v_3^{12}\} \cup F)$ for each facet, F, of $\mathcal{T}_3$ such that v_3^{12} is beyond that facet. We substitute v_3^{12} into each relevant inequality (i.e., the system of inequalities that describes $\mathcal{T}_3$) and if the result is negative then v_3^{12} lies beyond that facet. It is immediately clear that v_3^{12} satisfies inequalities 2.17-2.18, 2.23-2.25, 2.27-2.34 and 3.7-3.8. We can also see immediately that v_3^{12} violates inequalities 3.6 and 3.9. To see that inequality 2.26 is also satisfied see §3.8.2.3.

Therefore, we see that v_3^{12} is beyond two facets, and we need to compute the volume of the convex hull of v_3^{12} with each of these facets.

The extreme points that lie on the first facet are points v^1, v^3, v^8 and v_3^{11}. The polytope $\text{conv}\{v^1, v^3, v^8, v_3^{11}, v_3^{12}\}$ is a 4-simplex with volume:

$$b_1 a_2 (b_1 - a_1)^2 (b_2 - a_2)^2 (b_3 - a_3)^2 / \left(24(b_1 b_2 - a_1 a_2)\right).$$

The extreme points that lie on the second facet are points v^1, v^7, v^8 and v_3^{11}. The polytope $\text{conv}\{v^1, v^7, v^8, v_3^{11}, v_3^{12}\}$ is a 4-simplex with volume:

$$b_1 a_2 (b_1 - a_1)^2 (b_2 - a_2)^2 (b_3 - a_3)^2 / \left(24(b_1 b_2 - a_1 a_2)\right).$$

We now compute the additional facets; we take the four facets from adding each simplex and delete the facet that repeats. This leaves us with the following six facets to compute:

- The facet through points v^1, v^7, v^8 and v_3^{12} is 2.30.
- The facet through points v^1, v^7, v_3^{11} and v_3^{12} is 3.8.
- The facet through points v^7, v^8, v_3^{11} and v_3^{12} is 2.18.
- The facet through points v^1, v^3, v^8 and v_3^{12} is 2.31.
- The facet through points v^1, v^3, v_3^{11} and v_3^{12} is 2.17.

- The facet through points v^3, v^8, v_3^{11} and v_3^{12} is 3.7.

By adding and deleting the appropriate facets to and from system S_h, we see that we arrive at system S_3 (see §3.4.1 for a more detailed explanation).

Therefore, to compute the volume of $\mathcal{P}_3$, we sum the volume of $\mathcal{P}_h$ with that of the appropriate eight simplices, and we obtain our result. □

3.4.1 Keeping track of facets

Here, we briefly describe the details that confirm the $8+4=12$ extreme points we conjectured at the beginning of the proof are in fact the extreme points of polytope $\mathcal{P}_3$.

In the proof we start with system S_h, initially given in §2.1.2 and repeated here for convenience:

$$f - a_2a_3x_1 - a_1a_3x_2 - a_1a_2x_3 + 2a_1a_2a_3 \geq 0, \quad (2.17)$$

$$f - b_2b_3x_1 - b_1b_3x_2 - b_1b_2x_3 + 2b_1b_2b_3 \geq 0, \quad (2.18)$$

$$f - a_2b_3x_1 - a_1b_3x_2 - b_1a_2x_3 + a_1a_2b_3 + b_1a_2b_3 \geq 0, \quad (2.19)$$

$$f - b_2a_3x_1 - b_1a_3x_2 - a_1b_2x_3 + b_1b_2a_3 + a_1b_2a_3 \geq 0, \quad (2.20)$$

$$f - \tfrac{\eta_1}{b_1-a_1}x_1 - b_1a_3x_2 - b_1a_2x_3 + \left(\tfrac{\eta_1a_1}{b_1-a_1} + b_1b_2a_3 + b_1a_2b_3 - a_1b_2b_3\right) \geq 0, \quad (2.21)$$

$$f - \tfrac{\eta_2}{a_1-b_1}x_1 - a_1b_3x_2 - a_1b_2x_3 + \left(\tfrac{\eta_2b_1}{a_1-b_1} + a_1a_2b_3 + a_1b_2a_3 - b_1a_2a_3\right) \geq 0, \quad (2.22)$$

$$-f + a_2a_3x_1 + b_1a_3x_2 + b_1b_2x_3 - b_1b_2a_3 - b_1a_2a_3 \geq 0, \quad (2.23)$$

$$-f + b_2a_3x_1 + a_1a_3x_2 + b_1b_2x_3 - b_1b_2a_3 - a_1b_2a_3 \geq 0, \quad (2.24)$$

$$-f + a_2a_3x_1 + b_1b_3x_2 + b_1a_2x_3 - b_1a_2b_3 - b_1a_2a_3 \geq 0, \quad (2.25)$$

$$-f + b_2b_3x_1 + a_1a_3x_2 + a_1b_2x_3 - a_1b_2b_3 - a_1b_2a_3 \geq 0, \quad (2.26)$$

$$-f + a_2b_3x_1 + b_1b_3x_2 + a_1a_2x_3 - b_1a_2b_3 - a_1a_2b_3 \geq 0, \quad (2.27)$$

$$-f + b_2b_3x_1 + a_1b_3x_2 + a_1a_2x_3 - a_1b_2b_3 - a_1a_2b_3 \geq 0, \quad (2.28)$$

$$x_1 - a_1 \geq 0, \quad (2.29)$$

$$-x_1 + b_1 \geq 0, \quad (2.30)$$

$$x_2 - a_2 \geq 0, \quad (2.31)$$

$$-x_2 + b_2 \geq 0, \quad (2.32)$$

$$x_3 - a_3 \geq 0, \quad (2.33)$$

$$-x_3 + b_3 \geq 0, \quad (2.34)$$

where $\eta_1 = b_1b_2a_3 - a_1b_2b_3 - b_1a_2a_3 + b_1a_2b_3$ and $\eta_2 = a_1a_2b_3 - b_1a_2a_3 - a_1b_2b_3 + a_1b_2a_3$.

We then add in point v_3^9, and show that in doing this we need to remove the violated facet 2.25, shown below in red. We know that we do not need to add any additional facets because the facets that are generated by adding this point are already contained in system S_h. Namely, facets 2.23, 2.27, 2.30, and 2.31.

$$f - a_2a_3x_1 - a_1a_3x_2 - a_1a_2x_3 + 2a_1a_2a_3 \geq 0, \quad (2.17)$$
$$f - b_2b_3x_1 - b_1b_3x_2 - b_1b_2x_3 + 2b_1b_2b_3 \geq 0, \quad (2.18)$$
$$f - a_2b_3x_1 - a_1b_3x_2 - b_1a_2x_3 + a_1a_2b_3 + b_1a_2b_3 \geq 0, \quad (2.19)$$
$$f - b_2a_3x_1 - b_1a_3x_2 - a_1b_2x_3 + b_1b_2a_3 + a_1b_2a_3 \geq 0, \quad (2.20)$$
$$f - \tfrac{\eta_1}{b_1-a_1}x_1 - b_1a_3x_2 - b_1a_2x_3 + \left(\tfrac{\eta_1a_1}{b_1-a_1} + b_1b_2a_3 + b_1a_2b_3 - a_1b_2b_3\right) \geq 0, \quad (2.21)$$
$$f - \tfrac{\eta_2}{a_1-b_1}x_1 - a_1b_3x_2 - a_1b_2x_3 + \left(\tfrac{\eta_2b_1}{a_1-b_1} + a_1a_2b_3 + a_1b_2a_3 - b_1a_2a_3\right) \geq 0, \quad (2.22)$$
$$-f + a_2a_3x_1 + b_1a_3x_2 + b_1b_2x_3 - b_1b_2a_3 - b_1a_2a_3 \geq 0, \quad (2.23)$$
$$-f + b_2a_3x_1 + a_1a_3x_2 + b_1b_2x_3 - b_1b_2a_3 - a_1b_2a_3 \geq 0, \quad (2.24)$$
$$-f + a_2a_3x_1 + b_1b_3x_2 + b_1a_2x_3 - b_1a_2b_3 - b_1a_2a_3 \geq 0, \quad (2.25)$$
$$-f + b_2b_3x_1 + a_1a_3x_2 + a_1b_2x_3 - a_1b_2b_3 - a_1b_2a_3 \geq 0, \quad (2.26)$$
$$-f + a_2b_3x_1 + b_1b_3x_2 + a_1a_2x_3 - b_1a_2b_3 - a_1a_2b_3 \geq 0, \quad (2.27)$$
$$-f + b_2b_3x_1 + a_1b_3x_2 + a_1a_2x_3 - a_1b_2b_3 - a_1a_2b_3 \geq 0, \quad (2.28)$$
$$x_1 - a_1 \geq 0, \quad (2.29)$$
$$-x_1 + b_1 \geq 0, \quad (2.30)$$
$$x_2 - a_2 \geq 0, \quad (2.31)$$
$$-x_2 + b_2 \geq 0, \quad (2.32)$$
$$x_3 - a_3 \geq 0, \quad (2.33)$$
$$-x_3 + b_3 \geq 0, \quad (2.34)$$

Removing facet 2.25 leaves us with the system of inequalities shown below. We then add point v_3^{10}, and show that in doing this we need to remove the violated facet 2.26, shown below in red. Again, we note that we do not need to add any additional facets because the facets that are generated by adding this point are already contained in the system. Namely, facets 2.24, 2.28, 2.29, and 2.32.

$$f - a_2a_3x_1 - a_1a_3x_2 - a_1a_2x_3 + 2a_1a_2a_3 \geq 0, \quad (2.17)$$

$$f - b_2b_3x_1 - b_1b_3x_2 - b_1b_2x_3 + 2b_1b_2b_3 \geq 0, \quad (2.18)$$

$$f - a_2b_3x_1 - a_1b_3x_2 - b_1a_2x_3 + a_1a_2b_3 + b_1a_2b_3 \geq 0, \quad (2.19)$$

$$f - b_2a_3x_1 - b_1a_3x_2 - a_1b_2x_3 + b_1b_2a_3 + a_1b_2a_3 \geq 0, \quad (2.20)$$

$$f - \tfrac{\eta_1}{b_1-a_1}x_1 - b_1a_3x_2 - b_1a_2x_3 + \left(\tfrac{\eta_1a_1}{b_1-a_1} + b_1b_2a_3 + b_1a_2b_3 - a_1b_2b_3\right) \geq 0, \quad (2.21)$$

$$f - \tfrac{\eta_2}{a_1-b_1}x_1 - a_1b_3x_2 - a_1b_2x_3 + \left(\tfrac{\eta_2b_1}{a_1-b_1} + a_1a_2b_3 + a_1b_2a_3 - b_1a_2a_3\right) \geq 0, \quad (2.22)$$

$$-f + a_2a_3x_1 + b_1a_3x_2 + b_1b_2x_3 - b_1b_2a_3 - b_1a_2a_3 \geq 0, \quad (2.23)$$

$$-f + b_2a_3x_1 + a_1a_3x_2 + b_1b_2x_3 - b_1b_2a_3 - a_1b_2a_3 \geq 0, \quad (2.24)$$

$$-f + b_2b_3x_1 + a_1a_3x_2 + a_1b_2x_3 - a_1b_2b_3 - a_1b_2a_3 \geq 0, \quad (2.26)$$

$$-f + a_2b_3x_1 + b_1b_3x_2 + a_1a_2x_3 - b_1a_2b_3 - a_1a_2b_3 \geq 0, \quad (2.27)$$

$$-f + b_2b_3x_1 + a_1b_3x_2 + a_1a_2x_3 - a_1b_2b_3 - a_1a_2b_3 \geq 0, \quad (2.28)$$

$$x_1 - a_1 \geq 0, \quad (2.29)$$

$$-x_1 + b_1 \geq 0, \quad (2.30)$$

$$x_2 - a_2 \geq 0, \quad (2.31)$$

$$-x_2 + b_2 \geq 0, \quad (2.32)$$

$$x_3 - a_3 \geq 0, \quad (2.33)$$

$$-x_3 + b_3 \geq 0, \quad (2.34)$$

Removing facet 2.26 leaves us with the inequalities shown below. We then add point v_3^{11}, and show that in doing this we need to remove four violated facets. Namely, facets 2.19, 2.20, 2.21 and 2.22, shown below in red. We then generate the additional facets created by adding point v_3^{11}, and we note that four of these eight facets are already contained in the system (namely, facets 2.17, 2.18, 2.29, and 2.32). However, there are four additional facets that are not yet accounted for, facets 3.6, 3.7, 3.8 and 3.9. We add these to the system, shown below in blue.

$$f - a_2a_3x_1 - a_1a_3x_2 - a_1a_2x_3 + 2a_1a_2a_3 \geq 0, \quad (2.17)$$

$$f - b_2b_3x_1 - b_1b_3x_2 - b_1b_2x_3 + 2b_1b_2b_3 \geq 0, \quad (2.18)$$

$$f - a_2b_3x_1 - a_1b_3x_2 - b_1a_2x_3 + a_1a_2b_3 + b_1a_2b_3 \geq 0, \quad (2.19)$$

$$f - b_2a_3x_1 - b_1a_3x_2 - a_1b_2x_3 + b_1b_2a_3 + a_1b_2a_3 \geq 0, \quad (2.20)$$

$$f - \tfrac{\eta_1}{b_1-a_1}x_1 - b_1a_3x_2 - b_1a_2x_3 + \left(\tfrac{\eta_1a_1}{b_1-a_1} + b_1b_2a_3 + b_1a_2b_3 - a_1b_2b_3\right) \geq 0, \quad (2.21)$$

$$f - \tfrac{\eta_2}{a_1-b_1}x_1 - a_1b_3x_2 - a_1b_2x_3 + \left(\tfrac{\eta_2b_1}{a_1-b_1} + a_1a_2b_3 + a_1b_2a_3 - b_1a_2a_3\right) \geq 0, \quad (2.22)$$

$$-f + a_2a_3x_1 + b_1a_3x_2 + b_1b_2x_3 - b_1b_2a_3 - b_1a_2a_3 \geq 0, \quad (2.23)$$

$$-f + b_2a_3x_1 + a_1a_3x_2 + b_1b_2x_3 - b_1b_2a_3 - a_1b_2a_3 \geq 0, \quad (2.24)$$

$$-f + a_2b_3x_1 + b_1b_3x_2 + a_1a_2x_3 - b_1a_2b_3 - a_1a_2b_3 \geq 0, \quad (2.27)$$

$$-f + b_2b_3x_1 + a_1b_3x_2 + a_1a_2x_3 - a_1b_2b_3 - a_1a_2b_3 \geq 0, \quad (2.28)$$

$$\frac{1}{b_1b_2 - a_1a_2}\Big(- a_1^2a_2^2b_3 + a_1^2a_2b_3x_2 - a_1b_1a_2^2a_3 + a_1b_1a_2^2x_3$$
$$+a_1a_2^2b_3x_1 + a_1b_1a_2b_2a_3 + a_1b_1a_2a_3x_2 - a_1b_1a_2b_3x_2 - a_1b_1b_2a_3x_2$$
$$+b_1^2a_2b_2b_3 - b_1^2a_2b_2x_3 - b_1a_2b_2b_3x_1 - a_1a_2f + b_1b_2f\Big) \geq 0, \quad (3.6)$$

$$f - a_2b_3x_1 - a_1b_3x_2 - b_1b_2x_3 + a_1a_2b_3 + b_1b_2b_3 \geq 0, \quad (3.7)$$

$$f - b_2a_3x_1 - b_1a_3x_2 - a_1a_2x_3 + a_1a_2a_3 + b_1b_2a_3 \geq 0, \quad (3.8)$$

$$\frac{1}{b_1b_2 - a_1a_2}\Big(- a_1b_1a_2^2a_3 + a_1b_1a_2^2x_3 + a_1b_1a_2a_3x_2 - a_1b_1a_2b_2b_3 + a_1a_2b_2b_3x_1$$
$$+b_1a_2b_2a_3x_1 + b_1^2a_2b_2b_3 - b_1^2a_2b_2x_3 - b_1a_2b_2b_3x_1$$
$$+b_1^2b_2^2a_3 - b_1^2b_2a_3x_2 - b_1b_2^2a_3x_1 - a_1a_2f + b_1b_2f\Big) \geq 0, \quad (3.9)$$

$$x_1 - a_1 \geq 0, \quad (2.29)$$

$$-x_1 + b_1 \geq 0, \quad (2.30)$$

$$x_2 - a_2 \geq 0, \quad (2.31)$$

$$-x_2 + b_2 \geq 0, \quad (2.32)$$

$$x_3 - a_3 \geq 0, \quad (2.33)$$

$$-x_3 + b_3 \geq 0, \quad (2.34)$$

Adding and removing these facets leaves us with the inequalities shown below. Finally we add in point v_3^{12}, and show that in doing this we need to remove two violated facets (namely facets 3.6 and 3.9), shown below in red. We generate the additional facets created by adding point v_3^{12}, and note that each one is already contained in the system. These are facets 2.17, 2.18, 2.30, 2.31, 3.8, and 3.9.

$$f - a_2a_3x_1 - a_1a_3x_2 - a_1a_2x_3 + 2a_1a_2a_3 \geq 0, \quad (2.17)$$

$$f - b_2b_3x_1 - b_1b_3x_2 - b_1b_2x_3 + 2b_1b_2b_3 \geq 0, \quad (2.18)$$

$$-f + a_2a_3x_1 + b_1a_3x_2 + b_1b_2x_3 - b_1b_2a_3 - b_1a_2a_3 \geq 0, \quad (2.23)$$

$$-f + b_2a_3x_1 + a_1a_3x_2 + b_1b_2x_3 - b_1b_2a_3 - a_1b_2a_3 \geq 0, \quad (2.24)$$

$$-f + a_2b_3x_1 + b_1b_3x_2 + a_1a_2x_3 - b_1a_2b_3 - a_1a_2b_3 \geq 0, \quad (2.27)$$

$$-f + b_2b_3x_1 + a_1b_3x_2 + a_1a_2x_3 - a_1b_2b_3 - a_1a_2b_3 \geq 0, \quad (2.28)$$

$$\frac{1}{b_1b_2-a_1a_2}\Big(-a_1^2a_2^2b_3+a_1^2a_2b_3x_2-a_1b_1a_2^2a_3+a_1b_1a_2^2x_3$$
$$+a_1a_2^2b_3x_1+a_1b_1a_2b_2a_3+a_1b_1a_2a_3x_2-a_1b_1a_2b_3x_2-a_1b_1b_2a_3x_2$$
$$+b_1^2a_2b_2b_3-b_1^2a_2b_2x_3-b_1a_2b_2b_3x_1-a_1a_2f+b_1b_2f\Big)\geq 0, \quad (3.6)$$

$$f-a_2b_3x_1-a_1b_3x_2-b_1b_2x_3+a_1a_2b_3+b_1b_2b_3\geq 0, \quad (3.7)$$

$$f-b_2a_3x_1-b_1a_3x_2-a_1a_2x_3+a_1a_2a_3+b_1b_2a_3\geq 0, \quad (3.8)$$

$$\frac{1}{b_1b_2-a_1a_2}\Big(-a_1b_1a_2^2a_3+a_1b_1a_2^2x_3+a_1b_1a_2a_3x_2-a_1b_1a_2b_2b_3+a_1a_2b_2b_3x_1$$
$$+b_1a_2b_2a_3x_1+b_1^2a_2b_2b_3-b_1^2a_2b_2x_3-b_1a_2b_2b_3x_1$$
$$+b_1^2b_2^2a_3-b_1^2b_2a_3x_2-b_1b_2^2a_3x_1-a_1a_2f+b_1b_2f\Big)\geq 0, \quad (3.9)$$

$$x_1-a_1\geq 0, \quad (2.29)$$

$$-x_1+b_1\geq 0, \quad (2.30)$$

$$x_2-a_2\geq 0, \quad (2.31)$$

$$-x_2+b_2\geq 0, \quad (2.32)$$

$$x_3-a_3\geq 0, \quad (2.33)$$

$$-x_3+b_3\geq 0, \quad (2.34)$$

We are therefore left with the system of inequalities displayed below, and it is easy to check that this is exactly system S_3. From this we know that the twelve extreme points in the statement of Theorem 3.4 are in fact the extreme points of polytope $\mathcal{P}_3$, and that the volume computation is for the correct polytope: $\mathcal{P}_3$.

$$f-a_2a_3x_1-a_1a_3x_2-a_1a_2x_3+2a_1a_2a_3\geq 0, \quad (2.17)$$

$$f-b_2b_3x_1-b_1b_3x_2-b_1b_2x_3+2b_1b_2b_3\geq 0, \quad (2.18)$$

$$-f+a_2a_3x_1+b_1a_3x_2+b_1b_2x_3-b_1b_2a_3-b_1a_2a_3\geq 0, \quad (2.23)$$

$$-f+b_2a_3x_1+a_1a_3x_2+b_1b_2x_3-b_1b_2a_3-a_1b_2a_3\geq 0, \quad (2.24)$$

$$-f+a_2b_3x_1+b_1b_3x_2+a_1a_2x_3-b_1a_2b_3-a_1a_2b_3\geq 0, \quad (2.27)$$

$$-f+b_2b_3x_1+a_1b_3x_2+a_1a_2x_3-a_1b_2b_3-a_1a_2b_3\geq 0, \quad (2.28)$$

$$f-a_2b_3x_1-a_1b_3x_2-b_1b_2x_3+a_1a_2b_3+b_1b_2b_3\geq 0, \quad (3.7)$$

$$f-b_2a_3x_1-b_1a_3x_2-a_1a_2x_3+a_1a_2a_3+b_1b_2a_3\geq 0, \quad (3.8)$$

$$x_1-a_1\geq 0, \quad (2.29)$$

$$-x_1+b_1\geq 0, \quad (2.30)$$

$$x_2 - a_2 \geq 0, \tag{2.31}$$
$$-x_2 + b_2 \geq 0, \tag{2.32}$$
$$x_3 - a_3 \geq 0, \tag{2.33}$$
$$-x_3 + b_3 \geq 0, \tag{2.34}$$

3.5 Proof of Thm. 3.2

To compute the volume of polytope $\mathcal{P}_1$, we compute the volume of the convex hull of the 12 extreme points that we claim are exactly the extreme points of polytope $\mathcal{P}_1$. In computing the volume of this polytope, we also prove that these are the correct extreme points and that we have therefore computed the volume of $\mathcal{P}_1$.

The relevant points are the eight extreme points of $\mathcal{P}_h$, plus an additional four points. Because we have already computed the volume of $\mathcal{P}_h$, to compute the volume of $\mathcal{P}_1$, we need to compute the additional volume, compared with $\mathcal{P}_h$, added by these four extra extreme points. To show that this is indeed the volume of $\mathcal{P}_1$, we keep track of which facets need to be deleted and added to the system of inequalities as we go. When this is complete, we have exactly system S_1 and therefore we must also have the correct extreme points. Similar to in the corresponding section in the proof of Theorem 3.4, we provide more details concerning this part of the proof in §3.5.1.

We begin with system S_h which can be found in §2.1.2, and we use the same principles as we used in the previous proof to compute the volume of $\mathcal{P}_3$.

First, we argue that we can add v_1^9 to $\mathcal{P}_h$ separately, v_1^{10} to $\mathcal{P}_h$ separately and then v_1^{11} and v_1^{12} together. To do this, we show that the midpoint of the line segment between v_1^9 and all other additional points (v_1^{10}, v_1^{11} and v_1^{12}) intersects $\mathcal{P}_h$. We also show this is true for v_1^{10}.

As in the previous proof, we refer to the midpoint of the line between v_i^j and v_i^k as $M(v_i^j, v_i^k)$, and we show that the midpoint of each line satisfies each of the inequalities of $\mathcal{P}_h$ by substituting this point into each inequality and checking the result. See Table 3.2 for a summary of the resulting substitutions. The table notes whether non-negativity of the resulting quantity follows immediately (after factoring), or by using a technical lemma, after further explanation in §3.8.2, or after being rewritten in the way referenced in Figure 3.9. Because we have shown that each resulting quantity is non-negative, we know that the midpoint intersects $\mathcal{P}_h$, and therefore we can add v_1^9 separately, v_1^{10} separately, and then v_1^{11} and v_1^{12} together.

Ineq	$M(v_1^9, v_1^{10})$	$M(v_1^9, v_1^{11})$	$M(v_1^9, v_1^{12})$	$M(v_1^{10}, v_1^{11})$	$M(v_1^{10}, v_1^{12})$
2.17	immediate	immediate	immediate	immediate	immediate
2.18	immediate	immediate	immediate	immediate	immediate
2.19	see §3.8.2.4	see 3.10 and Lemma 3.9	immediate	see 3.11 and Lemma 3.9	see §3.8.2.10
2.20	see §3.8.2.5	see §3.8.2.6	see 3.11 and Lemma 3.9	immediate	see 3.12 and Lemma 3.9
2.21	see §3.8.2.16	see §3.8.2.7	by Lemma 3.9	by Lemma 3.9	see §3.8.2.11
2.22	see §3.8.2.17	see §3.8.2.8	by Lemma 3.9	by Lemma 3.9	see §3.8.2.12
2.23	immediate	immediate	immediate	immediate	immediate
2.24	see 3.13	see 3.14	see 3.15	immediate	see §3.8.2.13
2.25	immediate	immediate	immediate	immediate	immediate
2.26	immediate	immediate	immediate	immediate	immediate
2.27	see 3.16	See §3.8.2.9	immediate	see 3.15	see 3.17
2.28	immediate	immediate	immediate	immediate	immediate
2.29	immediate	immediate	immediate	immediate	immediate
2.30	immediate	immediate	immediate	immediate	immediate
2.31	immediate	immediate	immediate	immediate	immediate
2.32	immediate	immediate	immediate	immediate	immediate
2.33	immediate	immediate	immediate	immediate	immediate
2.34	immediate	immediate	immediate	immediate	immediate

Table 3.2: Summary of midpoint substitutions for Thm. 3.2

$$\frac{(b_3-a_3)(b_2-a_2)\,(b_3(b_1a_2-a_1b_2)+a_2(b_1a_3-a_1b_3))}{2(b_2b_3-a_2a_3)} \tag{3.10}$$

$$\frac{(b_1a_3-a_1b_3)(b_2-a_2)+(b_1a_2-a_1b_2)(b_3-a_3)}{2} \tag{3.11}$$

$$\frac{(b_3-a_3)(b_2-a_2)\,(b_2(b_1a_3-a_1b_3)+a_3(b_1a_2-a_1b_2))}{2(b_2b_3-a_2a_3)} \tag{3.12}$$

$$\frac{(b_3-a_3)(b_2-a_2)\,(b_1b_2(b_3-a_3)+a_1a_3(b_2-a_2))}{2(b_2b_3-a_2a_3)} \tag{3.13}$$

$$\frac{(b_3-a_3)(b_2-a_2)\,(b_2(b_1b_3-a_1a_3)+a_3(b_1b_2-a_1a_2))}{2(b_2b_3-a_2a_3)} \tag{3.14}$$

$$\frac{(b_1b_3-a_1a_3)(b_2-a_2)+(b_1b_2-a_1a_2)(b_3-a_3)}{2} \tag{3.15}$$

$$\frac{(b_3-a_3)(b_2-a_2)\,(b_1b_3(b_2-a_2)+a_1a_2(b_3-a_3))}{2(b_2b_3-a_2a_3)} \tag{3.16}$$

$$\frac{(b_3-a_3)(b_2-a_2)\,(b_3(b_1b_2-a_1a_2)+a_2(b_1b_3-a_1a_3))}{2(b_2b_3-a_2a_3)} \tag{3.17}$$

Figure 3.9: For Table 3.2

Computing the (additional) volume of conv($\mathcal{P}_h \cup \{v_1^9\}$). We now compute the additional volume of conv($\mathcal{P}_h \cup \{v_1^9\}$) compared to the volume of $\mathcal{P}_h$. To do this, we sum the volumes of conv($\{v_1^9\} \cup F$) for each facet, F, of $\mathcal{P}_h$ such that v_1^9 is beyond that facet. We substitute v_1^9 into the 18 relevant inequalities (2.17-2.34) and immediately see that it satisfies 2.17-2.19, 2.23-2.26 and 2.28-2.34. It is also immediate to see that inequality 2.27 is violated. To show that the remaining three inequalities are satisfied (2.20, 2.21 and 2.22) see §3.8.2.14, §3.8.2.18 and §3.8.2.19.

From this, we know that v_1^9 is beyond only one facet. The extreme points that lie on this facet are points v^2, v^3, v^6 and v^8. The polytope conv$\{v^2, v^3, v^6, v^8, v_1^9\}$ is a 4-simplex with volume:

$$a_2 b_3 (b_1 - a_1)^2 (b_2 - a_2)^2 (b_3 - a_3)^2 / \left(24(b_2 b_3 - a_2 a_3)\right).$$

The facets of conv($\mathcal{P}_h \cup \{v_1^9\}$) are the facets of $\mathcal{P}_h$ except inequality 2.27. We see this by computing the four additional facets that come from adding v_1^9 and noting that they are already contained in system S_h:

- The facet through points v^2, v^3, v^6 and v_1^9 is 2.28.
- The facet through points v^2, v^3, v^8 and v_1^9 is 2.31.
- The facet through points v^2, v^6, v^8 and v_1^9 is 2.25.
- The facet through points v^3, v^6, v^8 and v_1^9 is 2.34.

Computing the (additional) volume of conv($\mathcal{P}_h \cup \{v_1^{10}\}$). We now compute the additional volume of conv($\mathcal{P}_h \cup \{v_1^{10}\}$) compared to the volume of $\mathcal{P}_h$. To do this, we sum the volumes of conv($\{v_1^{10}\} \cup F$) for each facet, F, of $\mathcal{P}_h$ such that v_1^{10} is beyond that facet. We substitute v_1^{10} into the 18 relevant inequalities and immediately see that it satisfies 2.17, 2.18, 2.20, 2.23 and 2.25-2.34. It is also immediate to see that inequality 2.24 is violated. To show that the remaining three inequalities are satisfied (2.19, 2.21 and 2.22) see §3.8.2.14, §3.8.2.20 and §3.8.2.21.

From this, we know that v_1^{10} is beyond only one facet. The extreme points that lie on this facet are points v^2, v^4, v^6 and v^7. The polytope conv$\{v^2, v^4, v^6, v^7, v_3^{10}\}$ is a 4-simplex with volume:

$$b_2 a_3 (b_1 - a_1)^2 (b_2 - a_2)^2 (b_3 - a_3)^2 / \left(24(b_2 b_3 - a_2 a_3)\right).$$

The facets of conv($\mathcal{P}_h \cup \{v_1^{10}\}$) are the facets of $\mathcal{P}_h$ except inequality 2.24. We see this by computing the four additional facets that come from adding v_1^{10} and noting that they are already contained in system S_h:

- The facet through points v^2, v^4, v^6 and v_1^{10} is 2.26.
- The facet through points v^2, v^4, v^7 and v_1^{10} is 2.33.
- The facet through points v^2, v^6, v^7 and v_1^{10} is 2.23.
- The facet through points v^4, v^6, v^7 and v_1^{10} is 2.32.

Computing the (additional) volume of conv($\mathcal{P}_h \cup \{v_1^{11}\} \cup \{v_1^{12}\}$). We now compute the additional volume of conv($\mathcal{P}_h \cup \{v_1^{11}\} \cup \{v_1^{12}\}$) compared to the volume of $\mathcal{P}_h$. Because $L(v_1^{11}, v_1^{12})$ lies entirely outside of $\mathcal{P}_1$, we need to add them sequentially.

We first compute the additional volume of conv($\mathcal{P}_h \cup \{v_1^{11}\}$) compared to the volume of $\mathcal{P}_h$. As we have done previously, we sum the volumes of conv($\{v_1^{11}\} \cup F$) for each facet, F, of $\mathcal{P}_h$ such that v_1^{11} is beyond that facet. We substitute v_1^{11} into each relevant inequality and if the result is negative then v_1^{11} lies beyond that facet. It is immediate that v_1^{11} satisfies inequalities 2.17, 2.18, 2.23-2.26 and 2.28-2.34. In §3.8.2.15, we show that 2.27 is also satisfied. It is also immediate that v_1^{11} violates the three facets described by 2.20-2.22. We compute the volume of the convex hulls of v_1^{11} with each of these facets.

The extreme points that lie on the first facet are points v^1, v^4, v^5 and v^7. The polytope conv$\{v^1, v^4, v^5, v^7, v_1^{11}\}$ is a 4-simplex with volume:

$$b_2a_3(b_1 - a_1)^2(b_2 - a_2)^2(b_3 - a_3)^2 / \left(24(b_2b_3 - a_2a_3)\right).$$

The extreme points that lie on the second facet are points v^1, v^5, v^7 and v^8. The polytope conv$\{v^1, v^5, v^7, v^8, v_1^{11}\}$ is a 4-simplex with volume:

$$b_1a_3(b_1 - a_1)(b_2 - a_2)^3(b_3 - a_3)^2 / \left(24(b_2b_3 - a_2a_3)\right).$$

The extreme points that lie on the third facet are points v^1, v^3, v^4 and v^5. The polytope conv$\{v^1, v^3, v^4, v^5, v_1^{11}\}$ is a 4-simplex with volume:

$$a_1b_2(b_1 - a_1)(b_2 - a_2)^2(b_3 - a_3)^3 / \left(24(b_2b_3 - a_2a_3)\right).$$

Unlike in system S_3, we see immediately that there exists a fourth facet (described by 2.19) which, under certain circumstances, v_1^{11} is beyond. In particular, this is true if and only if $a_1b_2b_3 - b_1a_2a_3 > 0$. Therefore, we continue with two cases.

Case 1: $a_1b_2b_3 - b_1a_2a_3 > 0$ In this case there exists a fourth facet (described by 2.19) such that v_1^{11} is beyond this facet. The extreme points that lie on this fourth facet are points v^1, v^3, v^5 and v^8. The polytope $\text{conv}\{v^1, v^3, v^5, v^8, v_1^{11}\}$ is a 4-simplex with volume:

$$(a_1b_2b_3 - b_1a_2a_3)(b_1 - a_1)(b_2 - a_2)^2(b_3 - a_3)^2 / \left(24(b_2b_3 - a_2a_3)\right).$$

We now have a new polytope which is $\text{conv}(\mathcal{P}_h \cup \{v_3^{11}\})$ (in Case 1). We refer to this polytope as $\mathcal{T}_1^1$, and compute the facets of $\mathcal{T}_1^1$.

We begin with the facets of $\mathcal{P}_h$ and delete the four facets that v_1^{11} lies beyond (2.19-2.22). Let us call this system $\mathcal{T}_1^{1-}$. Now consider the four simplices we dealt with when computing the additional volume produced with v_1^{11}. Each of these simplices has 5 facets; one of which corresponds to a deleted facet of $\mathcal{P}_h$.

The remaining 4 facets of the first simplex are described by the planes through the following sets of points: $\{v^1, v^4, v^5, v_1^{11}\}$, $\{v^1, v^4, v^7, v_1^{11}\}$, $\{v^1, v^5, v^7, v_1^{11}\}$ and $\{v^4, v^5, v^7, v_1^{11}\}$.

The remaining 4 facets of the second simplex are described by the planes through the following sets of points: $\{v^1, v^5, v^7, v_1^{11}\}$, $\{v^1, v^5, v^8, v_1^{11}\}$, $\{v^1, v^7, v^8, v_1^{11}\}$ and $\{v^5, v^7, v^8, v_1^{11}\}$.

The remaining 4 facets of the third simplex are described by the planes through the following sets of points: $\{v^1, v^3, v^4, v_1^{11}\}$, $\{v^1, v^3, v^5, v_1^{11}\}$, $\{v^1, v^4, v^5, v_1^{11}\}$ and $\{v^3, v^4, v^5, v_1^{11}\}$.

The remaining 4 facets of the fourth simplex are described by the planes through the following sets of points: $\{v^1, v^3, v^5, v_1^{11}\}$, $\{v^1, v^3, v^8, v_1^{11}\}$, $\{v^1, v^5, v^8, v_1^{11}\}$ and $\{v^3, v^5, v^8, v_1^{11}\}$.

Consider these sixteen facets and exclude the facets that are shared by more than one simplex. This leaves eight facets.

We compute these eight facets to obtain the following:

- The facet through points v^1, v^3, v^8 and v_1^{11} is

$$\begin{aligned}\frac{1}{b_2b_3 - a_2a_3}\big(&-a_1a_2^2a_3b_3 + a_1a_2b_2a_3b_3 + a_1a_2a_3b_3x_2 - a_1b_2a_3b_3x_2 \\ &- b_1a_2^2a_3^2 + b_1a_2^2a_3x_3 + a_2^2a_3b_3x_1 + b_1a_2a_3^2x_2 - b_1a_2a_3b_3x_2 + \\ &b_1a_2b_2b_3^2 - b_1a_2b_2b_3x_3 - a_2b_2b_3^2x_1 - a_2a_3f + b_2b_3f\big) \geq 0.\end{aligned} \tag{3.18}$$

- The facet through points v^3, v^5, v^8 and v_1^{11} is

$$\begin{aligned}\frac{1}{b_2b_3-a_2a_3}\big(&-a_1a_2^2a_3b_3+a_1a_2a_3b_3x_2+a_1a_2b_2b_3x_3+a_1b_2^2b_3^2-\\&a_1b_2^2b_3x_3-a_1b_2b_3^2x_2+a_2^2a_3b_3x_1-b_1a_2b_2a_3b_3+b_1a_2b_2a_3x_3+\\&b_1a_2b_2b_3^2-b_1a_2b_2b_3x_3-a_2b_2b_3^2x_1-a_2a_3f+b_2b_3f\big)\geq 0.\end{aligned} \tag{3.19}$$

- The facet through points v^1, v^4, v^7 and v_1^{11} is 2.33.

- The facet through points v^4, v^5, v^7 and v_1^{11} is 2.32.

- The facet through points v^1, v^7, v^8 and v_1^{11} is

$$f-b_2b_3x_1-b_1a_3x_2-b_1a_2x_3+b_1a_2a_3+b_1b_2b_3\geq 0. \tag{3.20}$$

- The facet through points v^5, v^7, v^8 and v_1^{11} is 2.18.

- The facet through points v^1, v^3, v^4 and v_1^{11} is 2.17.

- The facet through points v^3, v^4, v^5 and v_1^{11} is

$$f-a_2a_3x_1-a_1b_3x_2-a_1b_2x_3+a_1a_2a_3+a_1b_2b_3\geq 0. \tag{3.21}$$

There are four inequalities that are not already contained in system $\mathcal{T}_1^{1-}$, we add these and in doing so obtain the system of inequalities that describes $\mathcal{T}_1^1 = \text{conv}(\mathcal{P}_h \cup \{v_1^{11}\})$ (in Case 1).

We now compute the additional volume of $\text{conv}(\mathcal{T}_1^1 \cup \{v_1^{12}\})$ compared to the volume of $\mathcal{T}_1^1$. As we have done previously, we sum the volumes of $\text{conv}(\{v_1^{12}\} \cup F)$ for each facet, F, of $\mathcal{T}_1^1$ such that v_1^{12} is beyond that facet. We substitute v_1^{12} into each relevant inequality (i.e., the system that describes $\mathcal{T}_1^1$) and if the result is negative then v_1^{12} lies beyond that facet. It is immediately clear that v_1^{12} satisfies inequalities 2.17, 2.18, 2.23, 2.25-2.34 and 3.20-3.21. We also see immediately that inequalities 3.18 and 3.19 are violated. To see that inequality 2.24 is also satisfied see §3.8.2.15.

Therefore, we see that v_1^{12} is beyond two facets, and we need to compute the volume of the convex hull of v_1^{12} with each of these facets.

The extreme points that lie on the first facet are points v^1, v^3, v^8 and v_1^{11}. The polytope $\text{conv}\{v^1, v^3, v^8, v_1^{11}, v_1^{12}\}$ is a 4-simplex with volume:

$$a_2b_3(b_1-a_1)^2(b_2-a_2)^2(b_3-a_3)^2/\left(24(b_2b_3-a_2a_3)\right).$$

The extreme points that lie on the second facet are points v^3, v^5, v^8 and v_1^{11}. The polytope $\text{conv}\{v^3, v^5, v^8, v_1^{11}, v_1^{12}\}$ is a 4-simplex with volume:

$$a_2 b_3 (b_1 - a_1)^2 (b_2 - a_2)^2 (b_3 - a_3)^2 / \left(24(b_2 b_3 - a_2 a_3)\right).$$

We now compute the additional facets; we take the four facets from adding each simplex and delete the facet that repeats. This leaves us with the following six facet defining inequalities to compute:

- The facet through points v^1, v^3, v^8 and v_1^{12} is 2.31.
- The facet through points v^1, v^3, v_1^{11} and v_1^{12} is 2.17.
- The facet through points v^1, v^8, v_1^{11} and v_1^{12} is 3.20.
- The facet through points v^3, v^5, v^8 and v_1^{12} is 2.34.
- The facet through points v^3, v^5, v_1^{11} and v_1^{12} is 3.21.
- The facet through points v^5, v^8, v_1^{11} and v_1^{12} is 2.18.

By adding and deleting the appropriate facets from system S_h, we see that we arrive at system S_1 (see §3.5.1 for a more detailed explanation).

Therefore, to compute the volume of $\mathcal{P}_1$, we sum the volume of $\mathcal{P}_h$ with that of the appropriate eight simplices, and we obtain our result for Case 1.

Case 2: $a_1 b_2 b_3 - b_1 a_2 a_3 \leq 0$ In this case it is immediate to see that v_1^{11} satisfies 2.19 and therefore lies beyond no further facets. This means that we now have a new polytope which is $\text{conv}(\mathcal{P}_h \cup \{v_3^{11}\})$ (in Case 2). We refer to this polytope as $\mathcal{T}_1^2$, and we compute the facets of $\mathcal{T}_1^2$.

We begin with the facets of $\mathcal{P}_h$ and delete the three facets that v_1^{11} lies beyond (2.20-2.22). Let us call this system $\mathcal{T}_1^{2-}$. Now consider the four simplices we dealt with when computing the additional volume produced with v_1^{11}. Each of these simplices has 5 facets; one of which corresponds to a deleted facet of $\mathcal{P}_h$.

The remaining 4 facets of the first simplex are described by the planes through the following sets of points: $\{v^1, v^4, v^5, v_1^{11}\}$, $\{v^1, v^4, v^7, v_1^{11}\}$, $\{v^1, v^5, v^7, v_1^{11}\}$ and $\{v^4, v^5, v^7, v_1^{11}\}$.

The remaining 4 facets of the second simplex are described by the planes through the following sets of points: $\{v^1, v^5, v^7, v_1^{11}\}$, $\{v^1, v^5, v^8, v_1^{11}\}$, $\{v^1, v^7, v^8, v_1^{11}\}$ and $\{v^5, v^7, v^8, v_1^{11}\}$.

The remaining 4 facets of the third simplex are described by the planes through the following sets of points: $\{v^1, v^3, v^4, v_1^{11}\}$, $\{v^1, v^3, v^5, v_1^{11}\}$, $\{v^1, v^4, v^5, v_1^{11}\}$ and $\{v^3, v^4, v^5, v_1^{11}\}$.

Consider these twelve facets and exclude the facets that are shared by more than one simplex. This leaves eight facets.

We compute these eight facets to obtain the following:

- The facet through points v^1, v^4, v^7 and v_1^{11} is 2.33.
- The facet through points v^4, v^5, v^7 and v_1^{11} is 2.32.
- The facet through points v^1, v^5, v^8 and v_1^{11} is

$$\begin{aligned}\frac{1}{b_2(b_1-a_1)}\big(&-a_1b_1a_2^2a_3+a_1b_1a_2a_3x_2+a_1b_1a_2b_2x_3-a_1b_1b_2^2b_3+\\&a_1b_2^2b_3x_1+b_1a_2^2a_3x_1+b_1^2a_2b_2a_3-b_1^2a_2a_3x_2-b_1a_2b_2a_3x_1+\\&b_1^2a_2b_2b_3-b_1^2a_2b_2x_3-b_1a_2b_2b_3x_1-a_1b_2f+b_1b_2f\big)\geq 0.\end{aligned} \tag{3.22}$$

- The facet through points v^1, v^7, v^8 and v_1^{11} is 3.20.
- The facet through points v^5, v^7, v^8 and v_1^{11} is 2.18.
- The facet through points v^1, v^3, v^4 and v_1^{11} is 2.17.
- The facet through points v^1, v^3, v^5 and v_1^{11} is

$$\begin{aligned}\frac{1}{a_3(b_1-a_1)}\big(&-a_1^2a_2a_3b_3-a_1^2b_2a_3b_3+a_1^2a_3b_3x_2+a_1^2b_2b_3x_3+\\&a_1b_1a_2a_3^2+a_1a_2a_3b_3x_1-a_1b_1a_3b_3x_2+a_1b_2a_3b_3x_1+a_1b_1b_2b_3^2-\\&a_1b_1b_2b_3x_3-a_1b_2b_3^2x_1-b_1a_2a_3^2x_1-a_1a_3f+b_1a_3f\big)\geq 0.\end{aligned} \tag{3.23}$$

- The facet through points v^3, v^4, v^5 and v_1^{11} is 3.21.

There are four inequalities that are not already contained in system $\mathcal{T}_1^{2-}$; we add these and in doing this, we obtain the system of inequalities that describes $T_1^2 = \text{conv}(\mathcal{P}_h \cup \{v_1^{11}\})$ (in Case 2).

We now compute the additional volume of $\text{conv}(\mathcal{T}_1^2 \cup \{v_1^{12}\})$ compared to the volume of $\mathcal{T}_1^2$. As we have done previously, we sum the volumes of $\text{conv}(\{v_1^{12}\} \cup F)$ for each facet, F, of $\mathcal{T}_1^2$ such that v_1^{12} is beyond that facet. We substitute v_1^{12} into

each relevant inequality (i.e., the system of inequalities that describes T_1^2 in Case 2) and if the result is negative then v_1^{12} lies beyond that facet. It is immediately clear that v_1^{12} satisfies inequalities 2.17, 2.18, 2.23, 2.25-2.34, 3.20 and 3.21. We also see immediately that v_1^{12} violates inequalities 2.19, 3.22 and 3.23. To see that 2.24 is also satisfied see §3.8.2.15.

Therefore, we know that v_1^{12} is beyond three facets, and we need to compute the volume of the convex hull of v_1^{12} with each of these facets.

The extreme points that lie on the first facet are points v^1, v^3, v^5 and v^8. The polytope $\text{conv}\{v^1, v^3, v^5, v^8, v_1^{12}\}$ is a 4-simplex with volume:

$$a_2 b_3 (b_1 - a_1)^2 (b_2 - a_2)^2 (b_3 - a_3)^2 / \left(24(b_2 b_3 - a_2 a_3)\right).$$

The extreme points that lie on the second facet are points v^1, v^5, v^8 and v_1^{11}. The polytope $\text{conv}\{v^1, v^5, v^8, v_1^{11}, v_1^{12}\}$ is a 4-simplex with volume:

$$(b_1 a_2 (b_1 - a_1)(b_2 - a_2)^2 (b_3 - a_3)^3 / \left(24(b_2 b_3 - a_2 a_3)\right).$$

The extreme points that lie on the third and final facet are points v^1, v^3, v^5 and v_1^{11}. The polytope $\text{conv}\{v^1, v^3, v^5, v_1^{11}, v_1^{12}\}$ is a 4-simplex with volume:

$$a_1 b_3 (b_1 - a_1)(b_2 - a_2)^3 (b_3 - a_3)^2 / \left(24(b_2 b_3 - a_2 a_3)\right).$$

We now compute the additional facets; we take the four facets from adding each simplex and delete the three facets that are repeated. This leaves us with the following six facet defining inequalities to compute:

- The facet through points v^1, v^3, v^8 and v_1^{12} is 2.31.
- The facet through points v^3, v^5, v^8 and v_1^{12} is 2.34.
- The facet through points v^1, v^8, v_1^{11} and v_1^{12} is 3.20.
- The facet through points v^5, v^8, v_1^{11} and v_1^{12} is 2.18.
- The facet through points v^1, v^3, v_1^{11} and v_1^{12} is 2.17.
- The facet through points v^3, v^5, v_1^{11} and v_1^{12} is 3.21.

By adding and deleting the appropriate facets from system S_h, we see that we also arrive at system S_1 in Case 2 (see §3.5.1 for a more detailed explanation).

Therefore, to compute the volume of $\mathcal{P}_1$, we sum the volume of $\mathcal{P}_h$ with that of the appropriate eight simplices, and we obtain our result for Case 2. □

3.5.1 Keeping track of facets

As we did at the end of the proof of Theorem 3.4, here, we briefly describe the details that confirm the $8 + 4 = 12$ extreme points we conjectured at the beginning of the proof are in fact the extreme points of polytope $\mathcal{P}_1$.

In the proof we start with system S_h, initially given in §2.1.2 and repeated here for convenience:

$$f - a_2a_3x_1 - a_1a_3x_2 - a_1a_2x_3 + 2a_1a_2a_3 \geq 0, \quad (2.17)$$

$$f - b_2b_3x_1 - b_1b_3x_2 - b_1b_2x_3 + 2b_1b_2b_3 \geq 0, \quad (2.18)$$

$$f - a_2b_3x_1 - a_1b_3x_2 - b_1a_2x_3 + a_1a_2b_3 + b_1a_2b_3 \geq 0, \quad (2.19)$$

$$f - b_2a_3x_1 - b_1a_3x_2 - a_1b_2x_3 + b_1b_2a_3 + a_1b_2a_3 \geq 0, \quad (2.20)$$

$$f - \tfrac{\eta_1}{b_1-a_1}x_1 - b_1a_3x_2 - b_1a_2x_3 + \left(\tfrac{\eta_1a_1}{b_1-a_1} + b_1b_2a_3 + b_1a_2b_3 - a_1b_2b_3\right) \geq 0, \quad (2.21)$$

$$f - \tfrac{\eta_2}{a_1-b_1}x_1 - a_1b_3x_2 - a_1b_2x_3 + \left(\tfrac{\eta_2b_1}{a_1-b_1} + a_1a_2b_3 + a_1b_2a_3 - b_1a_2a_3\right) \geq 0, \quad (2.22)$$

$$-f + a_2a_3x_1 + b_1a_3x_2 + b_1b_2x_3 - b_1b_2a_3 - b_1a_2a_3 \geq 0, \quad (2.23)$$

$$-f + b_2a_3x_1 + a_1a_3x_2 + b_1b_2x_3 - b_1b_2a_3 - a_1b_2a_3 \geq 0, \quad (2.24)$$

$$-f + a_2a_3x_1 + b_1b_3x_2 + b_1a_2x_3 - b_1a_2b_3 - b_1a_2a_3 \geq 0, \quad (2.25)$$

$$-f + b_2b_3x_1 + a_1a_3x_2 + a_1b_2x_3 - a_1b_2b_3 - a_1b_2a_3 \geq 0, \quad (2.26)$$

$$-f + a_2b_3x_1 + b_1b_3x_2 + a_1a_2x_3 - b_1a_2b_3 - a_1a_2b_3 \geq 0, \quad (2.27)$$

$$-f + b_2b_3x_1 + a_1b_3x_2 + a_1a_2x_3 - a_1b_2b_3 - a_1a_2b_3 \geq 0, \quad (2.28)$$

$$x_1 - a_1 \geq 0, \quad (2.29)$$

$$-x_1 + b_1 \geq 0, \quad (2.30)$$

$$x_2 - a_2 \geq 0, \quad (2.31)$$

$$-x_2 + b_2 \geq 0, \quad (2.32)$$

$$x_3 - a_3 \geq 0, \quad (2.33)$$

$$-x_3 + b_3 \geq 0, \quad (2.34)$$

where $\eta_1 = b_1b_2a_3 - a_1b_2b_3 - b_1a_2a_3 + b_1a_2b_3$ and $\eta_2 = a_1a_2b_3 - b_1a_2a_3 - a_1b_2b_3 + a_1b_2a_3$.

We then add in point v_1^9, and show that in doing this we need to remove the violated facet 2.27, shown below in red. We know that we do not need to add any additional facets because the facets that are generated by adding this point are already

contained in system S_h. Namely, facets 2.25, 2.28, 2.31, and 2.34.

$$f - a_2a_3x_1 - a_1a_3x_2 - a_1a_2x_3 + 2a_1a_2a_3 \geq 0, \quad (2.17)$$
$$f - b_2b_3x_1 - b_1b_3x_2 - b_1b_2x_3 + 2b_1b_2b_3 \geq 0, \quad (2.18)$$
$$f - a_2b_3x_1 - a_1b_3x_2 - b_1a_2x_3 + a_1a_2b_3 + b_1a_2b_3 \geq 0, \quad (2.19)$$
$$f - b_2a_3x_1 - b_1a_3x_2 - a_1b_2x_3 + b_1b_2a_3 + a_1b_2a_3 \geq 0, \quad (2.20)$$
$$f - \tfrac{\eta_1}{b_1-a_1}x_1 - b_1a_3x_2 - b_1a_2x_3 + \left(\tfrac{\eta_1a_1}{b_1-a_1} + b_1b_2a_3 + b_1a_2b_3 - a_1b_2b_3\right) \geq 0, \quad (2.21)$$
$$f - \tfrac{\eta_2}{a_1-b_1}x_1 - a_1b_3x_2 - a_1b_2x_3 + \left(\tfrac{\eta_2b_1}{a_1-b_1} + a_1a_2b_3 + a_1b_2a_3 - b_1a_2a_3\right) \geq 0, \quad (2.22)$$
$$-f + a_2a_3x_1 + b_1a_3x_2 + b_1b_2x_3 - b_1b_2a_3 - b_1a_2a_3 \geq 0, \quad (2.23)$$
$$-f + b_2a_3x_1 + a_1a_3x_2 + b_1b_2x_3 - b_1b_2a_3 - a_1b_2a_3 \geq 0, \quad (2.24)$$
$$-f + a_2a_3x_1 + b_1b_3x_2 + b_1a_2x_3 - b_1a_2b_3 - b_1a_2a_3 \geq 0, \quad (2.25)$$
$$-f + b_2b_3x_1 + a_1a_3x_2 + a_1b_2x_3 - a_1b_2b_3 - a_1b_2a_3 \geq 0, \quad (2.26)$$
$$-f + a_2b_3x_1 + b_1b_3x_2 + a_1a_2x_3 - b_1a_2b_3 - a_1a_2b_3 \geq 0, \quad (2.27)$$
$$-f + b_2b_3x_1 + a_1b_3x_2 + a_1a_2x_3 - a_1b_2b_3 - a_1a_2b_3 \geq 0, \quad (2.28)$$
$$x_1 - a_1 \geq 0, \quad (2.29)$$
$$-x_1 + b_1 \geq 0, \quad (2.30)$$
$$x_2 - a_2 \geq 0, \quad (2.31)$$
$$-x_2 + b_2 \geq 0, \quad (2.32)$$
$$x_3 - a_3 \geq 0, \quad (2.33)$$
$$-x_3 + b_3 \geq 0, \quad (2.34)$$

Removing facet 2.27 leaves us with the system of inequalities shown below. We then add point v_1^{10}, and show that in doing this we need to remove the violated facet 2.24, shown below in red. Again, we note that we do not need to add any additional facets because the facets that are generated by adding this point are already contained in the system. Namely, facets 2.23, 2.26, 2.32, and 2.33.

$$f - a_2a_3x_1 - a_1a_3x_2 - a_1a_2x_3 + 2a_1a_2a_3 \geq 0, \quad (2.17)$$
$$f - b_2b_3x_1 - b_1b_3x_2 - b_1b_2x_3 + 2b_1b_2b_3 \geq 0, \quad (2.18)$$
$$f - a_2b_3x_1 - a_1b_3x_2 - b_1a_2x_3 + a_1a_2b_3 + b_1a_2b_3 \geq 0, \quad (2.19)$$
$$f - b_2a_3x_1 - b_1a_3x_2 - a_1b_2x_3 + b_1b_2a_3 + a_1b_2a_3 \geq 0, \quad (2.20)$$
$$f - \tfrac{\eta_1}{b_1-a_1}x_1 - b_1a_3x_2 - b_1a_2x_3 + \left(\tfrac{\eta_1a_1}{b_1-a_1} + b_1b_2a_3 + b_1a_2b_3 - a_1b_2b_3\right) \geq 0, \quad (2.21)$$

$$f - \tfrac{\eta_2}{a_1-b_1}x_1 - a_1b_3x_2 - a_1b_2x_3 + \left(\tfrac{\eta_2 b_1}{a_1-b_1} + a_1a_2b_3 + a_1b_2a_3 - b_1a_2a_3\right) \geq 0, \quad (2.22)$$

$$-f + a_2a_3x_1 + b_1a_3x_2 + b_1b_2x_3 - b_1b_2a_3 - b_1a_2a_3 \geq 0, \quad (2.23)$$

$$-f + b_2a_3x_1 + a_1a_3x_2 + b_1b_2x_3 - b_1b_2a_3 - a_1b_2a_3 \geq 0, \quad (2.24)$$

$$-f + a_2a_3x_1 + b_1b_3x_2 + b_1a_2x_3 - b_1a_2b_3 - b_1a_2a_3 \geq 0, \quad (2.25)$$

$$-f + b_2b_3x_1 + a_1a_3x_2 + a_1b_2x_3 - a_1b_2b_3 - a_1b_2a_3 \geq 0, \quad (2.26)$$

$$-f + b_2b_3x_1 + a_1b_3x_2 + a_1a_2x_3 - a_1b_2b_3 - a_1a_2b_3 \geq 0, \quad (2.28)$$

$$x_1 - a_1 \geq 0, \quad (2.29)$$

$$-x_1 + b_1 \geq 0, \quad (2.30)$$

$$x_2 - a_2 \geq 0, \quad (2.31)$$

$$-x_2 + b_2 \geq 0, \quad (2.32)$$

$$x_3 - a_3 \geq 0, \quad (2.33)$$

$$-x_3 + b_3 \geq 0, \quad (2.34)$$

Removing facet 2.24 leaves us with the inequalities shown below. We then add point v_1^{11}, and at this stage we need to split our analysis into two separate cases, however, as we will see, both cases ultimately result in the same system of facets.

Case 1: $a_1b_2b_3 - b_1a_2a_3 > 0$ Given that we are in Case 1, adding point v_1^{11} requires that we remove four violated facets. Namely, facets 2.19, 2.20, 2.21 and 2.22, shown below in red. We then generate the additional facets created by adding point v_1^{11}, and we note that four of these eight facets are already contained in the system (namely, facets 2.17, 2.18, 2.32, and 2.33). However, there are four additional facets that are not yet accounted for, facets 3.18, 3.19, 3.20 and 3.21. We add these to the system, shown below in blue.

$$f - a_2a_3x_1 - a_1a_3x_2 - a_1a_2x_3 + 2a_1a_2a_3 \geq 0, \quad (2.17)$$

$$f - b_2b_3x_1 - b_1b_3x_2 - b_1b_2x_3 + 2b_1b_2b_3 \geq 0, \quad (2.18)$$

$$f - a_2b_3x_1 - a_1b_3x_2 - b_1a_2x_3 + a_1a_2b_3 + b_1a_2b_3 \geq 0, \quad (2.19)$$

$$f - b_2a_3x_1 - b_1a_3x_2 - a_1b_2x_3 + b_1b_2a_3 + a_1b_2a_3 \geq 0, \quad (2.20)$$

$$f - \tfrac{\eta_1}{b_1-a_1}x_1 - b_1a_3x_2 - b_1a_2x_3 + \left(\tfrac{\eta_1 a_1}{b_1-a_1} + b_1b_2a_3 + b_1a_2b_3 - a_1b_2b_3\right) \geq 0, \quad (2.21)$$

$$f - \tfrac{\eta_2}{a_1-b_1}x_1 - a_1b_3x_2 - a_1b_2x_3 + \left(\tfrac{\eta_2 b_1}{a_1-b_1} + a_1a_2b_3 + a_1b_2a_3 - b_1a_2a_3\right) \geq 0, \quad (2.22)$$

$$-f + a_2a_3x_1 + b_1a_3x_2 + b_1b_2x_3 - b_1b_2a_3 - b_1a_2a_3 \geq 0, \quad (2.23)$$

$$-f + a_2a_3x_1 + b_1b_3x_2 + b_1a_2x_3 - b_1a_2b_3 - b_1a_2a_3 \geq 0, \quad (2.25)$$

$$-f + b_2b_3x_1 + a_1a_3x_2 + a_1b_2x_3 - a_1b_2b_3 - a_1b_2a_3 \geq 0, \tag{2.26}$$

$$-f + b_2b_3x_1 + a_1b_3x_2 + a_1a_2x_3 - a_1b_2b_3 - a_1a_2b_3 \geq 0, \tag{2.28}$$

$$x_1 - a_1 \geq 0, \tag{2.29}$$

$$-x_1 + b_1 \geq 0, \tag{2.30}$$

$$x_2 - a_2 \geq 0, \tag{2.31}$$

$$-x_2 + b_2 \geq 0, \tag{2.32}$$

$$x_3 - a_3 \geq 0, \tag{2.33}$$

$$-x_3 + b_3 \geq 0, \tag{2.34}$$

$$\begin{aligned}\frac{1}{b_2b_3 - a_2a_3}\Big(&- a_1a_2^2a_3b_3 + a_1a_2b_2a_3b_3 + a_1a_2a_3b_3x_2 - a_1b_2a_3b_3x_2 \\ &-b_1a_2^2a_3^2 + b_1a_2^2a_3x_3 + a_2^2a_3b_3x_1 + b_1a_2a_3^2x_2 - b_1a_2a_3b_3x_2 \\ &+b_1a_2b_2b_3^2 - b_1a_2b_2b_3x_3 - a_2b_2b_3^2x_1 - a_2a_3f + b_2b_3f\Big) \geq 0,\end{aligned} \tag{3.18}$$

$$\begin{aligned}\frac{1}{b_2b_3 - a_2a_3}\Big(&- a_1a_2^2a_3b_3 + a_1a_2a_3b_3x_2 + a_1a_2b_2b_3x_3 + a_1b_2^2b_3^2 \\ &-a_1b_2^2b_3x_3 - a_1b_2b_3^2x_2 + a_2^2a_3b_3x_1 - b_1a_2b_2a_3b_3 + b_1a_2b_2a_3x_3 \\ &+b_1a_2b_2b_3^2 - b_1a_2b_2b_3x_3 - a_2b_2b_3^2x_1 - a_2a_3f + b_2b_3f\Big) \geq 0,\end{aligned} \tag{3.19}$$

$$f - b_2b_3x_1 - b_1a_3x_2 - b_1a_2x_3 + b_1a_2a_3 + b_1b_2b_3 \geq 0, \tag{3.20}$$

$$f - a_2a_3x_1 - a_1b_3x_2 - a_1b_2x_3 + a_1a_2a_3 + a_1b_2b_3 \geq 0, \tag{3.21}$$

Adding and removing these facets leaves us with the inequalities shown below. Finally, we add in point v_1^{12}, and show that in doing this we need to remove two violated facets (namely facets 3.18 and 3.19), shown below in red. We generate the additional facets created by adding point v_1^{12}, and note that each one is already contained in the system. These are facets 2.17, 2.18, 2.31, 2.34, 3.20, and 3.21.

$$f - a_2a_3x_1 - a_1a_3x_2 - a_1a_2x_3 + 2a_1a_2a_3 \geq 0, \tag{2.17}$$

$$f - b_2b_3x_1 - b_1b_3x_2 - b_1b_2x_3 + 2b_1b_2b_3 \geq 0, \tag{2.18}$$

$$-f + a_2a_3x_1 + b_1a_3x_2 + b_1b_2x_3 - b_1b_2a_3 - b_1a_2a_3 \geq 0, \tag{2.23}$$

$$-f + a_2a_3x_1 + b_1b_3x_2 + b_1a_2x_3 - b_1a_2b_3 - b_1a_2a_3 \geq 0, \tag{2.25}$$

$$-f + b_2b_3x_1 + a_1a_3x_2 + a_1b_2x_3 - a_1b_2b_3 - a_1b_2a_3 \geq 0, \tag{2.26}$$

$$-f + b_2b_3x_1 + a_1b_3x_2 + a_1a_2x_3 - a_1b_2b_3 - a_1a_2b_3 \geq 0, \tag{2.28}$$

$$x_1 - a_1 \geq 0, \tag{2.29}$$
$$-x_1 + b_1 \geq 0, \tag{2.30}$$
$$x_2 - a_2 \geq 0, \tag{2.31}$$
$$-x_2 + b_2 \geq 0, \tag{2.32}$$
$$x_3 - a_3 \geq 0, \tag{2.33}$$
$$-x_3 + b_3 \geq 0, \tag{2.34}$$
$$\begin{aligned}\frac{1}{b_2b_3 - a_2a_3}\Big(&- a_1a_2^2a_3b_3 + a_1a_2b_2a_3b_3 + a_1a_2a_3b_3x_2 - a_1b_2a_3b_3x_2 \\ &-b_1a_2^2a_3^2 + b_1a_2^2a_3x_3 + a_2^2a_3b_3x_1 + b_1a_2a_3^2x_2 - b_1a_2a_3b_3x_2 \\ &+b_1a_2b_2b_3^2 - b_1a_2b_2b_3x_3 - a_2b_2b_3^2x_1 - a_2a_3f + b_2b_3f\Big) \geq 0,\end{aligned} \tag{3.18}$$
$$\begin{aligned}\frac{1}{b_2b_3 - a_2a_3}\Big(&- a_1a_2^2a_3b_3 + a_1a_2a_3b_3x_2 + a_1a_2b_2b_3x_3 + a_1b_2^2b_3^2 \\ &-a_1b_2^2b_3x_3 - a_1b_2b_3^2x_2 + a_2^2a_3b_3x_1 - b_1a_2b_2a_3b_3 + b_1a_2b_2a_3x_3 \\ &+b_1a_2b_2b_3^2 - b_1a_2b_2b_3x_3 - a_2b_2b_3^2x_1 - a_2a_3f + b_2b_3f\Big) \geq 0,\end{aligned} \tag{3.19}$$
$$f - b_2b_3x_1 - b_1a_3x_2 - b_1a_2x_3 + b_1a_2a_3 + b_1b_2b_3 \geq 0, \tag{3.20}$$
$$f - a_2a_3x_1 - a_1b_3x_2 - a_1b_2x_3 + a_1a_2a_3 + a_1b_2b_3 \geq 0, \tag{3.21}$$

We are therefore left with the system of inequalities displayed below and it is easy to check that this is exactly system S_1.

$$f - a_2a_3x_1 - a_1a_3x_2 - a_1a_2x_3 + 2a_1a_2a_3 \geq 0, \tag{2.17}$$
$$f - b_2b_3x_1 - b_1b_3x_2 - b_1b_2x_3 + 2b_1b_2b_3 \geq 0, \tag{2.18}$$
$$-f + a_2a_3x_1 + b_1a_3x_2 + b_1b_2x_3 - b_1b_2a_3 - b_1a_2a_3 \geq 0, \tag{2.23}$$
$$-f + a_2a_3x_1 + b_1b_3x_2 + b_1a_2x_3 - b_1a_2b_3 - b_1a_2a_3 \geq 0, \tag{2.25}$$
$$-f + b_2b_3x_1 + a_1a_3x_2 + a_1b_2x_3 - a_1b_2b_3 - a_1b_2a_3 \geq 0, \tag{2.26}$$
$$-f + b_2b_3x_1 + a_1b_3x_2 + a_1a_2x_3 - a_1b_2b_3 - a_1a_2b_3 \geq 0, \tag{2.28}$$
$$x_1 - a_1 \geq 0, \tag{2.29}$$
$$-x_1 + b_1 \geq 0, \tag{2.30}$$
$$x_2 - a_2 \geq 0, \tag{2.31}$$
$$-x_2 + b_2 \geq 0, \tag{2.32}$$
$$x_3 - a_3 \geq 0, \tag{2.33}$$

$$-x_3 + b_3 \geq 0, \tag{2.34}$$
$$f - b_2b_3x_1 - b_1a_3x_2 - b_1a_2x_3 + b_1a_2a_3 + b_1b_2b_3 \geq 0, \tag{3.20}$$
$$f - a_2a_3x_1 - a_1b_3x_2 - a_1b_2x_3 + a_1a_2a_3 + a_1b_2b_3 \geq 0, \tag{3.21}$$

Case 2: $a_1b_2b_3 - b_1a_2a_3 \leq 0$ Given that we are in Case 2, adding point v_1^{11} requires that we remove three violated facets. Namely, facets 2.20, 2.21 and 2.22, shown below in red. We then generate the additional facets created by adding point v_1^{11}, and we note that four of these eight facets are already contained in the system (namely, facets 2.17, 2.18, 2.32, and 2.33). However, there are four additional facets that are not yet accounted for, facets 3.20, 3.21, 3.22 and 3.23. We add these to the system, shown below in blue.

$$f - a_2a_3x_1 - a_1a_3x_2 - a_1a_2x_3 + 2a_1a_2a_3 \geq 0, \tag{2.17}$$
$$f - b_2b_3x_1 - b_1b_3x_2 - b_1b_2x_3 + 2b_1b_2b_3 \geq 0, \tag{2.18}$$
$$f - a_2b_3x_1 - a_1b_3x_2 - b_1a_2x_3 + a_1a_2b_3 + b_1a_2b_3 \geq 0, \tag{2.19}$$
$$f - b_2a_3x_1 - b_1a_3x_2 - a_1b_2x_3 + b_1b_2a_3 + a_1b_2a_3 \geq 0, \tag{2.20}$$
$$f - \tfrac{\eta_1}{b_1-a_1}x_1 - b_1a_3x_2 - b_1a_2x_3 + \left(\tfrac{\eta_1a_1}{b_1-a_1} + b_1b_2a_3 + b_1a_2b_3 - a_1b_2b_3\right) \geq 0, \tag{2.21}$$
$$f - \tfrac{\eta_2}{a_1-b_1}x_1 - a_1b_3x_2 - a_1b_2x_3 + \left(\tfrac{\eta_2b_1}{a_1-b_1} + a_1a_2b_3 + a_1b_2a_3 - b_1a_2a_3\right) \geq 0, \tag{2.22}$$
$$-f + a_2a_3x_1 + b_1a_3x_2 + b_1b_2x_3 - b_1b_2a_3 - b_1a_2a_3 \geq 0, \tag{2.23}$$
$$-f + a_2a_3x_1 + b_1b_3x_2 + b_1a_2x_3 - b_1a_2b_3 - b_1a_2a_3 \geq 0, \tag{2.25}$$
$$-f + b_2b_3x_1 + a_1a_3x_2 + a_1b_2x_3 - a_1b_2b_3 - a_1b_2a_3 \geq 0, \tag{2.26}$$
$$-f + b_2b_3x_1 + a_1b_3x_2 + a_1a_2x_3 - a_1b_2b_3 - a_1a_2b_3 \geq 0, \tag{2.28}$$
$$x_1 - a_1 \geq 0, \tag{2.29}$$
$$-x_1 + b_1 \geq 0, \tag{2.30}$$
$$x_2 - a_2 \geq 0, \tag{2.31}$$
$$-x_2 + b_2 \geq 0, \tag{2.32}$$
$$x_3 - a_3 \geq 0, \tag{2.33}$$
$$-x_3 + b_3 \geq 0, \tag{2.34}$$
$$f - b_2b_3x_1 - b_1a_3x_2 - b_1a_2x_3 + b_1a_2a_3 + b_1b_2b_3 \geq 0, \tag{3.20}$$
$$f - a_2a_3x_1 - a_1b_3x_2 - a_1b_2x_3 + a_1a_2a_3 + a_1b_2b_3 \geq 0, \tag{3.21}$$

$$\frac{1}{b_2(b_1-a_1)}\Big(-a_1b_1a_2^2a_3+a_1b_1a_2a_3x_2+a_1b_1a_2b_2x_3-a_1b_1b_2^2b_3$$
$$+a_1b_2^2b_3x_1+b_1a_2^2a_3x_1+b_1^2a_2b_2a_3-b_1^2a_2a_3x_2-b_1a_2b_2a_3x_1$$
$$+b_1^2a_2b_2b_3-b_1^2a_2b_2x_3-b_1a_2b_2b_3x_1-a_1b_2f+b_1b_2f\Big)\geq 0, \quad (3.22)$$
$$\frac{1}{a_3(b_1-a_1)}\Big(-a_1^2a_2a_3b_3-a_1^2b_2a_3b_3+a_1^2a_3b_3x_2+a_1^2b_2b_3x_3$$
$$+a_1b_1a_2a_3^2+a_1a_2a_3b_3x_1-a_1b_1a_3b_3x_2+a_1b_2a_3b_3x_1+a_1b_1b_2b_3^2$$
$$-a_1b_1b_2b_3x_3-a_1b_2b_3^2x_1-b_1a_2a_3^2x_1-a_1a_3f+b_1a_3f\Big)\geq 0, \quad (3.23)$$

Adding and removing these facets leaves us with the inequalities shown below. Finally, we add in point v_1^{12}, and show that in doing this we need to remove three violated facets (namely facets 2.19, 3.22 and 3.23), shown below in red. We generate the additional facets created by adding point v_1^{12}, and note that each one is already contained in the system. These are facets 2.17, 2.18, 2.31, 2.34, 3.20, and 3.21.

$$f-a_2a_3x_1-a_1a_3x_2-a_1a_2x_3+2a_1a_2a_3\geq 0, \quad (2.17)$$
$$f-b_2b_3x_1-b_1b_3x_2-b_1b_2x_3+2b_1b_2b_3\geq 0, \quad (2.18)$$
$$f-a_2b_3x_1-a_1b_3x_2-b_1a_2x_3+a_1a_2b_3+b_1a_2b_3\geq 0, \quad (2.19)$$
$$-f+a_2a_3x_1+b_1a_3x_2+b_1b_2x_3-b_1b_2a_3-b_1a_2a_3\geq 0, \quad (2.23)$$
$$-f+a_2a_3x_1+b_1b_3x_2+b_1a_2x_3-b_1a_2b_3-b_1a_2a_3\geq 0, \quad (2.25)$$
$$-f+b_2b_3x_1+a_1a_3x_2+a_1b_2x_3-a_1b_2b_3-a_1b_2a_3\geq 0, \quad (2.26)$$
$$-f+b_2b_3x_1+a_1b_3x_2+a_1a_2x_3-a_1b_2b_3-a_1a_2b_3\geq 0, \quad (2.28)$$
$$x_1-a_1\geq 0, \quad (2.29)$$
$$-x_1+b_1\geq 0, \quad (2.30)$$
$$x_2-a_2\geq 0, \quad (2.31)$$
$$-x_2+b_2\geq 0, \quad (2.32)$$
$$x_3-a_3\geq 0, \quad (2.33)$$
$$-x_3+b_3\geq 0, \quad (2.34)$$
$$f-b_2b_3x_1-b_1a_3x_2-b_1a_2x_3+b_1a_2a_3+b_1b_2b_3\geq 0, \quad (3.20)$$
$$f-a_2a_3x_1-a_1b_3x_2-a_1b_2x_3+a_1a_2a_3+a_1b_2b_3\geq 0, \quad (3.21)$$

$$\frac{1}{b_2(b_1-a_1)}\Big(-a_1b_1a_2^2a_3+a_1b_1a_2a_3x_2+a_1b_1a_2b_2x_3-a_1b_1b_2^2b_3$$
$$+a_1b_2^2b_3x_1+b_1a_2^2a_3x_1+b_1^2a_2b_2a_3-b_1^2a_2a_3x_2-b_1a_2b_2a_3x_1$$
$$+b_1^2a_2b_2b_3-b_1^2a_2b_2x_3-b_1a_2b_2b_3x_1-a_1b_2f+b_1b_2f\Big)\geq 0, \tag{3.22}$$
$$\frac{1}{a_3(b_1-a_1)}\Big(-a_1^2a_2a_3b_3-a_1^2b_2a_3b_3+a_1^2a_3b_3x_2+a_1^2b_2b_3x_3$$
$$+a_1b_1a_2a_3^2+a_1a_2a_3b_3x_1-a_1b_1a_3b_3x_2+a_1b_2a_3b_3x_1+a_1b_1b_2b_3^2$$
$$-a_1b_1b_2b_3x_3-a_1b_2b_3^2x_1-b_1a_2a_3^2x_1-a_1a_3f+b_1a_3f\Big)\geq 0, \tag{3.23}$$

We are therefore left with the system of inequalities displayed below and it is easy to check that this is exactly system S_1.

$$f-a_2a_3x_1-a_1a_3x_2-a_1a_2x_3+2a_1a_2a_3\geq 0, \tag{2.17}$$
$$f-b_2b_3x_1-b_1b_3x_2-b_1b_2x_3+2b_1b_2b_3\geq 0, \tag{2.18}$$
$$-f+a_2a_3x_1+b_1a_3x_2+b_1b_2x_3-b_1b_2a_3-b_1a_2a_3\geq 0, \tag{2.23}$$
$$-f+a_2a_3x_1+b_1b_3x_2+b_1a_2x_3-b_1a_2b_3-b_1a_2a_3\geq 0, \tag{2.25}$$
$$-f+b_2b_3x_1+a_1a_3x_2+a_1b_2x_3-a_1b_2b_3-a_1b_2a_3\geq 0, \tag{2.26}$$
$$-f+b_2b_3x_1+a_1b_3x_2+a_1a_2x_3-a_1b_2b_3-a_1a_2b_3\geq 0, \tag{2.28}$$
$$x_1-a_1\geq 0, \tag{2.29}$$
$$-x_1+b_1\geq 0, \tag{2.30}$$
$$x_2-a_2\geq 0, \tag{2.31}$$
$$-x_2+b_2\geq 0, \tag{2.32}$$
$$x_3-a_3\geq 0, \tag{2.33}$$
$$-x_3+b_3\geq 0, \tag{2.34}$$
$$f-b_2b_3x_1-b_1a_3x_2-b_1a_2x_3+b_1a_2a_3+b_1b_2b_3\geq 0, \tag{3.20}$$
$$f-a_2a_3x_1-a_1b_3x_2-a_1b_2x_3+a_1a_2a_3+a_1b_2b_3\geq 0, \tag{3.21}$$

Therefore, in both cases we know that the twelve extreme points in the statement of Theorem 3.2 are in fact the extreme points of polytope $\mathcal{P}_1$, and the volume computation is for the correct polytope: $\mathcal{P}_1$.

3.6 Proof of Thm. 3.3

A mapping from the proof of Theorem 3.4 allows us to claim Theorem 3.3 immediately. □

3.7 Concluding remarks and future work

Our results geometrically quantify the tradeoff between different convexifications of trilinear monomials. Of course it would be nice to use our results to develop guidelines for attacking trilinear monomials within an sBB code. In doing so, it should prove important to develop guidelines for how our results could be applied to formulations having many trilinear monomials overlapping on the same variables. In Chapter 4, we will see that our results are very robust for scenarios where there is a high degree of overlap between trilinear monomials. Another important issue is how to effectively make branching decisions in the context of our relaxations. Guided by our volume results, we have made some significant progress in this direction (see [55] and Chapter 5).

It would be natural and certainly difficult to extend our work to multilinear monomials having $n > 3$. In particular, advances for the important case of $n = 4$ could have immediate impact; [11] found, via experiments, that composing a trilinear and bilinear convexification in the manner suggested by $(x_i x_j)x_k x_l$ was a good strategy. They further observed sensitivity to the bounds on the variables, but they reached no clear conclusion on how to factor in that aspect. Restricting to this type of convexification, we could apply our results by substituting $w \in [a_i a_j, b_i b_j]$ to arrive at the trilinear monomial $w x_k x_l$, which can then be analyzed and relaxed according to our methodology. Of course, for a general quadrilinear monomial, there are six choices of which pair of variables will be treated as $\{x_i, x_j\}$, so we could analyze all six possibilities and take the best overall.

Also, there is the possibility of extending our results on trilinear monomials to (i) box domains that are not necessarily non-negative, (ii) domains other than boxes, and (iii) other low-dimensional functions.

With regard to (i), this is likely to be conceptually very straightforward; the analysis should be very similar to what we have completed here. However, it is not hard to imagine that it would be quite laborious. Without the convenient assumption of non-negativity on the variable bounds, there would be multiple cases to consider, and as we have seen, the analysis for just one case is lengthy.

When it comes to (iii), a possible example of an interesting low-dimensional function that may be amenable to analysis is considered in [20]. They study a family of MINLO problems referred to as *indicator-induced* $\{0,1\}$*-mixed-integer non-linear programs*. These contain binary indicator variables which each control a subset of the decision variables. When an indicator variable is 'turned off' it forces some of the decision variables to assume a fixed value, and when it is 'turned on' it forces them to belong to a convex set. These indicator-induced MINLO problems are very useful and can be used to model a wide range of interesting applications. For example, the quadratic cost uncapacitated facility location problem in [19]; the network design problem under queuing delay first considered by [7]; portfolio optimization problems such as [41], and job scheduling problems as in [2]. The problem substructures induced by this modelling technique have (at least) two alternative intuitive convexifications. The simplest instances of these substructures are in three dimensions, but unlike the convexifications for trilinear monomials they are not linear. It is known that the simpler convexification is dominated by the other, but to our knowledge there has been no analytic comparisons on how much worse it can be. Comparing the volumes of the two sets would give an explanation for the computational differences observed, and additionally give interesting insights into what kind of differences in volume have a real impact on the implementation of algorithms.

In general, we hope that our work is just a first step in using volume to better understand and mathematically quantify the tradeoffs involved in developing sBB strategies for factorable formulations.

3.8 Technical lemmas

Throughout the proofs, we have repeatedly claimed that certain quantities are non-negative for every choice of $a_1, a_2, a_3, b_1, b_2, b_3$, such that, $0 < a_i < b_i$, for all i and

$$a_1b_2b_3 + b_1a_2a_3 \leq b_1a_2b_3 + a_1b_2a_3 \leq b_1b_2a_3 + a_1a_2b_3.$$

In this section, we provide proofs for the cases that are not immediate. As will become apparent, we need to demonstrate that many different 6-variable polynomials are non-negative on the relevant parameter space. Generally, such demonstrations can be tricky global-optimization problems, and in many cases "sum-of-squares" proofs are not available; rather, we often make somewhat ad hoc arguments. Still, we can place some efficiency on all of this by establishing some technical lemmas.

3.8.1 Useful lemmas

We begin with the following lemmas that will be helpful in establishing the non-negativity of certain quantities:

Lemma 3.9. For all choices of parameters that meet our assumptions, we have: $b_1a_2 - a_1b_2 \geq 0$, $b_1a_3 - a_1b_3 \geq 0$ and $b_2a_3 - a_2b_3 \geq 0$.

Proof. $(b_3 - a_3)(b_1a_2 - a_1b_2) = b_1a_2b_3 + a_1b_2a_3 - a_1b_2b_3 - b_1a_2a_3 \geq 0$ by our original assumptions Ω. This implies $b_1a_2 - a_1b_2 \geq 0$, because $b_3 - a_3 > 0$. $b_1a_3 - a_1b_3 \geq 0$ and $b_2a_3 - a_2b_3 \geq 0$ follow from Ω in a similar way.

□

Lemma 3.10. Let $A, B, C, D, E, F \in \mathbb{R}$ with $A \geq B \geq C \geq 0$ and $D \geq 0$, $E \geq 0$, $F \leq 0$. Also let, $D + E + F = 0$. Then $AD + BE + CF \geq 0$.

Proof. $AD + BE + CF = AD + BE - C(D + E) \geq BD + BE - CD - CE = (B - C)(D + E) \geq 0$.

□

Lemma 3.11. Let $A, B, C, D, E, F \in \mathbb{R}$ with $A \geq B \geq C \geq 0$ and $D \geq 0$, $E \leq 0$, $F \leq 0$. Also let, $D + E + F = 0$. Then $AD + BE + CF \geq 0$.

Proof. $AD + BE + CF = -A(E + F) + BE + CF = E(B - A) + F(C - A) \geq 0$.

□

Lemma 3.12. 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. $AC + BD \geq B(C + D) \geq 0$.

□

3.8.2 Proving non-negativity

3.8.2.1

Substituting $M(v_3^9, v_3^{11})$ into inequality 2.25 of the convex hull, we obtain:

$$\big(-a_1^2a_2^2a_3 + a_1^2a_2b_2b_3 + a_1a_2^2a_3b_1 + 2a_1a_2a_3b_1b_2 - 2a_1a_2b_1b_2b_3 \\ -a_1a_3b_1b_2^2 - a_2^2a_3b_1^2 + a_2^2b_1^2b_3 - a_2b_1^2b_2b_3 + b_1^2b_2^2b_3\big) \Big/ 2(b_1b_2 - a_1a_2),$$

the numerator of which can be rewritten as

$$b_1b_2\Big((b_1b_3-a_1a_3)(b_2-a_2)\Big)+b_1a_2\Big((b_1a_2-a_1b_2)(b_3-a_3)\Big)+a_1a_2\Big((b_2b_3-a_2a_3)(a_1-b_1)\Big),$$

which is non-negative by Lemmas 3.9 and 3.10.

3.8.2.2

Substituting $M(v_3^{10}, v_3^{12})$ into inequality 2.26 of the convex hull, we obtain:

$$\begin{aligned}\big(&-a_1^2a_2^2a_3 + a_1^2a_2a_3b_2 - a_1^2a_3b_2^2 + a_1^2b_2^2b_3 + a_1a_2^2b_1b_3 + 2a_1a_2a_3b_1b_2\\ &-2a_1a_2b_1b_2b_3 - a_1b_1b_2^2b_3 - a_2a_3b_1^2b_2 + b_1^2b_2^2b_3\big)\Big/2(b_1b_2 - a_1a_2),\end{aligned}$$

the numerator of which can be rewritten as

$$b_1b_2\Big((b_2b_3-a_2a_3)(b_1-a_1)\Big)+a_1b_2\Big((b_1a_2-a_1b_2)(a_3-b_3)\Big)+a_1a_2\Big((b_1b_3-a_1a_3)(a_2-b_2)\Big),$$

which is non-negative by Lemmas 3.9 and 3.11.

3.8.2.3

Substituting point v_3^{11} into inequality 2.25 of the convex hull, or substituting v_3^{12} into inequality 2.26, we obtain:

$$\begin{aligned}\big(&-a_1^2a_2^2a_3 + a_1^2a_2b_2b_3 + a_1a_2^2b_1b_3 + 3a_1a_2a_3b_1b_2 - 3a_1a_2b_1b_2b_3\\ &-a_1a_3b_1b_2^2 - a_2a_3b_1^2b_2 + b_1^2b_2^2b_3\big)\Big/(b_1b_2 - a_1a_2).\end{aligned}$$

The numerator can be rewritten as

$$\begin{aligned}&b_3\Big(a_1a_2(a_1b_2 - b_1b_2) + b_1a_2(a_1a_2 - a_1b_2) + b_1b_2(b_1b_2 - a_1a_2)\Big)\\ &+ a_3\Big(a_1a_2(b_1b_2 - a_1a_2) + b_1a_2(a_1b_2 - b_1b_2) + b_1b_2(a_1a_2 - a_1b_2)\Big) \;=: b_3Y + a_3Z.\end{aligned}$$

Then we can see $Y + Z = (b_2 - a_2)(b_1 - a_1)(b_1b_2 - a_1a_2)$, which is non-negative. Furthermore, by Lemma 3.11, we have $Y \geq 0$. Therefore, by Lemma 3.12 the numerator is non-negative.

3.8.2.4

Substituting $M(v_1^9, v_1^{10})$ into inequality 2.19 of the convex hull, we obtain:

$$\begin{aligned}\big(&-2a_1a_2^2a_3b_3 + a_1a_2^2b_3^2 - a_1a_2a_3^2b_2 + 4a_1a_2a_3b_2b_3 - a_1a_2b_2b_3^2 - a_1b_2^2b_3^2 + a_2^2a_3^2b_1\\ &+a_2^2a_3b_1b_3 - a_2^2b_1b_3^2 - 4a_2a_3b_1b_2b_3 + 2a_2b_1b_2b_3^2 + a_3b_1b_2^2b_3\big)\Big/2(b_2b_3 - a_2a_3),\end{aligned}$$

the numerator of which can be rewritten as

$$b_2b_3\Big(b_2(b_1a_3-a_1b_3)+a_2(b_1b_3-2b_1a_3+a_1a_3)\Big)+a_2b_3\Big((b_1-a_1)(b_2-a_2)(b_3-a_3)\Big)$$
$$+a_2a_3\Big(b_2(2a_1b_3-b_1b_3-a_1a_3)+a_2(b_1a_3-a_1b_3)\Big) =: b_2b_3X+a_2b_3Y+a_2a_3Z.$$

Now, we write $X =: b_2V + a_2W$ and see that $V + W = (b_1 - a_1)(b_3 - a_3) \geq 0$ and, by Lemma 3.9, $V \geq 0$. Therefore $X \geq 0$ by Lemma 3.12. Because $X + Y + Z = 0$, by Lemma 3.10 we have that the numerator is non-negative.

3.8.2.5

Substituting $M(v_1^9, v_1^{10})$ into inequality 2.20 of the convex hull, we obtain:

$$\big(-a_1a_2^2a_3b_3-2a_1a_2a_3^2b_2+4a_1a_2a_3b_2b_3+a_1a_3^2b_2^2-a_1a_3b_2^2b_3-a_1b_2^2b_3^2+a_2^2a_3^2b_1$$
$$+a_2a_3^2b_1b_2-4a_2a_3b_1b_2b_3+a_2b_1b_2b_3^2-a_3^2b_1b_2^2+2a_3b_1b_2^2b_3\big)\Big/2(b_2b_3-a_2a_3),$$

the numerator of which can be rewritten as

$$b_2b_3\Big(b_3(b_1a_2-a_1b_2)+a_3(b_1b_2-2b_1a_2+a_1a_2)\Big)+b_2a_3\Big((b_1-a_1)(b_2-a_2)(b_3-a_3)\Big)$$
$$+a_2a_3\Big(b_3(2a_1b_2-b_1b_2-a_1a_2)+a_3(b_1a_2-a_1b_2)\Big) =: b_2b_3X+b_2a_3Y+a_2a_3Z.$$

Now, we write $X =: b_3V + a_3W$ and see that $V + W = (b_1 - a_1)(b_2 - a_2) \geq 0$ and, by Lemma 3.9 $V \geq 0$. Therefore $X \geq 0$ by Lemma 3.12. Because $X + Y + Z = 0$, by Lemma 3.10 we have that the numerator is non-negative.

3.8.2.6

Substituting $M(v_1^9, v_1^{11})$ into inequality 2.20 of the convex hull, we obtain:

$$\big(-a_1a_2^2a_3b_3+2a_1a_2a_3b_2b_3-a_1a_3^2b_2^2+a_1a_3b_2^2b_3-a_1b_2^2b_3^2+a_2^2a_3^2b_1-a_2a_3^2b_1b_2$$
$$-2a_2a_3b_1b_2b_3+a_2b_1b_2b_3^2+a_3^2b_1b_2^2\big)\Big/2(b_2b_3-a_2a_3),$$

the numerator of which can be rewritten as

$$b_2b_3\Big((b_1a_2-a_1b_2)(b_3-a_3)\Big)+b_2a_3\Big((b_2a_3-a_2b_3)(b_1-a_1)\Big)+a_2a_3\Big((b_1a_3-a_1b_3)(a_2-b_2)\Big),$$

which is non-negative by Lemmas 3.10 and 3.9.

3.8.2.7

Substituting $M(v_1^9, v_1^{11})$ into inequality 2.21 of the convex hull, we obtain:

$$(b_2 - a_2)\left(a_1a_2a_3b_3 - a_1b_2b_3^2 - a_2a_3^2b_1 - a_2a_3b_1b_3 + a_2b_1b_3^2 + a_3^2b_1b_2\right)\Big/2(b_2b_3 - a_2a_3),$$

where the second multiplicand of the numerator can be rewritten as

$$b_3\Big(b_3(b_1a_2 - a_1b_2)\Big) + a_3\Big(a_1a_2b_3 - b_1a_2a_3 + b_1b_2a_3 - b_1a_2b_3\Big) =: b_3Y + a_3Z,$$

now $Y + Z = (b_1a_3 - a_1b_3)(b_2 - a_2) \geq 0$ (Lemma 3.9), and $Y \geq 0$ (Lemma 3.9), therefore by Lemma 3.12 we have that $b_3Y + a_3Z$ is non-negative.

3.8.2.8

Substituting $M(v_1^9, v_1^{11})$ into inequality 2.22 of the convex hull, we obtain:

$$(b_3 - a_3)\left(a_1a_2^2b_3 - a_1a_2b_2b_3 + a_1a_3b_2^2 - a_1b_2^2b_3 - a_2^2a_3b_1 + a_2b_1b_2b_3\right)\Big/2(b_2b_3 - a_2a_3),$$

where the second multiplicand of the numerator can be rewritten as

$$b_2\Big(b_3(b_1a_2 - a_1b_2) + a_1(b_2a_3 - a_2b_3)\Big) + a_2\Big(a_2(a_1b_3 - b_1a_3)\Big) =: b_2Y + a_2Z,$$

where $Y + Z = (b_1a_2 - a_1b_2)(b_3 - a_3) \geq 0$ and $Y \geq 0$ (both Lemma 3.9), therefore by Lemma 3.12 we have that this term is non-negative.

3.8.2.9

Substituting $M(v_1^9, v_1^{11})$ into inequality 2.27 of the convex hull, we obtain:

$$\begin{aligned}\big(&-a_1a_2^2a_3^2 + a_1a_2^2a_3b_3 - a_1a_2^2b_3^2 + 2a_1a_2a_3b_2b_3 - a_1a_3b_2^2b_3 + a_2^2b_1b_3^2\\ &+a_2a_3^2b_1b_2 - 2a_2a_3b_1b_2b_3 - a_2b_1b_2b_3^2 + b_1b_2^2b_3^2\big)\Big/2(b_2b_3 - a_2a_3),\end{aligned}$$

the numerator of which can be rewritten as

$$b_2b_3\Big((b_1b_3 - a_1a_3)(b_2 - a_2)\Big) + a_2b_3\Big((b_2a_3 - a_2b_3)(a_1 - b_1)\Big) + a_2a_3\Big((b_1b_2 - a_1a_2)(a_3 - b_3)\Big),$$

which is non-negative by Lemmas 3.11 and 3.9.

3.8.2.10

Substituting $M(v_1^{10}, v_1^{12})$ into inequality 2.19 of the convex hull, we obtain:

$$\begin{aligned}\big(&-a_1a_2^2b_3^2 - a_1a_2a_3^2b_2 + 2a_1a_2a_3b_2b_3 + a_1a_2b_2b_3^2 - a_1b_2^2b_3^2 + a_2^2a_3^2b_1\\ &-a_2^2a_3b_1b_3 + a_2^2b_1b_3^2 - 2a_2a_3b_1b_2b_3 + a_3b_1b_2^2b_3\big) \Big/ 2(b_2b_3 - a_2a_3),\end{aligned}$$

the numerator of which can be rewritten as

$$b_2b_3\Big((b_1a_3-a_1b_3)(b_2-a_2)\Big)+a_2b_3\Big((b_2a_3-a_2b_3)(a_1-b_1)\Big)+a_2a_3\Big((b_1a_2-a_1b_2)(a_3-b_3)\Big),$$

which is non-negative by Lemma 3.11 and Lemma 3.9.

3.8.2.11

Substituting $M(v_1^{10}, v_1^{12})$ into inequality 2.21 of the convex hull, we obtain:

$$(b_3 - a_3)\left(a_1a_2a_3b_2 - a_1b_2^2b_3 - a_2^2a_3b_1 + a_2^2b_1b_3 - a_2a_3b_1b_2 + a_3b_1b_2^2\right) \Big/ 2(b_2b_3 - a_2a_3),$$

where the second multiplicand of the numerator can be rewritten as

$$b_2\Big(b_2(b_1a_3 - a_1b_3)\Big) + a_2\Big(b_1(a_2b_3 - b_2a_3) + a_3(a_1b_2 - b_1a_2)\Big) \;=: b_2Y + a_2Z,$$

where $Y+Z = (b_1a_2-a_1b_2)(b_3-a_3) \geq 0$ and $Y \geq 0$ (both by Lemma 3.9). Therefore, by Lemma 3.12 we have that $b_2Y + a_2Z$ is non-negative.

3.8.2.12

Substituting $M(v_1^{10}, v_1^{12})$ into inequality 2.22 of the convex hull, we obtain:

$$(b_2 - a_2)\left(a_1a_2b_3^2 + a_1a_3^2b_2 - a_1a_3b_2b_3 - a_1b_2b_3^2 - a_2a_3^2b_1 + a_3b_1b_2b_3\right) \Big/ 2(b_2b_3 - a_2a_3),$$

where the second multiplicand of the numerator can be rewritten as

$$b_3\Big(a_1(a_2b_3 - b_2a_3) + b_2(b_1a_3 - a_1b_3)\Big) + a_3\Big(a_3(a_1b_2 - b_1a_2)\Big) \;=: b_3Y + a_3Z,$$

where $Y + Z = (b_1a_3 - a_1b_3)(b_2 - a_2) \geq 0$ and $Z \leq 0 \implies Y \geq 0$ (both by Lemma 3.9). Therefore by Lemma 3.12 we have that $b_3Y + a_3Z$ is non-negative.

3.8.2.13

Substituting $M(v_1^{10}, v_1^{12})$ into inequality 2.24 of the convex hull, we obtain:

$$\big(-a_1a_2^2a_3^2 + a_1a_2a_3^2b_2 + 2a_1a_2a_3b_2b_3 - a_1a_2b_2b_3^2 - a_1a_3^2b_2^2 + a_2^2a_3b_1b_3 \\ -2a_2a_3b_1b_2b_3 + a_3^2b_1b_2^2 - a_3b_1b_2^2b_3 + b_1b_2^2b_3^2\big) \Big/ 2(b_2b_3 - a_2a_3),$$

the numerator of which simplifies to

$$b_2b_3\Big((b_1b_2 - a_1a_2)(b_3 - a_3)\Big) + b_2a_3\Big((b_2a_3 - a_2b_3)(b_1 - a_1)\Big) + a_2a_3\Big((b_1b_3 - a_1a_3)(a_2 - b_2)\Big),$$

which is non-negative by Lemmas 3.9 and 3.10.

3.8.2.14

Substituting point v_1^9 into inequality 2.20 of the convex hull, or substituting point v_1^{10} into inequality 2.19, we obtain:

$$\big(-a_1a_2^2a_3b_3 - a_1a_2a_3^2b_2 + 3a_1a_2a_3b_2b_3 - a_1b_2^2b_3^2 + a_2^2a_3^2b_1 \\ -3a_2a_3b_1b_2b_3 + a_2b_1b_2b_3^2 + a_3b_1b_2^2b_3\big) \Big/ (b_2b_3 - a_2a_3),$$

the numerator of which can be rewritten as

$$b_3\Big(b_2(b_2(b_1a_3 - a_1b_3) + a_2(a_1a_3 + b_1b_3 - 2b_1a_3))\Big) \\ + a_3\Big(a_2(b_2(2a_1b_3 - a_1a_3 - b_1b_3) + a_2(b_1a_3 - a_1b_3))\Big) \;=: b_3Y + a_3Z.$$

Now, we write $Y =: b_2^2V + a_2b_2W$ and see that $V + W = (b_1 - a_1)(b_3 - a_3) \geq 0$ and, by Lemma 3.9 $V \geq 0$. Therefore $Y \geq 0$ by Lemma 3.12. Because $Y + Z = (b_1a_3 - a_1b_3)(b_2 - a_2)^2 \geq 0$ (Lemma 3.9), by Lemma 3.12 we have that the numerator is non-negative.

3.8.2.15

Substituting point v_1^{11} into inequality 2.27 of the convex hull, or substituting point v_1^{12} into 2.24, we obtain:

$$\big(-a_1a_2^2a_3^2 + 3a_1a_2a_3b_2b_3 - a_1a_2b_2b_3^2 - a_1a_3b_2^2b_3 + a_2^2a_3b_1b_3 \\ +a_2a_3^2b_1b_2 - 3a_2a_3b_1b_2b_3 + b_1b_2^2b_3^2\big) \Big/ (b_2b_3 - a_2a_3),$$

the numerator of which can be rewritten as

$$b_1\Big(b_2b_3(b_2b_3 - a_2a_3) + b_2a_3(a_2a_3 - a_2b_3) + a_2a_3(a_2b_3 - b_2b_3)\Big) \\ + a_1\Big(b_2b_3(a_2a_3 - a_2b_3) + b_2a_3(a_2b_3 - b_2b_3) + a_2a_3(b_2b_3 - a_2a_3)\Big) \;=: b_1Y + a_1Z,$$

where: $Y + Z = (b_2b_3 - a_2a_3)(b_2 - a_2)(b_3 - a_3) \geq 0$, and by Lemma 3.11 we have $Y \geq 0$. Therefore, by Lemma 3.12 we have that the numerator is non-negative.

3.8.2.16

Substituting $M(v_1^9, v_1^{10})$ into inequality 2.21 of the convex hull, we obtain

$$(2b_2b_3 - a_2b_3 - b_2a_3)(a_1a_2a_3 - a_1b_2b_3 - 2b_1a_2a_3 + b_1a_2b_3 + b_1b_2a_3)\Big/2(b_2b_3 - a_2a_3),$$

where the second multiplicand of the numerator can be rewritten as

$$b_2(b_1a_3 - a_1b_3) + a_2(a_1a_3 - 2b_1a_3 + b_1b_3),$$

which is non-negative by Lemmas 3.9 and 3.12.

3.8.2.17

Substituting $M(v_1^9, v_1^{10})$ into inequality 2.22 of the convex hull, we obtain

$$(a_2b_3 + b_2a_3 - 2a_2a_3)(a_1a_2b_3 + a_1b_2a_3 - 2a_1b_2b_3 - b_1a_2a_3 + b_1b_2b_3)\Big/2(b_2b_3 - a_2a_3),$$

where the second multiplicand of the numerator can be rewritten as

$$b_2(b_1b_3 - 2a_1b_3 + a_1a_3) + a_2(a_1b_3 - b_1a_3),$$

which is non-negative by Lemmas 3.9 and 3.12.

3.8.2.18

Substituting point v_1^9 into inequality 2.21 of the convex hull, we obtain

$$b_3(b_2 - a_2)\Big(b_2(b_1a_3 - a_1b_3) + a_2(a_1a_3 - 2b_1a_3 + b_1b_3)\Big)\Big/(b_2b_3 - a_2a_3),$$

which is non-negative by Lemmas 3.9 and 3.12.

3.8.2.19

Substituting point v_1^9 into inequality 2.22 of the convex hull, we obtain

$$a_2(b_3 - a_3)\Big(b_2(b_1b_3 - 2a_1b_3 + a_1a_3) + a_2(a_1b_3 - b_1a_3)\Big)\Big/(b_2b_3 - a_2a_3),$$

which is non-negative by Lemmas 3.9 and 3.12.

3.8.2.20

Substituting point v_1^{10} into inequality 2.21 of the convex hull, we obtain

$$b_2(b_3 - a_3)\Big(b_2(b_1a_3 - a_1b_3) + a_2(a_1a_3 - 2b_1a_3 + b_1b_3)\Big)\Big/(b_2b_3 - a_2a_3),$$

which is non-negative by Lemmas 3.9 and 3.12.

3.8.2.21

Substituting point v_1^{10} into inequality 2.22 of the convex hull, we obtain

$$a_3(b_2 - a_2)\Big(b_2(b_1b_3 - 2a_1b_3 + a_1a_3) + a_2(a_1b_3 - b_1a_3)\Big)\Big/(b_2b_3 - a_2a_3),$$

which is non-negative by Lemmas 3.9 and 3.12.

CHAPTER 4

Experimental Justification of Volume

4.1 Introduction

In this chapter, we experimentally validate our choice of volume as a comparison measure for alternative convexifications of triple products. In §4.2, we recall the important tradeoff between the tightness and simplicity of a convexification. In §4.3, we formally introduce box cubic programming problems, which will be important throughout the chapter. In §4.4, we address the relationship between the volume of a relaxation and the result of optimizing over it. In §4.5, we provide the details of our computational experiments and results. Finally, in §4.6, we make some concluding remarks. The computational work for this chapter was completed with the assistance of Han Yu, a University of Michigan masters student (see [57]).

4.2 Measuring relaxations via volume in mathematical optimization

We have seen that in the context of mathematical optimization, there is often a natural tradeoff in the tightness of a convexification and the difficulty of optimizing over it. This idea was emphasized by [24] in the context of mixed-integer nonlinear optimization (MINLO) (see also the recent work [13]). Of course this is also a well-known phenomenon for difficult 0/1 linear-optimization problems, where very tight relaxations are available via extremely heavy semidefinite-programming relaxations (e.g., the Lasserre hierarchy), and the most effective relaxation for branch-and-bound/cut may well not be the tightest. Earlier, again in the context of mathematical optimization, [25] introduced the idea of using volume as a measure of the tightness of a convex relaxation (for fixed-charge and vertex packing problems). Most of that

mathematical work was asymptotic, seeking to understand the quality of families of relaxations with a growing number of variables, but some of it was also substantiated experimentally in [24].

4.3 Box cubic programming problems

In Chapter 3 we applied the idea of [25], but in the context of the low-dimensional relaxations of basic functions that arise in sBB. We derived analytic expressions for the volume of $\mathcal{P}_h$ as well as all three of the natural double-McCormick relaxations. The expressions are formulae in the six constants $0 \leq a_i < b_i$, $i = 1, 2, 3$. In doing so, we quantify the quality of the various relaxations and provide recommendations for which to use. The results of Chapter 3 are theoretical. Their utility for guiding modelers and sBB implementers depends on the belief that volume is a good measure of the quality of a relaxation. Morally, this belief is based on the idea that with no prior information on the form of an objective function, the solution of a relaxation should be assumed to occur with a uniform density on the feasible region. The contribution of this chapter is to experimentally validate the robustness of this theory in the context of a particular use case, optimizing multilinear cubics over boxes (box cubic programming or 'boxcup'). There is considerable literature on techniques for optimizing quadratics, much of which is developed and validated in the context of so-called box quadratic programming or 'boxqp' problems, where we minimize $\sum_{i,j} q_{ij} x_i x_j$ over a box domain in $\mathbb{R}^n$. So our boxcup problems, for which we minimize $\sum_{i,j,k} q_{ijk} x_i x_j x_k$ over a box domain in $\mathbb{R}^n$, are natural and defined in the same spirit and for the same purpose as the boxqp problems.

A main result of Chapter 3 is Corollary 3.5; an ordering of the three natural relaxations of individual trilinear monomials by volume. But this result is for $n = 3$. Our experiments validate the theory as applied to our use case. We demonstrate that in the setting of boxcup problems, the average objective discrepancy between relaxations very closely follows the prediction of the theory, when volumes are appropriately combined (summing the 4-th root of the volume, across the chosen relaxations of each trilinear monomial). Moreover and very importantly, we are able to demonstrate that these results are robust against sparsity of the cubic forms.

[25] defined the *idealized radius* of a polytope in $\mathbb{R}^d$ as essentially the d-th root of its volume (up to some constants depending on d). For a polytope that is very much like a ball in shape, we can expect that this quantity is (proportional to) the "average width" of the polytope. The average width arises by looking at 'max minus min',

averaged over all normalized linear objectives. So, the implicit prediction of Corollary 3.5 is that the *idealized radius* should (linearly) predict the expected 'max minus min' for normalized linear objectives. We have validated this experimentally, and looked further into the *idealized radial distance* between pairs of relaxations, finding an even higher degree of linear association.

Finally, in the important case $a_1 = a_2 = 0$, $b_1 = b_2 = 1$, Corollary 3.7 shows that the two worst relaxations have the same volume, and the greatest difference in volume between $\mathcal{P}_h = \mathcal{P}_3$ and the (two) worst relaxations occurs when $a_3 = b_3/3$. We present results of experiments that clearly show that these predictions via volume are again borne out on boxcup problems.

All in all, in this chapter we present convincing experimental evidence that volume is a good predictor for quality of relaxation in the context of sBB. Our results strongly suggest that the theoretical results of Chapter 3 are important in devising decompositions of complex functions in the context of factorable formulations and therefore our results help inform both modelers and implementers of sBB.

4.4 From volume to objective function gap

For a convex body $C \subset \mathbb{R}^d$, we denote its *volume* (i.e., Lebesgue measure) by $\text{vol}(C)$. Volume seems like an awkward measure to compare relaxations, when typically we are interested in objective-function gaps. Following [25], the *idealized radius* of a convex body $C \subset \mathbb{R}^d$ is

$$\rho(C) := \left(\text{vol}(C)/\text{vol}(B_d)\right)^{1/d},$$

where B_d is the (Euclidean) unit ball in $\mathbb{R}^d$. $\rho(C)$ is simply the radius of a ball having the same volume as C. The *idealized radial distance* between convex bodies C_1 and C_2 is simply $|\rho(C_1) - \rho(C_2)|$. If C_1 and C_2 are concentric balls, say with $C_1 \subset C_2$, then the idealized radial distance between them is the (radial) height of C_2 above C_1. The *mean semi-width* of C is simply

$$\frac{1}{2}\int_{\|c\|=1} \left(\max_{x\in C} c'x - \min_{x\in C} c'x\right) d\psi,$$

where ψ is the $(d-1)$-dimensional Lebesgue measure on the boundary of B_d, normalized so that ψ on the entire boundary is unity. If C is itself a ball, then (i) its idealized radius is in fact its radius, and (ii) its width in any unit-norm direction c is

constant, and so (iii) its (idealized) radius is equal to its mean semi-width.

Key point: What we can hope is that our relaxations of individual trilinear monomials (in a model with many overlapping trilinear monomials) are round enough so that choosing them to be of small volume (which is proportional to and monotone increasing in its idealized radius raised to the power d) is a good proxy for choosing the overall relaxation by *mean width* (which is the same as mean objective-value range). It is this that we investigate experimentally.

4.5 Computational experiments

4.5.1 Box cubic programming problems and four relaxations

Our experiments are aimed at the following natural problem which is concerned with optimizing a linear function on trinomials. Let H be a 3-uniform hypergraph on n vertices. Each hyperedge of H is a set of three vertices, and we denote the set of hyperedges by $E(H)$. If H is complete, then $|E(H)| = \binom{n}{3}$. We associate with each vertex i a variable $x_i \in [a_i, b_i]$, and with each hyperedge (i, j, k) the trinomial $x_i x_j x_k$ and a coefficient q_{ijk} $(1 \leq i < j < k \leq n)$. We now formulate the associated box cubic programming problem:

$$\min_{x \in \mathbb{R}^n} \left\{ \sum_{(i,j,k) \in E(H)} q_{ijk}\, x_i\, x_j\, x_k \;:\; x_i \in [a_i, b_i],\ i = 1, 2, \ldots, n \right\}. \qquad \text{(BOXCUP)}$$

The name is in analogy with the well-known *boxqp*, where just two terms (rather than three) are multiplied ('box' refers to the feasible region and 'qp' refers to 'quadratic program').

(BOXCUP) is a difficult non-convex global-optimization problem. Our goal here is not to solve instances of this problem, but rather to solve a number of different *relaxations* of the problem and see how the results of these experiments correlate with the volume results of Chapter 3. In this way, we seek to determine if the guidance of Chapter 3 is relevant to modelers and those implementing sBB.

We have seen how for a single trilinear term $f = x_i x_j x_k$, we can build four distinct relaxations: the convex hull of the feasible points, $\mathcal{P}_h$, and three relaxations arising from double McCormick: $\mathcal{P}_1$, $\mathcal{P}_2$ and $\mathcal{P}_3$. To obtain a relaxation of (BOXCUP), we choose a relaxation $\mathcal{P}_\ell$, for some $\ell = 1, 2, 3, h$ and apply this same relaxation method to *each* trinomial of (BOXCUP). We therefore obtain 4 distinct linear relaxations of

the form:

$$\min_{(x,f)\in\mathscr{P}_\ell} \left\{ \sum_{(i,j,k)\in E(H)} q_{ijk}\, f_{ijk} \right\}.$$

where $\mathscr{P}_\ell$, $\ell = 1, 2, 3, h$ is the polytope in dimension $|E(H)| + n$ arising from using relaxation $\mathcal{P}_\ell$ on each trinomial. This linear relaxation is a linear inequality system involving the n variables x_i ($i = 1, 2, \ldots, n$), and the $|E(H)|$ new 'function variables' f_{ijk}. Each such 'function variable' models a product $x_i\, x_j\, x_k$.

For our experiments, we randomly generate box bounds $[a_i, b_i]$ on x_i, for each $i = 1, \ldots, n$ independently, by choosing (uniformly) random pairs of integers $0 \leq a_i < b_i \leq 10$. With each realization of these bounds, we get relaxation feasible regions $\mathcal{P}_\ell$, for $\ell = 1, 2, 3, h$.

4.5.2 Three scenarios for the hypergraph H

We designed our experiments with the idea of gaining some understanding of whether our conclusions would depend on how much the trinomials overlap. So we looked at three scenarios for the hypergraph H of (BOXCUP), all with $|E(H)| = 20$ trinomials:

- Our **dense** scenario has H being a *complete* 3-uniform hypergraph on $n = 6$ vertices ($\binom{6}{3} = 20$). We note that each of the $n = 6$ variables appears in $\binom{6-1}{3-1} = 10$ of the 20 trinomials, so there is considerable overlap in variables between trinomials.

- Our **sparse** scenario has hyperedges: $\{1, 2, 3\}$, $\{2, 3, 4\}$, $\{3, 4, 5\}$ $\ldots$ $\{18, 19, 20\}$, $\{19, 20, 1\}$, $\{20, 1, 2\}$. Here we have $n = 20$ variables and each variable is in only 3 of the trinomials.

- Our **very sparse** scenario has $n = 30$ variables and each variable is in only 2 of the trinomials. We have the 10 hyperedges with the form: $\{1, 2, 3\}$, $\{4, 5, 6\}$ $\ldots$ $\{25, 26, 27\}$, $\{28, 29, 30\}$, and 10 hyperedges that we obtain by 'switching' the last node and the first node from pairs of these edges i.e., $\{1, 2, 4\}$, $\{3, 5, 6\}$ $\ldots$ $\{25, 26, 28\}$, $\{27, 29, 30\}$.

For each scenario, we generate 30 *sets* of bounds $[a_i, b_i]$ on x_i ($i = 1, \ldots, n$).

To control the variation in our results, and considering that the scaling of

$$Q := \{q_{ijk} \ : \ \{i, j, k\} \in E(H)\}$$

is arbitrary, we generate 100,000 random Q with $|E(H)|$ entries, uniformly distributed on the unit sphere in $\mathbb{R}^{|E(H)|}$.

Then, for each Q, we both minimize and maximize $\sum_{i<j<k} q_{ijk} f_{ijk}$ over each $\mathscr{P}_\ell, \ell = 1, 2, 3, h$ and each set of bounds.

4.5.3 Quality of relaxations

For each Q we take the difference in the optimal values, i.e. the maximum value minus the minimum value; this can be thought of as the width of the polytope in the direction Q. We then average these widths for each $\mathscr{P}_\ell$, $\ell = 1, 2, 3, h$, across the 100,000 realizations of Q (which results in very small standard errors), and we refer to this quantity $\omega(\mathscr{P}_\ell)$ as the *quasi mean width* of the relaxation. It is not quite the geometric mean width, because we do not have objective terms for all variables in (BOXCUP) (i.e., we have no objective terms $\sum_{i=1}^{n} c_i x_i$).

We seek to investigate how well the volume formulae, comparing the volumes of the polytopes $\mathcal{P}_\ell$ ($\ell = 1, 2, 3, h$), can be used to predict the quality of the relaxations $\mathscr{P}_\ell$ ($\ell = 1, 2, 3, h$) as measured by their quasi mean width.

Figure 4.1 consists of a plot for each scenario: dense, sparse, and very sparse. Each plot illustrates the difference in quasi mean width between $\mathscr{P}_3$ (using the 'best' double McCormick) and each of the other relaxations. Each point represents a choice of bounds and the instances are sorted by $\omega(\mathscr{P}_1) - \omega(\mathscr{P}_h)$. In all three plots $\omega(\mathscr{P}_h) - \omega(\mathscr{P}_3)$ is non-positive, which is to be expected because $\mathscr{P}_h$ is contained in each of the three double-McCormick relaxations. Furthermore the plots illustrate that the general trend is for $\omega(\mathscr{P}_2) - \omega(\mathscr{P}_3)$ and $\omega(\mathscr{P}_1) - \omega(\mathscr{P}_3)$ to be positive and also for $\omega(\mathscr{P}_1) - \omega(\mathscr{P}_3)$ to be greater than $\omega(\mathscr{P}_2) - \omega(\mathscr{P}_3)$. This agrees with Corollary 3.5 and gives strong validation for the use of volume to measure the strength of different relaxations. It confirms that given a choice of the double-McCormick relaxations, $\mathscr{P}_3$ is the one to choose.

However, there are a few exceptions to the general trend and these exceptions are most apparent in the very sparse case. In both the dense case and the sparse case we only see a deviation from the trend on a small number of occasions when $\mathscr{P}_2$ is very slightly better than $\mathscr{P}_3$. In each of these cases, the difference seems to be so small that we can really regard $\mathscr{P}_3$ and $\mathscr{P}_2$ as being equivalent from a practical viewpoint. In the very sparse case, the general trend is still followed, but we see a few more cases where $\mathscr{P}_2$ is slightly better than $\mathscr{P}_3$. We also see that in a few instances, $\mathscr{P}_1$ is better than $\mathscr{P}_2$ and occasionally even slightly better than $\mathscr{P}_3$.

However, it is important to note that when we consider the sparse and very sparse

cases, the differences in quasi mean width between *any* two of the relaxations is much smaller than these differences in the dense case. If we were to take the sparsity of H to the extreme and run our experiments with $n = 60$ and each variable only in one trinomial, the difference in quasi mean width between *any* two of the polytopes will become zero for these boxcup problems. Therefore, it is not surprising that our results diverge from the general trend as H becomes sparser.

Using the common technique of 'performance profiles' (see [14]), we can illustrate the differences in quasi mean width of the three double-McCormick relaxations in another way. We obtained the matlab code "perf.m" which was adapted to create these plots from the link contained in [54].

Figure 4.2 shows a performance profile for each of the dense, sparse and very sparse scenarios. For each choice of bounds, $\mathscr{P}_h$ gives the least quasi mean width (because it is contained in each of the other relaxations). Our performance profiles display the fraction of instances where the quasi mean width of $\mathscr{P}_\ell$ is within a factor α of the mean width of $\mathscr{P}_h$, for $\ell = 1, 2, 3$. The plots are natural log plots where the horizonal axis is $\tau := \ln(\alpha)$. Using this measure, we see that the trend in *all* cases is that $\mathscr{P}_3$ dominates $\mathscr{P}_2$ which in turn dominates $\mathscr{P}_1$. In the very sparse case, we see that $\mathscr{P}_3$ and $\mathscr{P}_2$ are very close for small factors α. In general, all three relaxations are within a small factor of the hull. Displaying the results in this manner gives us a way to see quickly which relaxation performs best for the majority of instances. Again, we see agreement with the prediction of Corollary 3.5 and confirmation that $\mathscr{P}_3$ is the best double-McCormick relaxation.

4.5.4 Validating the relationship between volume and objective gap

Using the volume formulae, we calculate the volume of the relaxation for each individual trinomial, $\mathcal{P}_\ell$. We then we take the fourth root of these volumes and sum over all $|E(H)|$ trinomials to obtain a kind of 'aggregated idealized radius' for each relaxation and each set of bounds. Restricting our attention to the dense scenario, in Figure 4.3, we compare these aggregated idealized radii with quasi mean width, across all relaxations $\mathscr{P}_\ell, \ell = 1, 2, 3, h$ and each set of bounds (each point in each scatter plot corresponds to a choice of bounds). We see a high R^2 coefficient in all cases, so we may conclude that volume really is a good predictor of relaxation width.

We also compute the difference in width between polytope pairs: $\mathscr{P}_h$ and $\mathscr{P}_3$, $\mathscr{P}_3$ and $\mathscr{P}_2$, $\mathscr{P}_2$ and $\mathscr{P}_1$ for each direction Q. We then average these width differences for each polytope and each set of bounds, across the 100,000 realizations of Q. We refer to this result as the *quasi mean width difference* of the pair of polytopes. In

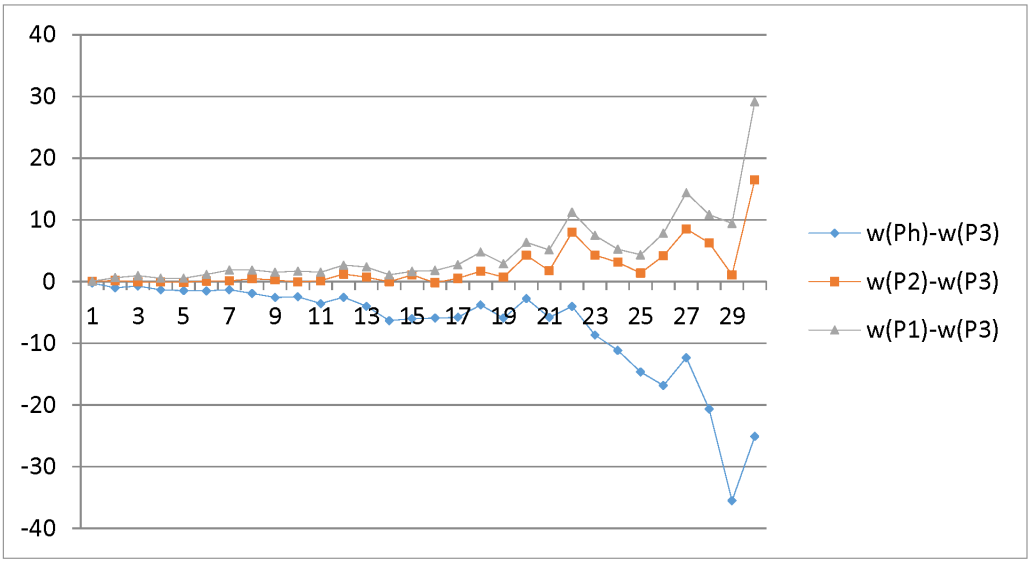
Figure 4.4, we similarly compare aggregated idealized radial differences with quasi mean width differences. We see even higher R^2 coefficients, validating volume as an excellent predictor of average objective gap between pairs of relaxations.

4.5.5 A worst case

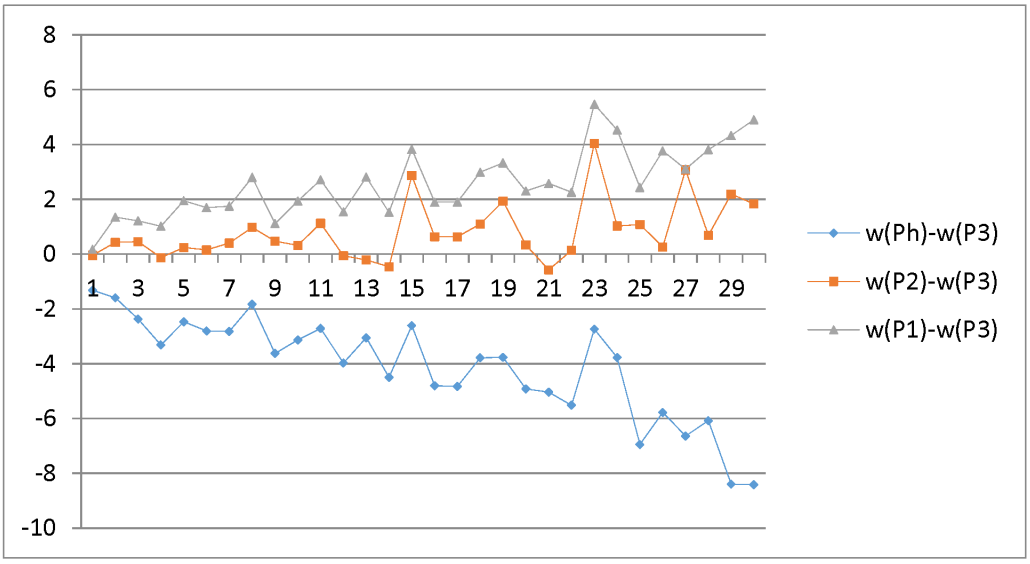
Our final set of experiments relate to a 'worst case' as described in Corollary 3.7. In the important special case of $a_1 = a_2 = 0$ and $b_1 = b_2 = 1$, the two 'bad' double-McCormick relaxations have the same volume and the 'good' double McCormick is exactly the hull. In addition, the greatest difference in volume between the hull and the bad relaxations occurs when $a_3 = b_3/3$.

We compute the same results as we have discussed before (i.e. the differences in quasi mean width between the relaxations) with $n = 6$, but now instead of using random bounds, for each trinomial we fix $a_1 = a_2 = 0$ and $b_1 = b_2 = 1$. We also fix b_3 and run the experiments for $a_3 = 1, 2, \ldots, b_3 - 1$. Here, we only consider the $\binom{5}{3} = 10$ trinomials that have the form $x_j x_k x_6$.

Figure 4.5 displays a plot of these results for $b_3 = 30, 60, 90, 120$ and 150. From the inequality systems we know that $\mathscr{P}_h$ is exactly $\mathscr{P}_3$, therefore we are interested in the comparison between: $\mathscr{P}_2$ and $\mathscr{P}_3$, and $\mathscr{P}_2$ and $\mathscr{P}_1$. From the plots of these differences, we see exactly what we would expect given the volume formulae. The difference in mean width between $\mathscr{P}_2$ and $\mathscr{P}_1$ is very small; from a practical standpoint it is essentially zero. The difference in mean width between $\mathscr{P}_2$ and $\mathscr{P}_3$ is always positive, indicating again that $\mathcal{P}_3$ is the best choice of double-McCormick relaxation. In addition, we observe that the maximum difference falls close to $a_3 = b_3/3$ in all cases, demonstrating again that volume is a good predictor of how well a relaxation behaves.

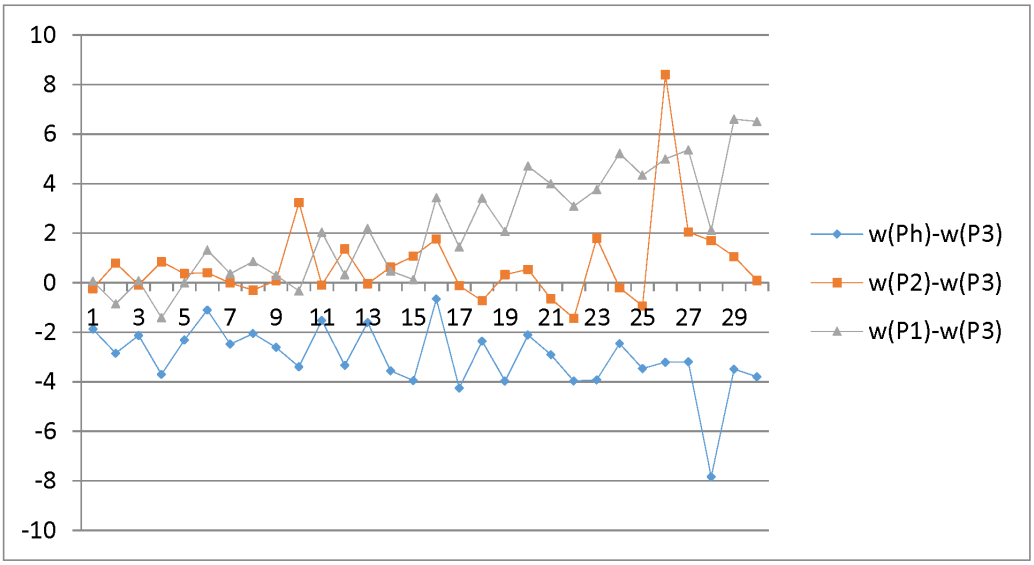


(a) dense case



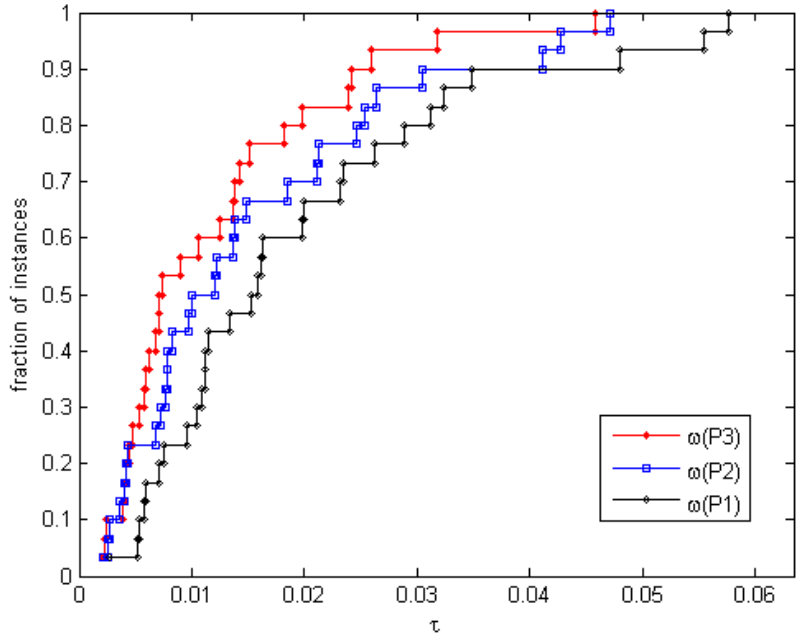
(b) sparse case

Figure 4.1: Quasi-mean-width differences



(c) very-sparse case

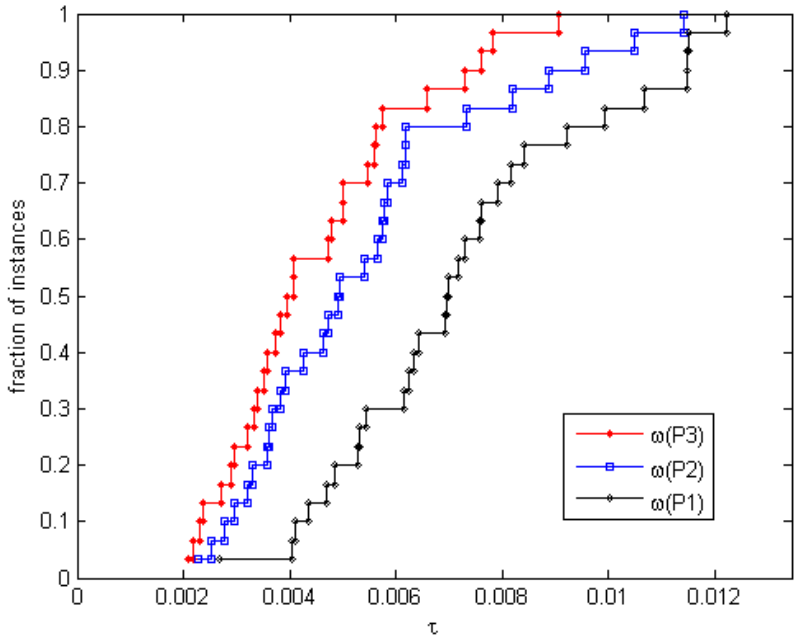
Figure 4.1: Quasi-mean-width differences



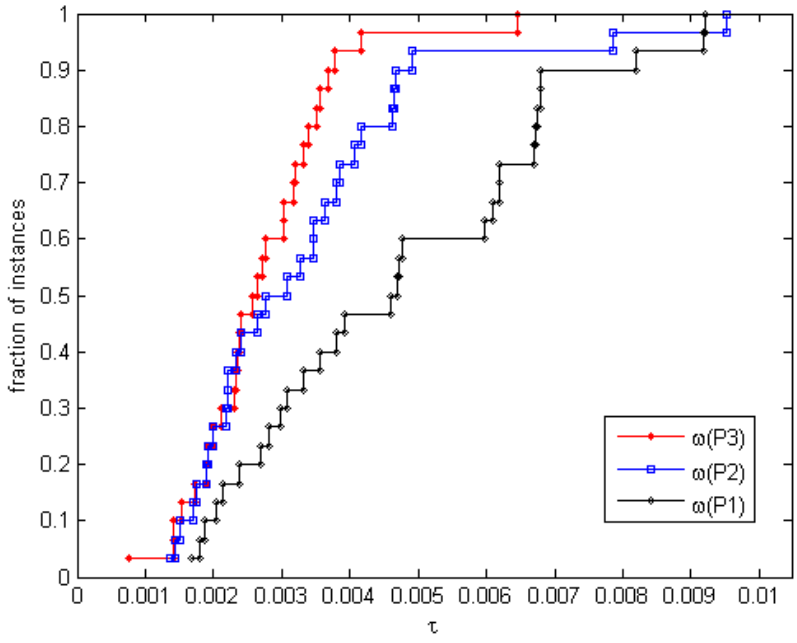
(a) dense case

Figure 4.2: Quasi-mean-width performance profiles

Displays the fraction of instances where the quasi mean width of $\mathscr{P}_\ell$ is within a factor $\alpha = e^\tau$ of the mean width of $\mathscr{P}_h$. Note that for small τ, $e^\tau \approx 1 + \tau$.



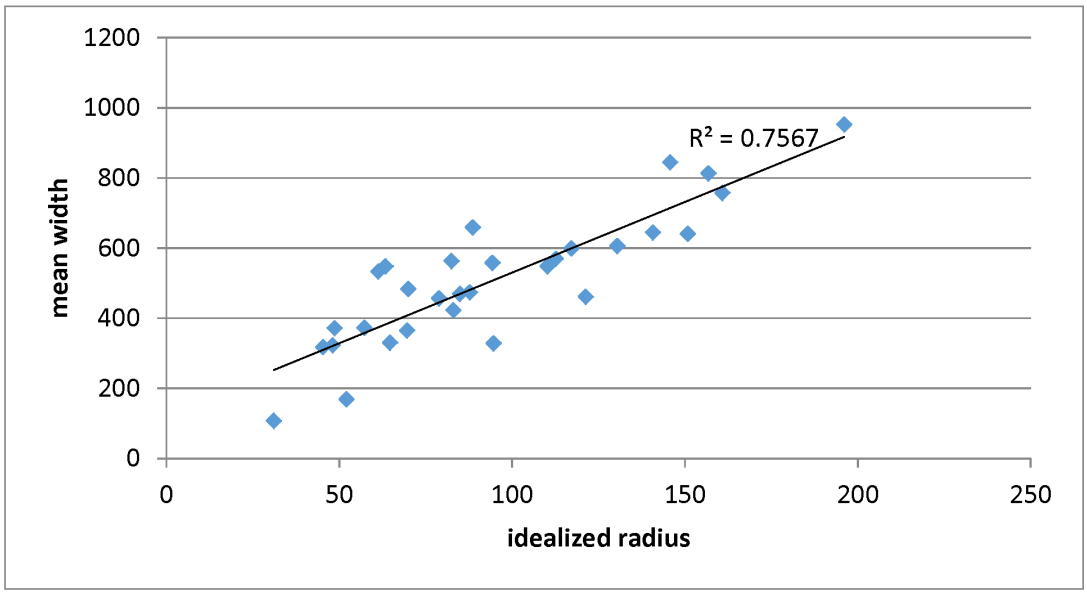
(b) sparse case



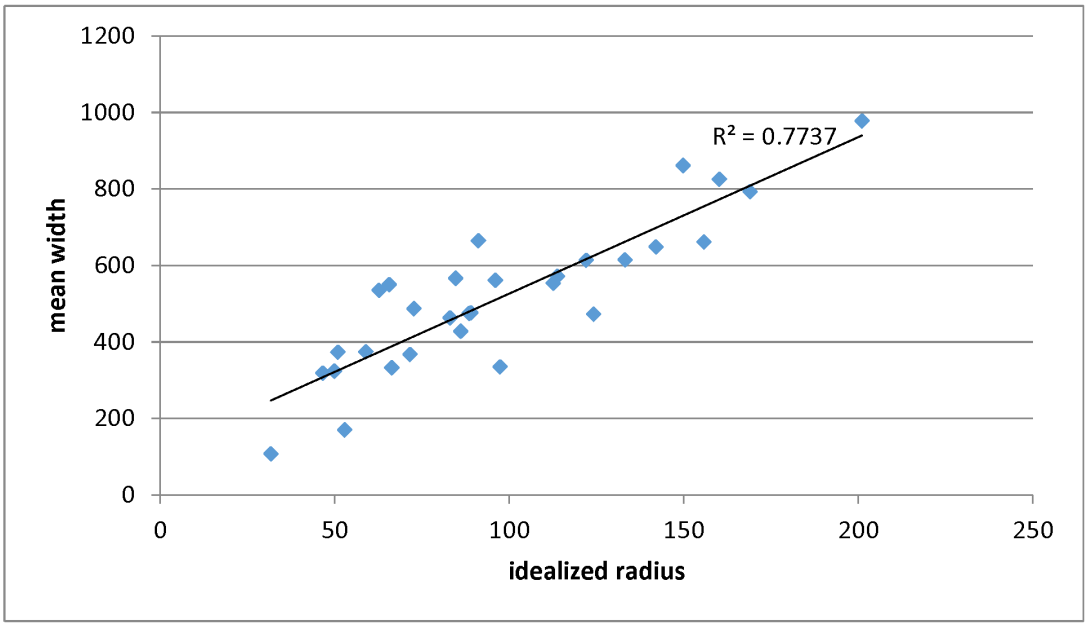
(c) very-sparse case

Figure 4.2: Quasi-mean-width performance profiles

Displays the fraction of instances where the quasi mean width of $\mathscr{P}_\ell$ is within a factor $\alpha = e^\tau$ of the mean width of $\mathscr{P}_h$. Note that for small τ, $e^\tau \approx 1 + \tau$.

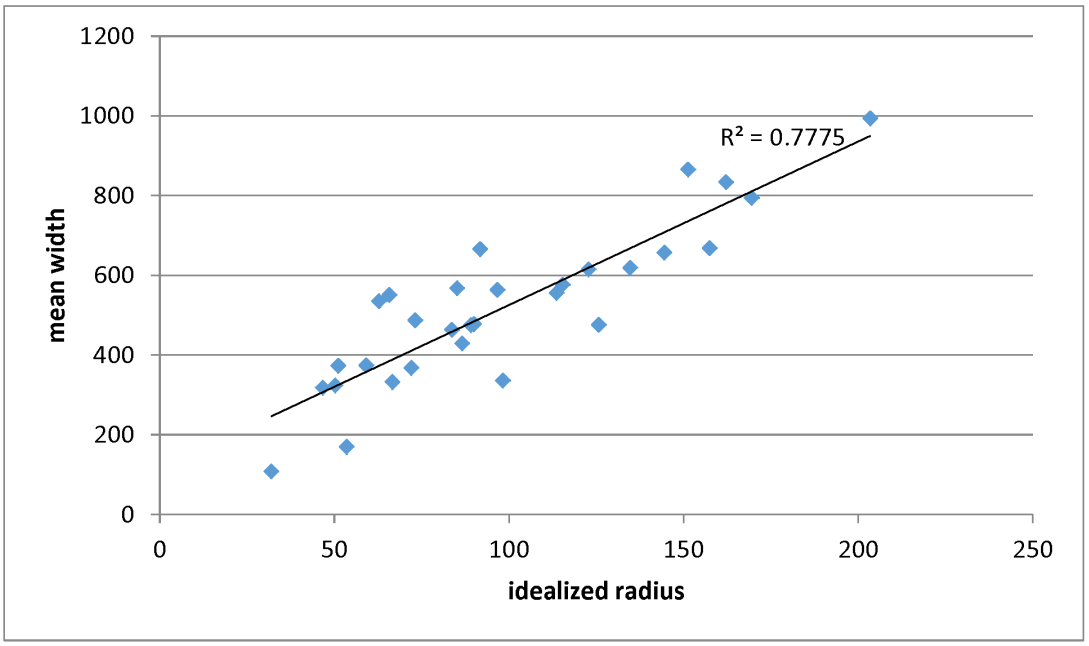


(a) $\mathscr{P}_h$

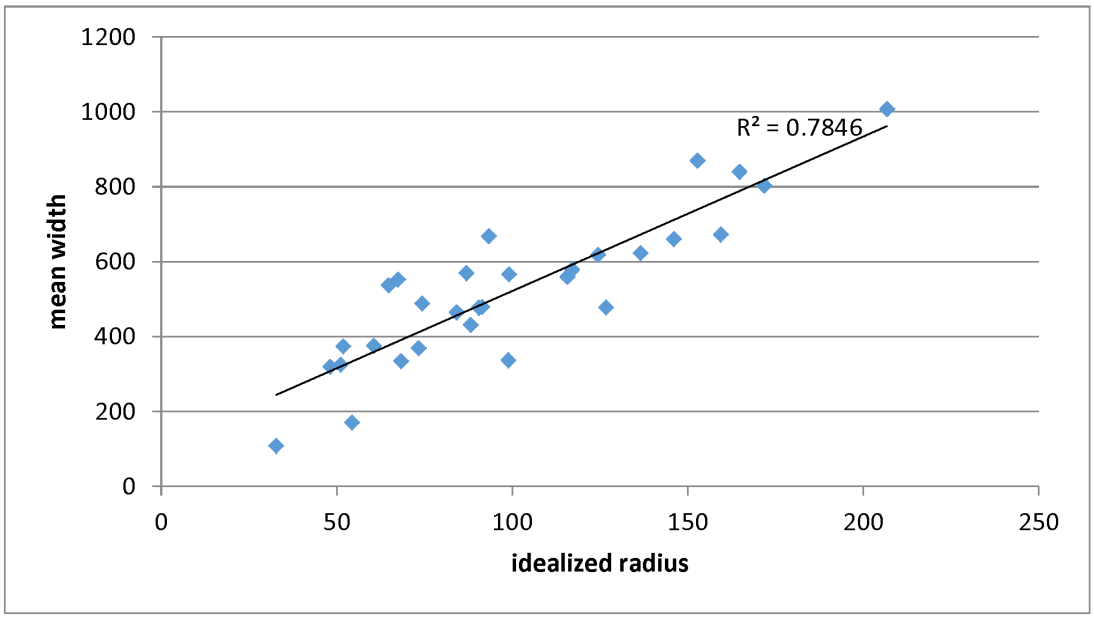


(b) $\mathscr{P}_3$

Figure 4.3: Idealized radius predicting quasi mean width

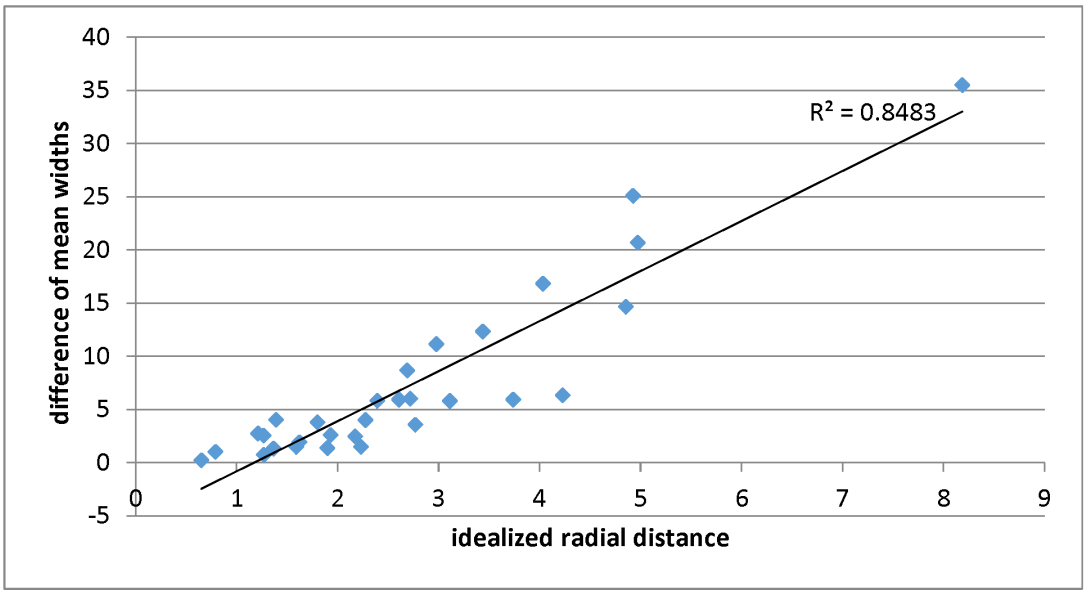


(c) $\mathscr{P}_2$

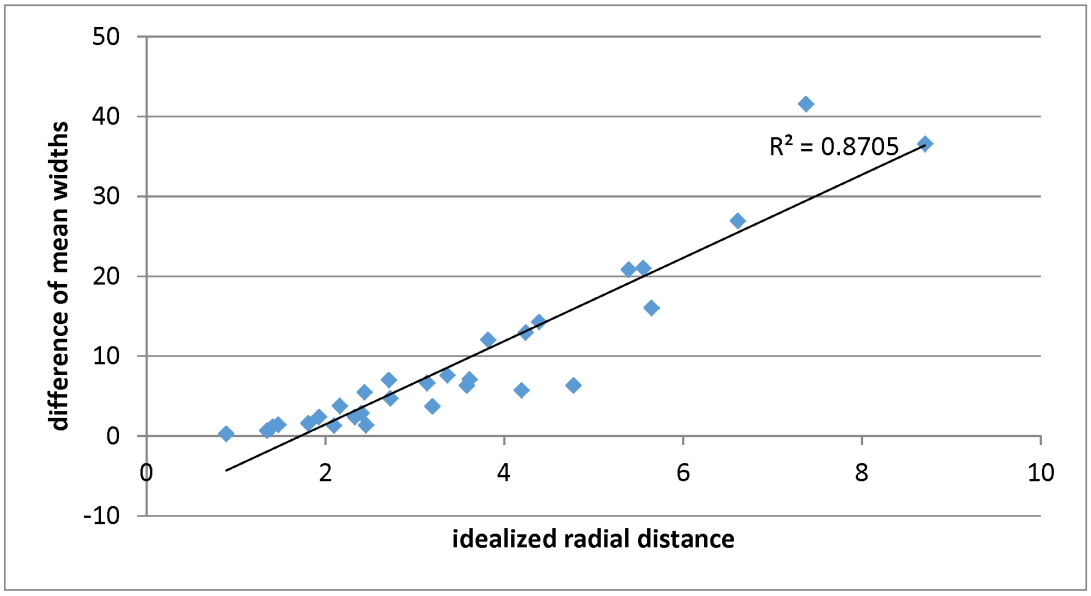


(d) $\mathscr{P}_1$

Figure 4.3: Idealized radius predicting quasi mean width

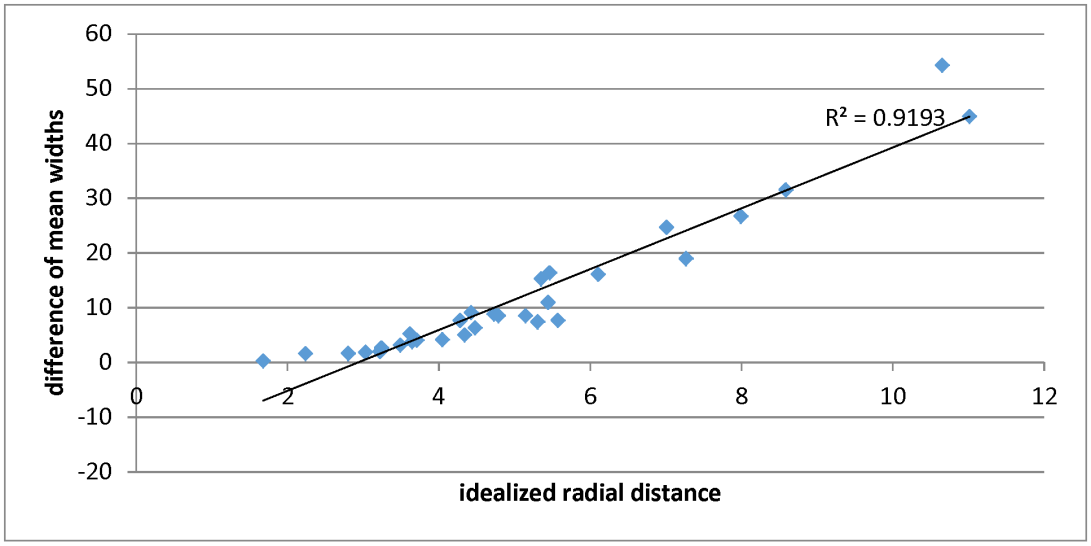


(a) $\mathscr{P}_3$ vs. $\mathscr{P}_h$



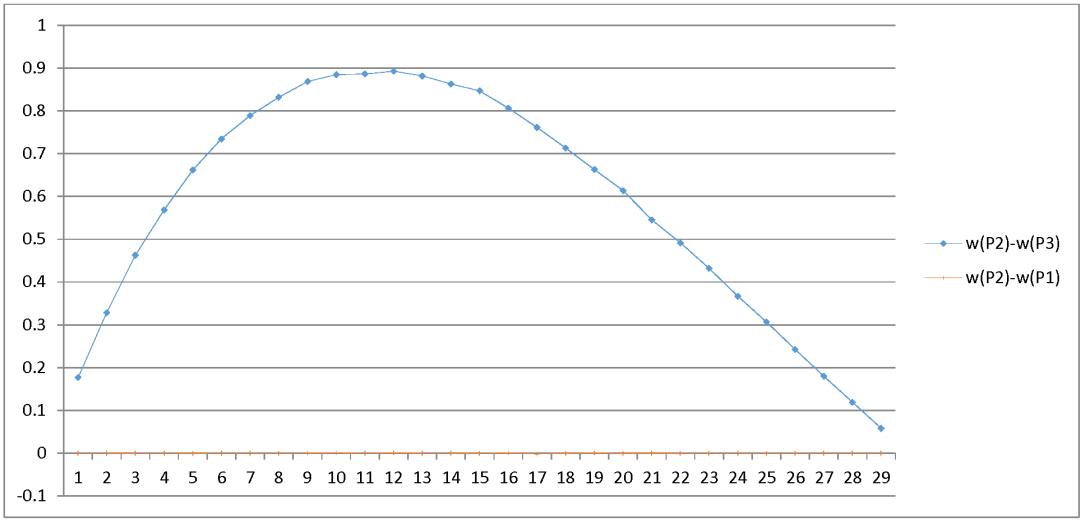
(b) $\mathscr{P}_2$ vs. $\mathscr{P}_h$

Figure 4.4: Idealized radial distance predicting quasi mean width difference



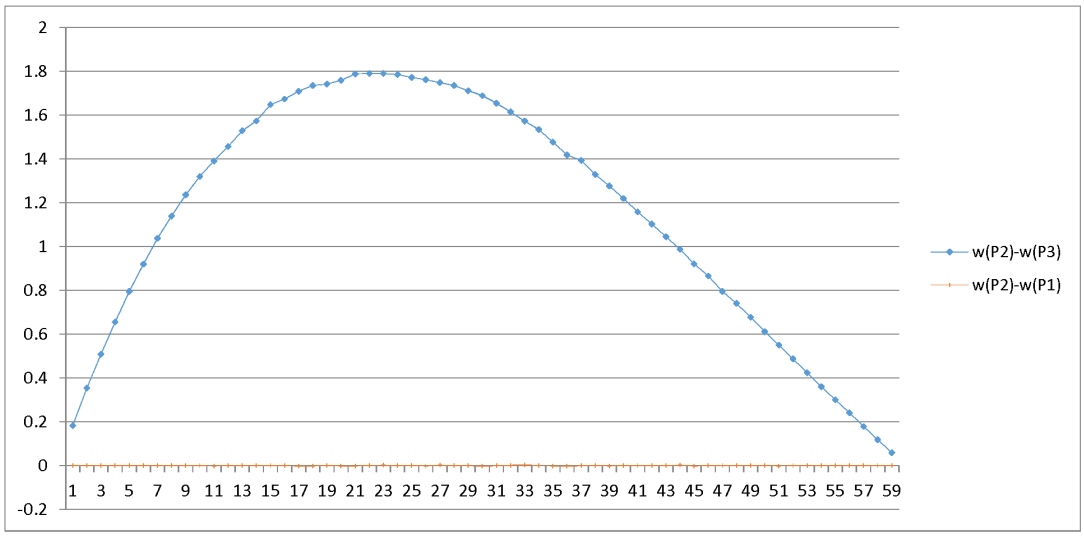
(c) $\mathscr{P}_1$ vs. $\mathscr{P}_h$

Figure 4.4: Idealized radial distance predicting quasi mean width difference

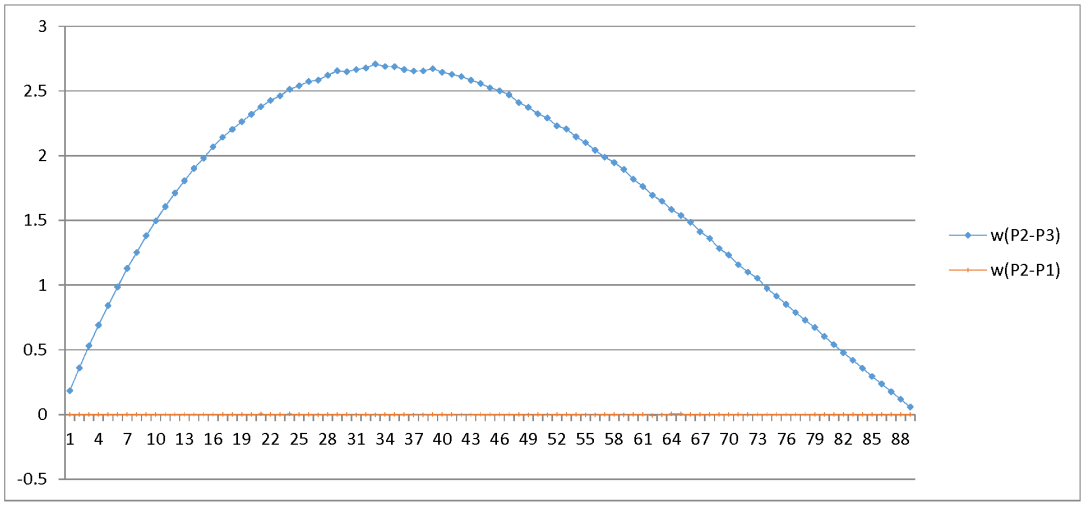


(a) $b_3 = 30$

Figure 4.5: Worst-case analysis for a_3 ($a_1 = a_2 = 0$, $b_1 = b_2 = 1$)

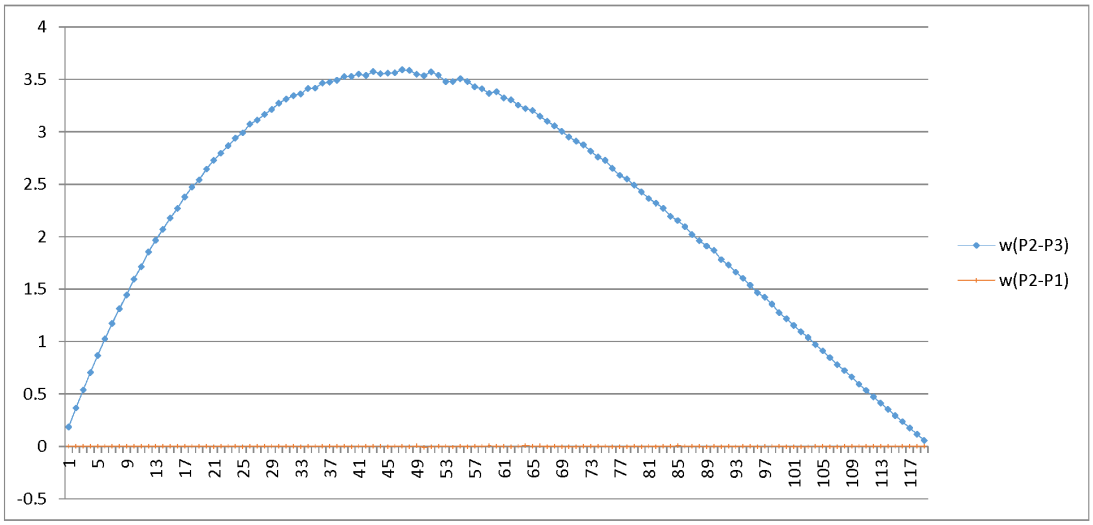


(b) $b_3 = 60$

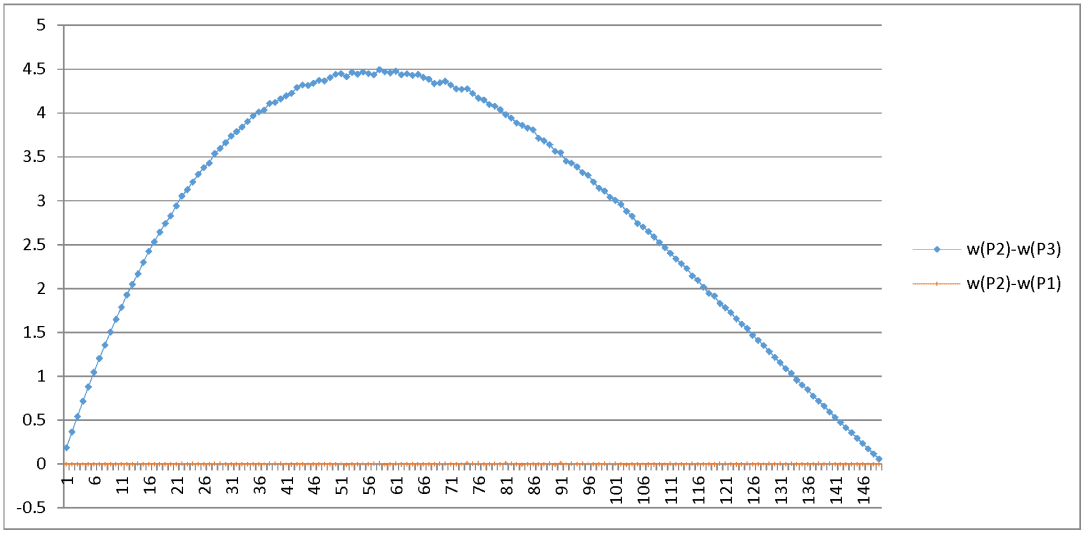


(c) $b_3 = 90$

Figure 4.5: Worst-case analysis for a_3 ($a_1 = a_2 = 0$, $b_1 = b_2 = 1$)



(d) $b_3 = 120$



(e) $b_3 = 150$

Figure 4.5: Worst-case analysis for a_3 ($a_1 = a_2 = 0$, $b_1 = b_2 = 1$)

4.6 Concluding remarks and future work

In this chapter, we experimentally validated our claim that volume is a good measure of objective function gap for triple products. We compared the performance of alternative convexifications for triple products in boxcup problems (BOXCUP) with varying levels of sparseness. We provided evidence that our theoretical results from Chapter 3 really do have the potential to improve algorithm performance once their implications are implemented.

An obvious direction for further work is to complete similar experiments on problems where triple products occur, but *not* as boxcup problems. This would validate our results for triple products in other contexts. Moreover, it would be nice to complete experiments comparing the sBB solution time for various problem instances when each of the four different triple-product convexification methods are used. In particular, it would be interesting to computationally explore if and how the variable bounds impact whether the full convex hull or the best double McCormick is the preferred choice of convexification.

CHAPTER 5

Using Volume to Guide Branching-Point Selection

5.1 Introduction

Along with utilizing good convex relaxations, other important issues in the effective implementation of sBB are the choice of branching variable and the selection of the branching point. In this final chapter, we use our volume results to analyze branching-point selection for triple products. We consider branching on variable x_1, and for the case when the full convex hull is used as the choice of convexification, we obtain an algorithm to compute the branching point that results in the least total volume for the two subproblems. For the case when the best double-McCormick relaxation is used, we present a partial analysis. All the technical lemmas, theorems, and propositions referenced throughout this chapter are stated and established in §5.5.

In §5.2, we briefly describe the current branching practices of software, before presenting our results in §5.3. We make some concluding remarks in §5.4, and finally, as noted earlier, §5.5 contains the technical theorems and lemmas referred to throughout the chapter. This work expands on and corrects the results of [55].

5.2 Current branching practice

There has been extensive computational research into branching-point selection (e.g., see [6]). 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 value of the variable at the solution of the current relaxation, and/or restricting the branching choice to a central part of the interval. A typical way to achieve this (see [60]) is to choose the branching

point as follows:

$$\max\Big\{a_i + \beta(a_i - b_i),\ \min\Big\{b_i - \beta(b_i - a_i),\ \alpha\hat{x}_i + (1-\alpha)(a_i + b_i)/2\Big\}\Big\}, \qquad (5.1)$$

where $\hat{x}_i$ 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

$$[a_i + \beta(a_i - b_i),\ b_i - \beta(b_i - a_i)]$$

to the weighted combination $\alpha\hat{x}_i + (1-\alpha)(a_i + b_i)/2$ (of the current value and the interval midpoint), thus explicitly ruling out branching in the bottom and top β fraction of the interval. Note that if $\beta \leq (1-\alpha)/2$, then the explicit restriction is redundant, 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 software packages use a variety of values for the parameters α and β. The method (mostly) employed by `SCIP` (see [61] 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 β. The current *default* settings of `ANTIGONE` ([36] and [39]), `BARON` ([48]) and `Couenne` (see [6] 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 Table 5.1.

Solver	α	β	
`SCIP`	1.00	0.20	$\not\leq (1-\alpha)/2 = 0.0$
`ANTIGONE`	0.75	0.10	$\leq (1-\alpha)/2 = 0.125$
`BARON`	0.70	0.01	$\leq (1-\alpha)/2 = 0.15$
`Couenne`	0.25	0.20	$\leq (1-\alpha)/2 = 0.375$

Table 5.1: Default parameter settings

The alternatives are based somewhat on intuition, and of course on 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 that sometimes supersede selecting a branching point according to the formula (5.1); in particular, functional forms involved, the solution of the current relaxation, available incumbent solutions, comple-

mentarity considerations, etc. Our work in this chapter is based solely on analyzing the volumes of relaxations for triple products, with the goal of helping to inform and in some cases mathematically support the choice of a branching point.

5.3 Results

In this section, and as we have been doing throughout, we focus on trilinear monomials. As before, for the variables $x_i \in [a_i, b_i]$, $i = 1, 2, 3$, we assume that the following conditions hold:

$$\begin{aligned} &0 \leq a_i < b_i \text{ for } i = 1, 2, 3, \quad \text{and} \\ &a_1b_2b_3 + b_1a_2a_3 \;\leq\; b_1a_2b_3 + a_1b_2a_3 \;\leq\; b_1b_2a_3 + a_1a_2b_3, \end{aligned} \tag{Ω}$$

We also recall that because we are only considering non-negative bounds, the latter part of Ω is equivalent to:

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

Consider what happens when we pick a branching variable, and branch at a given point: we obtain two subproblems, now with different bounds on the branching variable. The upper bound of the branching variable in the left subproblem becomes the value of the branching point, as does the lower bound of the branching variable in the right subproblem. We reconvexify the two subproblems using our chosen method of convexification (i.e. the full hull or a double McCormick), and we can sum the volumes from both subproblems 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 Figure 1.2, which illustrates reconvexifying after branching in sBB. 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.

Throughout this section, we focus on what happens when we branch on variable x_1. We chose to analyze x_1 because some early investigation suggested that branching on this variable (given that we branch at an optimal point) may result in the least total volume when compared with branching on x_2 or x_3. This intuition is merely based on partial results; however, we do believe that following this work on x_1, it will also be possible to complete the analysis of branching on variable x_2 and variable x_3. For the convex-hull convexification, we can see from the structure of the volume

function (see Theorem 3.1), that once the analysis has been completed for one of the variables x_2 or x_3, completing the last one will be trivial.

We can compute the volume of the relaxation for each of the subproblems using the appropriate theorem from Chapter 3. 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 Ω. To illustrate this, consider the left subproblem obtained by branching on variable x_1 at some point $c_1 \in [a_1, b_1]$. For this left subproblem, 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' to b_1, then Ω will likely remain satisfied, however as c_1 becomes smaller, there comes a point where eventually 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 think about the right subproblem 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 likely remain satisfied, however, as c_1 becomes larger, eventually the labeling must change. This critical point for the right subproblem is at $c_1 = \frac{b_1 a_2}{b_2}$.

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

$$\frac{b_1 a_2}{b_2} \leq \frac{a_1 b_2}{a_2} \quad \Longleftrightarrow \quad \frac{a_2^2}{b_2^2} \leq \frac{a_1}{b_1} \quad \Longleftrightarrow \quad b_1 a_2^2 \leq a_1 b_2^2,$$

and second when

$$\frac{b_1 a_2}{b_2} > \frac{a_1 b_2}{a_2} \quad \Longleftrightarrow \quad \frac{a_2^2}{b_2^2} > \frac{a_1}{b_1} \quad \Longleftrightarrow \quad b_1 a_2^2 > a_1 b_2^2.$$

The case of equality, i.e., $\frac{b_1 a_2}{b_2} = \frac{a_1 b_2}{a_2}$, is arbitrarily included with Case 1. In fact, when equality holds, the analysis that follows in the remainder of this chapter 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 5.2. 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.

5.3.1 The convex-hull convexification

In this section, we assume that the convexification used is the full convex hull. We note that because of the structure of the volume function of the convex hull, (see Theorem 3.1), the second and third variables are interchangeable. This means that

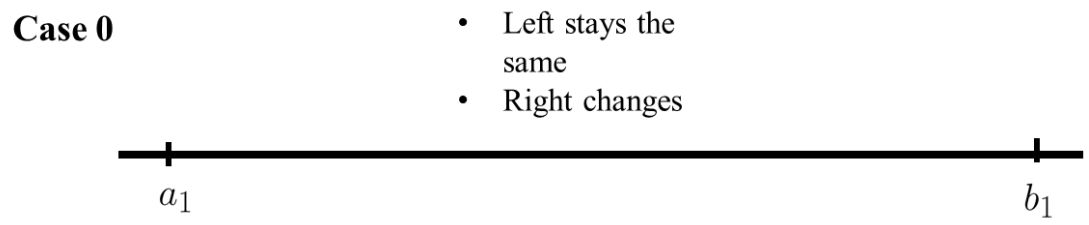
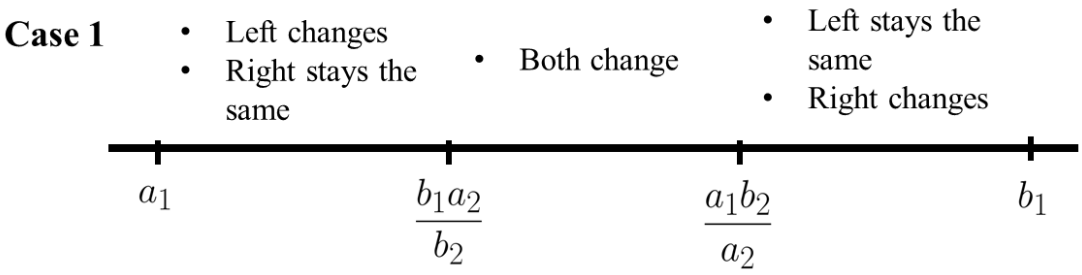
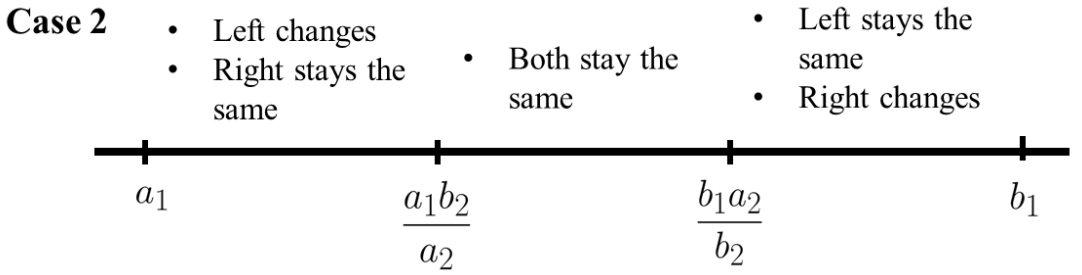


Figure 5.1: Variable labeling as the branching point varies in Case 0



(a) Case 1



(b) Case 2

Figure 5.2: Variable labeling as the branching point varies in Case 1 and Case 2

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.

5.3.1.1 Case 0: $a_2 = 0$

From the condition Ω, we know that $a_2 = 0 \Rightarrow a_1 = 0$. In this special case, the labeling for the left subproblem does not change no matter how small the upper bound becomes. Conversely, the labeling for the right subproblem changes as soon

as the lower bound becomes positive. We therefore have the picture shown in Figure 5.1, and only 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 (later defined as q_3), is the midpoint of the interval.

5.3.1.2 Case 1: $\frac{b_1 a_2}{b_2} \leq \frac{a_1 b_2}{a_2}$

We define

$$V(l_1, u_1, l_2, u_2, l_3, u_3) := (u_1 - l_1)(u_2 - l_2)(u_3 - l_3) \times (u_1(5u_2u_3 - l_2u_3 - u_2l_3 - 3l_2l_3) + l_1(5l_2l_3 - u_2l_3 - l_2u_3 - 3u_2u_3))/24$$

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 subproblems after branching at point c_1, is given by the following parameterized function:

$$TV(c_1) := \begin{cases} V_1(c_1) & a_1 \leq c_1 \leq \frac{b_1 a_2}{b_2}; \\ V_2(c_1) & \frac{b_1 a_2}{b_2} \leq c_1 \leq \frac{a_1 b_2}{a_2}; \\ V_3(c_1) & \frac{a_1 b_2}{a_2} \leq c_1 \leq b_1, \end{cases} \tag{5.2}$$

where:

$$\begin{aligned} V_1(c_1) &:= V(a_2, b_2, a_1, c_1, a_3, b_3) + V(c_1, b_1, a_2, b_2, a_3, b_3), \\ V_2(c_1) &:= V(a_2, b_2, a_1, c_1, a_3, b_3) + V(a_2, b_2, c_1, b_1, a_3, b_3), \\ V_3(c_1) &:= V(a_1, c_1, a_2, b_2, a_3, b_3) + V(a_2, b_2, c_1, b_1, a_3, b_3). \end{aligned}$$

This is a piecewise-quadratic function in c_1. It is straightforward to check that the function is continuous over its domain. Furthermore, by observing that the leading coefficient of each piece is non-negative for all parameter values satisfying Ω, we conclude that each piece is convex.

The leading coefficient of $V_1(c_1)$ is: $\dfrac{(b_3 - a_3)(b_2 - a_2)(8b_2b_3 - 6a_2a_3 - 2a_2b_3)}{24} \geq 0.$

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

$$\frac{(b_3 - a_3)(b_2 - a_2)(6b_2b_3 + 2b_2a_3 - 6a_2a_3 - 2a_2b_3)}{24} \geq 0.$$

The leading coefficient of $V_3(c_1)$ is: $\dfrac{(b_3 - a_3)(b_2 - a_2)(6b_2b_3 + 2b_2a_3 - 8a_2a_3)}{24} \geq 0.$

Figure 5.3 gives some intuition of what this function *could* look like.

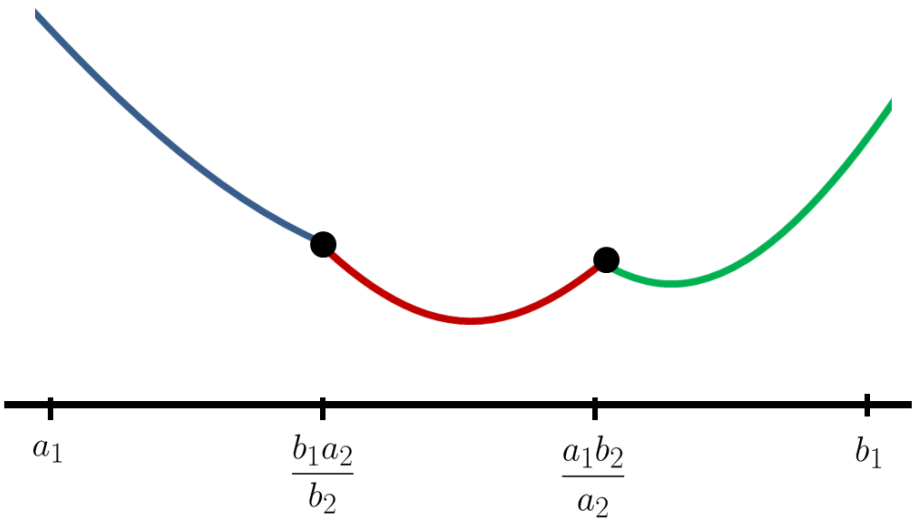


Figure 5.3: Illustration of a continuous piecewise-quadratic function

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 obtain 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 subproblems, 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 5.8 (see §5.5), 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:

$$q_1 - \frac{a_1 + b_1}{2} = \frac{(b_3 - a_3)(b_1a_2 - a_1b_2)}{2(4b_2b_3 - a_2b_3 - 3a_2a_3)} \geq 0, \tag{5.3}$$

and

$$q_3 - \frac{a_1 + b_1}{2} = \frac{(a_3 - b_3)(b_1a_2 - a_1b_2)}{2(3b_2b_3 + b_2a_3 - 4a_2a_3)} \leq 0. \tag{5.4}$$

We therefore have:

$$q_1 \geq q_2 = \frac{a_1 + b_1}{2} \geq q_3. \tag{5.5}$$

From this, we can observe that if $q_3 \geq \frac{a_1b_2}{a_2}$, then $q_2 \geq q_3 \geq \frac{a_1b_2}{a_2}$, 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 $V_3(q_3) \leq V_3\left(\frac{a_1b_2}{a_2}\right)$, and by Lemma 5.5 (see §5.5) we know that $V_3\left(\frac{a_1b_2}{a_2}\right) \leq V_2\left(\frac{b_1a_2}{b_2}\right)$.

If this does not occur, i.e. $q_3 < \frac{a_1b_2}{a_2}$, then if $\frac{b_1a_2}{b_2} \leq \frac{a_1+b_1}{2} \leq \frac{a_1b_2}{a_2}$, 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 $\frac{a_1b_2}{a_2}$ is the minimizer by Lemma 5.5 (see §5.5).

As an interesting side point, we also note that if it were possible to have $q_1 \leq \frac{b_1 a_2}{b_2}$, then $q_3 \leq q_2 \leq q_1 \leq \frac{b_1 a_2}{b_2}$, 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 definition of q_1 as the minimizer of V_1, we have that $V_1(q_1) \leq V_1\left(\frac{b_1 a_2}{b_2}\right)$, and by Proposition 5.4 (see §5.5) we know that $V_1(q_1) \leq V_2\left(\frac{a_1 b_2}{a_2}\right)$. However, by Lemma 5.8 (see §5.5) we have already discarded this case.

5.3.1.3 Case 2: $\frac{b_1 a_2}{b_2} > \frac{a_1 b_2}{a_2}$

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 subproblems after branching at point c_1, is given by the following parameterized function (this is similar, but distinct, from the function in Case 1):

$$\widehat{TV}(c_1) := \begin{cases} V_1(c_1) & a_1 \leq c_1 \leq \frac{b_1 a_2}{b_2}; \\ V_4(c_1) & \frac{b_1 a_2}{b_2} \leq c_1 \leq \frac{a_1 b_2}{a_2}; \\ V_3(c_1) & \frac{a_1 b_2}{a_2} \leq c_1 \leq b_1, \end{cases} \tag{5.6}$$

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 non-negative for all parameter values satisfying Ω, we know that each piece is convex.

$$\text{The leading coefficient of } V_4(c_1) \text{ is: } \frac{(b_3 - a_3)(b_2 - a_2)(8b_2 b_3 - 8a_2 a_3)}{24} \geq 0.$$

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 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 5.9 (see §5.5), 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 $q_2 \geq q_3 \geq \frac{b_1 a_2}{b_2}$, 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 5.7 (see §5.5) 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 $\frac{a_1 b_2}{a_2} \leq \frac{a_1 + b_1}{2} \leq \frac{b_1 a_2}{b_2}$, 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_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 5.7 (see §5.5).

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 $q_3 \leq q_2 \leq q_1 \leq \frac{a_1 b_2}{a_2}$, 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 5.6 (see §5.5) we know that $V_1(q_1) \leq V_4\left(\frac{b_1 a_2}{b_2}\right)$. However, by Lemma 5.9 (see §5.5) we have already discarded this case.

5.3.1.4 Algorithm for obtaining the optimal branching point

Using our analysis, we can now specify a formal algorithm for obtaining the branching point that will minimize the total volume when branching on variable x_1.

Data: $(a_1, b_1, a_2, b_2, a_3, b_3)$

Result: Branching point for variable x_1 resulting in least total volume

initialization

$q_2 := \frac{a_1+b_1}{2}$

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

if $a_2 = 0$ **then** `// Case 0 (see §5.3.1.1)`

  return $q_3 (= q_2)$

end

else

  if $\frac{a_1b_2}{a_2} \geq \frac{b_1a_2}{b_2}$ **then** `// Case 1 (see §5.3.1.2)`

    if $q_3 \geq \frac{a_1b_2}{a_2}$ **then**

      return q_3

    end

    else if $q_2 \geq \frac{b_1a_2}{b_2}$ **and** $q_2 \leq \frac{a_1b_2}{a_2}$ **then**

      return q_2

    else

      return $\frac{a_1b_2}{a_2}$

  end

  else `// Case 2 (see §5.3.1.3)`

    if $q_3 \geq \frac{b_1a_2}{b_2}$ **then**

      return q_3

    end

    else if $q_2 \geq \frac{a_1b_2}{a_2}$ **and** $q_2 \leq \frac{b_1a_2}{b_2}$ **then**

      return q_2

    else

      return $\frac{b_1a_2}{b_2}$

  end

end

Algorithm 1: Optimal branching point for x_1

5.3.1.5 Some examples

We can illustrate these piecewise-quadratic functions for the possible outcomes of Algorithm 1. In this illustration, we focus on Case 1, and therefore Figure 5.4 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 $\frac{a_1 b_2}{a_2}$.

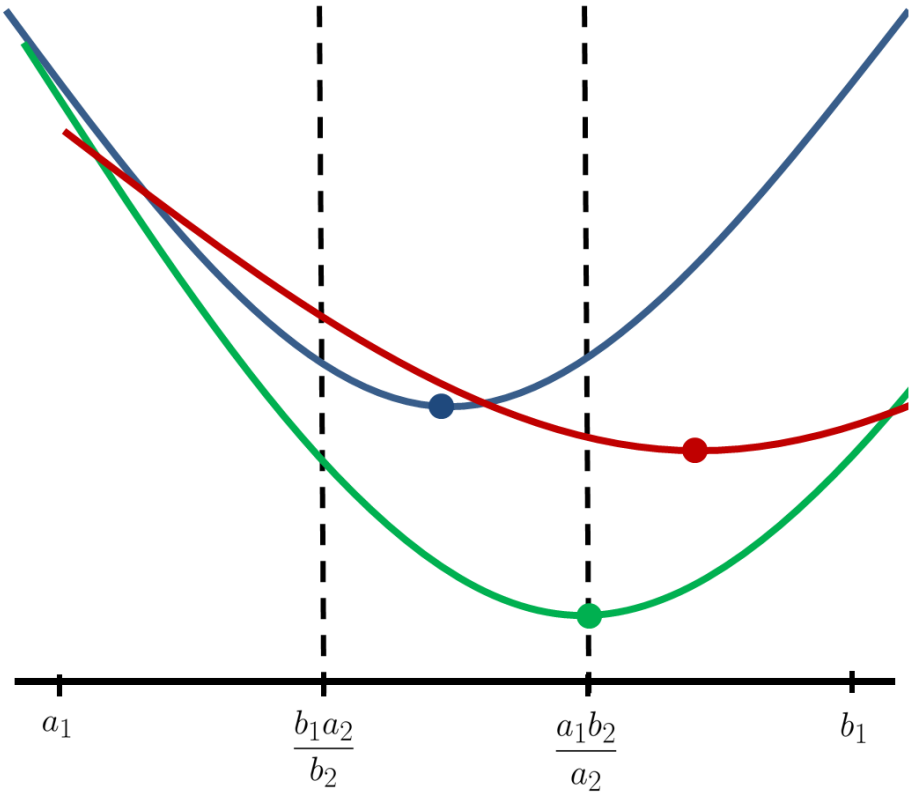


Figure 5.4: Picture to illustrate the possible outcomes of Algorithm 1 in Case 1

It is important to note that each of the cases in Algorithm 1 actually *can* occur. Unfortunately, the plots of the 'real' functions do not display the key details as clearly as our illustration, so we do not include them here. However, 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 ($a_1 = 1$, $b_1 = 34$, $a_2 = 2$, $b_2 = 35$, $a_3 = 12$, $b_3 = 35$).
- An example of a green curve (minimum occurs at $\frac{a_1 b_2}{a_2}$) is ($a_1 = 1$, $b_1 = 8$, $a_2 = 5$, $b_2 = 22$, $a_3 = 1$, $b_3 = 4$).

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 $\frac{b_1 a_2}{b_2}$ 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 5.5 we can see the plot of this function and the minimum, which falls at the midpoint. In Case 0 we always have $q_1 = q_2 = q_3 = \frac{a_1 + b_1}{2}$.

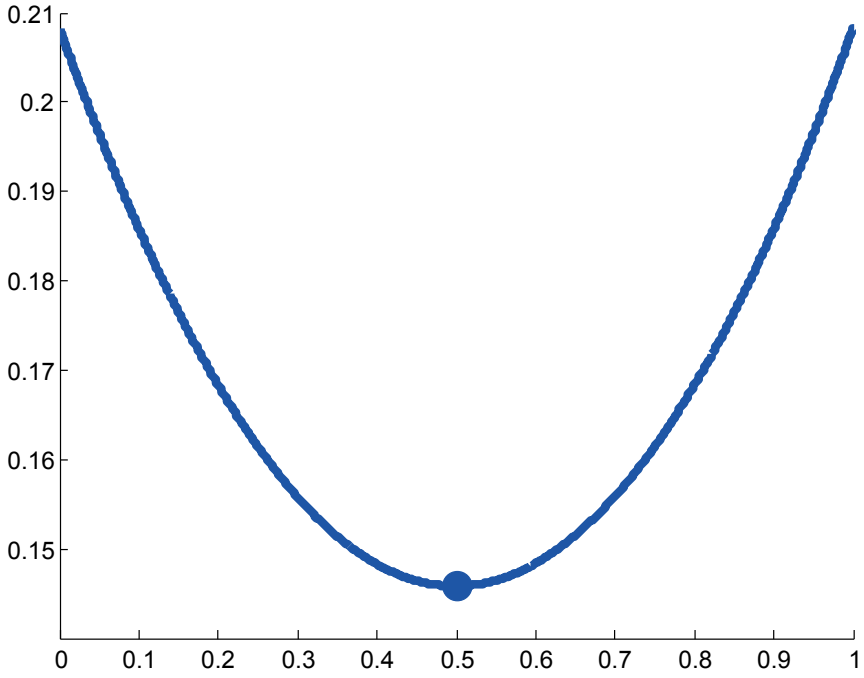


Figure 5.5: Plot of the total volume function for parameter values: $(a_1 = 0, b_1 = 1, a_2 = 0, b_2 = 1, a_3 = 0, b_3 = 1)$

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

We have seen that each piece of $TV(c_1)$ and $\widehat{TV}(c_1)$ is a convex quadratic function. However, this does not imply that the functions are convex over the whole domain,

$[a_1, b_1]$, and in fact, we hinted at the possibility of non-convexity in our sketch of Figure 5.3. Nevertheless, as we show in the following theorem, with a bit more work, we are able to demonstrate that $TV(c_1)$ and $\widehat{TV}(c_1)$ *are* convex over the domain, $[a_1, b_1]$. Therefore, a more appropriate picture for Figure 5.3 would be the illustration in Figure 5.6.

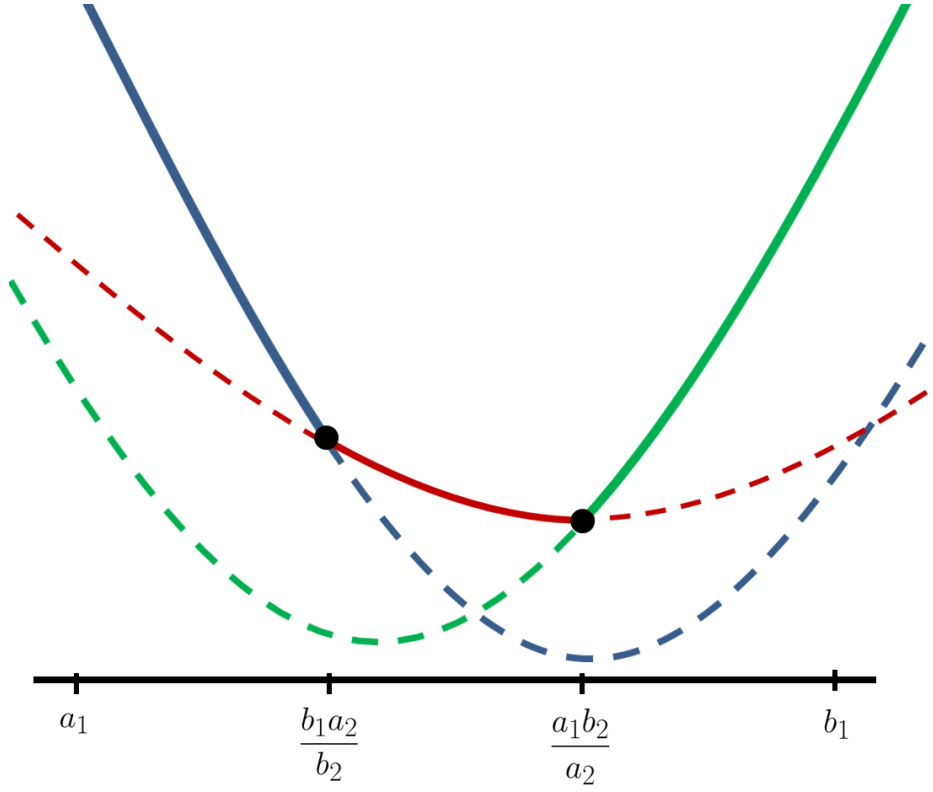


Figure 5.6: Illustration of a globally convex piecewise-quadratic function

Theorem 5.1. Given that the upper and lower bound parameters respect the labeling Ω, the functions $TV(c_1)$ and $\widehat{TV}(c_1)$ 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 $\widehat{TV}(c_1)$ 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 $\frac{b_1 a_2}{b_2}$:

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

Secondly, we compare the first derivatives of V_2 and V_3 at breakpoint $\frac{a_1 b_2}{a_2}$:

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

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 $\widehat{TV}(c_1)$: First, we compare the first derivatives of V_1 and V_4 at the breakpoint $\frac{a_1 b_2}{a_2}$:

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

Secondly, we compare the first derivatives of V_4 and V_3 at the breakpoint $\frac{b_1 a_2}{b_2}$:

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

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

5.3.1.7 Bounds on where the optimal branching point can occur

We have seen in §5.2 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 this minimizer can fall close to either of the endpoints. We want to know how likely it is that solvers are routinely precluding the "best" branching point. The following propositions give some insight to this issue and show that, in fact, software is unlikely to be cutting off the optimal branching point.

Proposition 5.2. The branching point for variable x_1 that obtains the least 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 $\frac{a_1 b_2}{a_2} \geq \frac{b_1 a_2}{b_2}$, then we are in Case 1. If $q_3 \geq \frac{a_1 b_2}{a_2}$, then q_3 is the minimizer, but we know that $q_3 \leq \frac{a_1+b_1}{2}$ (see 5.4). 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) $\frac{a_1 b_2}{a_2}$ is the minimizer, and (ii) it must be that either that $\frac{a_1+b_1}{2} > \frac{a_1 b_2}{a_2}$, in which case our claim is valid, or $\frac{a_1+b_1}{2} < \frac{b_1 a_2}{b_2} \leq \frac{a_1 b_2}{a_2}$. We will show by contradiction that this cannot be the case.

For contradiction 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:

$$\begin{aligned} 2b_1 a_2 - b_1 b_2 - a_1 b_2 = b_1(a_2 - b_2) + (b_1 a_2 - a_1 b_2) > 0, \quad \text{and} \\ 2a_1 b_2 - a_1 a_2 - b_1 a_2 = a_1(b_2 - a_2) + (a_1 b_2 - b_1 a_2) > 0. \end{aligned}$$

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). 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 $q_3 \leq \frac{a_1+b_1}{2}$ (see 5.4). 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 $\frac{a_1+b_1}{2} > \frac{b_1 a_2}{b_2}$, in which case our claim is valid, or $\frac{a_1+b_1}{2} < \frac{a_1 b_2}{a_2} < \frac{b_1 a_2}{b_2}$. 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. □

This proposition 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 evidence to back up software's current choice of branching point, as discussed in §5.2. The following proposition 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).

Proposition 5.3. Given upper and lower bound parameters (a_1, b_1, a_2, b_2, a_3, b_3) satisfying Ω, and $a_2 \neq 0$. The branching point for variable x_1 that obtains the least 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

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

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:

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

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

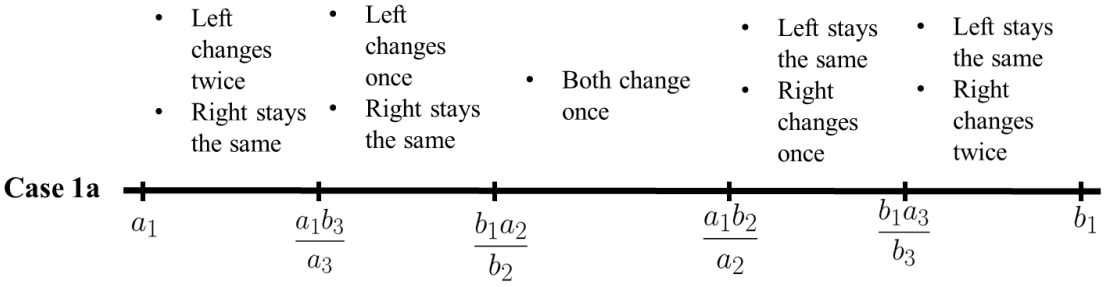
□

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 ϵ, and is therefore not particularly informative, given that we can make ϵ as close to zero as we wish. However, we have checked millions of examples and are yet to find an example where the minimizer occurs less than ~ 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.

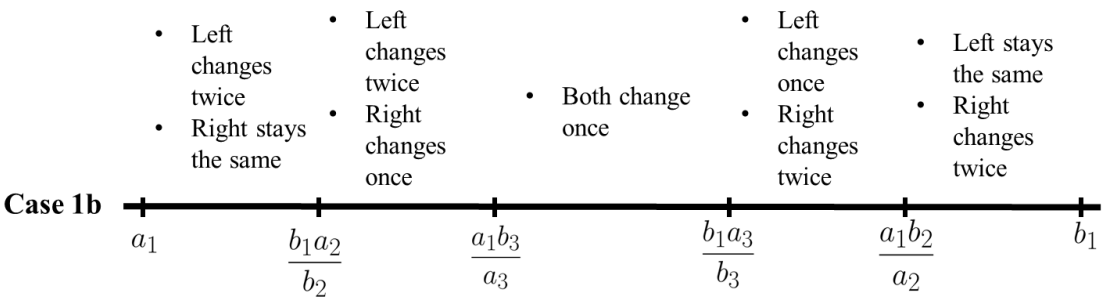
5.3.2 The best double-McCormick convexification

Unlike `ANTIGONE` ([37]) and `BARON` ([49]), some software does not use the explicit convex hull for trilinear monomials, but instead employs repeated McCormick convexifications to obtain a relaxation; see `SCIP` ([61]) and `Couenne` ([6]). Here, we describe some partial branching-point analysis for the best double-McCormick relaxation $\mathcal{P}_3$ (the relaxation with the least volume). We do note that currently `SCIP` and `Couenne` do not always use $\mathcal{P}_3$; rather, their choice of which double McCormick is arbitrary. However, because of our results from Chapter 3, we choose to focus on $\mathcal{P}_3$. As we did in §5.3.1, we assume throughout that we branch on variable x_1.

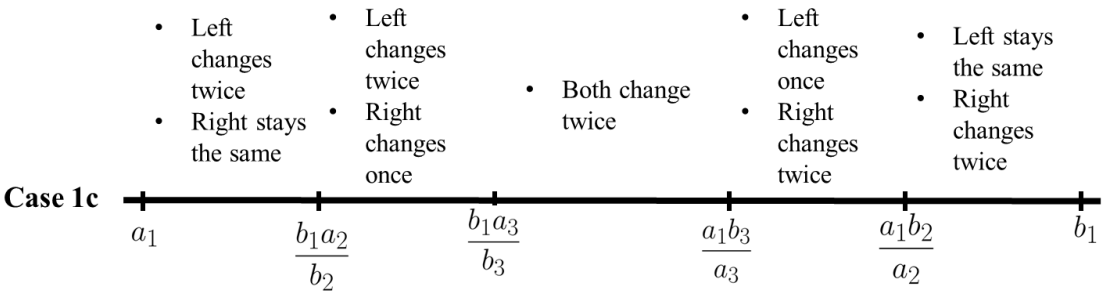
When analyzing the branching point for the double-McCormick convexification, it is important to note that variables x_2 and x_3 are *not* interchangeable. This is because of the structure of the volume function of $\mathcal{P}_3$ (see Theorem 3.4). This means that we do need to consider the what happens when the bounds of x_1 vary enough for x_1 to be relabeled as x_3, and the case analysis becomes more complicated than in the convex-hull analysis. The original Case 1 (recall Figure 5.2) splits into three further cases, while Case 2 remains as only one case. The number of pieces of the piecewise functions increase for *all* cases. See Figure 5.7 for an illustration (as before, the $a_3 = 0$ and/or $a_2 = 0$ cases will need to be handled separately).



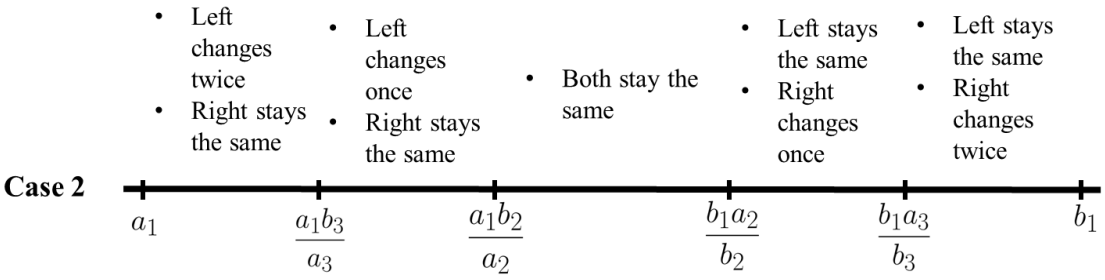
(a) Case 1a



(b) Case 1b



(c) Case 1c



(d) Case 2

Figure 5.7: Case analysis for $\mathcal{P}_3$

It is clear that the for the double-McCormick convexification, even when only considering the case analysis, the situation is much more complex. We would like to perform the same analysis that we did for the convex-hull case. Namely: construct the appropriate piecewise functions for each case; establish the convexity of each piece (for each case); find a closed form expression for the global minimizer of each piece. Using this information, we would then be able to specify an algorithm to: (i) check if the global minimizer falls in the appropriate segment of the interval, and if not, check the end points to find the minimizer over that segment and then (ii) compare the minimums of each segment to find the global minimizer. This could be done for each case.

However, because in this section we present only a partial analysis, we will focus on the segment where the branching point is such that *none* of the variables need to be relabeled. This is the middle segment of Case 2. We define the relevant function for this segment of the interval, and note that it is no longer a quadratic. However, we are able to establish that it is convex, and therefore for a given problem we will be able to efficiently calculate the minimizer (and minimum). Unfortunately, in general, we are unable to get a closed-form expression for the minimizer, but with a bit of work, we are able to obtain a rather small window for where the minimum will occur.

To begin, we define

$$\begin{aligned} V_{DM}(l_1,u_1,l_2,u_2,l_3,u_3) := &\Bigg((u_1-l_1)(u_2-l_2)(u_3-l_3) \\ &\times \frac{(u_1(5u_2u_3-l_2u_3-u_2l_3-3l_2l_3)+l_1(5l_2l_3-u_2l_3-l_2u_3-3u_2u_3))}{24} \Bigg) \\ &+ \frac{(u_1-l_1)(u_2-l_2)^2(u_3-l_3)^2\left(5(l_1u_1u_2-l_1u_1l_2)+3(u_1^2l_2-l_1^2u_2)\right)}{24(u_1u_2-l_1l_2)}, \end{aligned}$$

to be the volume of the relaxation $\mathcal{P}_3$, with variable lower bounds l_i and upper bounds, u_i, for $i = 1 \dots 3$.

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 subproblems after branching at point c_1 (given that we have $\frac{a_1b_2}{a_2} \leq c_1 \leq \frac{b_1a_2}{b_2}$), is described by the following parameterized function:

$$TV_{DM}(c_1) = V_{DM}(a_1,c_1,a_2,b_2,a_3,b_3) + V_{DM}(c_1,b_1,a_2,b_2,a_3,b_3).$$

This function is *not* a quadratic, but it is convex. See Theorem 5.10 (§5.5) for the proof of convexity.

As noted, we are unable to obtain a closed form expression for the minimizer of this function, but we are able to obtain a small window for where the minimum will occur. In particular, we establish that the minimizer must be strictly greater than the midpoint. For this proof see Proposition 5.11 (§5.5). We also establish that the furthest right this minimizer can occur is $1/\sqrt{3} \approx 0.5773503$ of the way through the interval. For this proof see Theorem 5.12 (§5.5).

Furthermore, we can find examples where the minimum occurs very close to the midpoint of the interval. For example, with $a_i = 1000$, $b_i = 1001$, for all i, the minimum occurs at $c_1 \approx 1000.500008$. We can also find examples where the minimum occurs exactly $1/\sqrt{3}$ of the way through the interval. For the special case of $a_i = 0$ and $b_i = 1$ for $i = 1, 2, 3$, we find that the minimum of the function is obtained when branching at $c_1 = 1/\sqrt{3} \approx 0.5773503$. Therefore, our result that the minimizer occurs somewhere between 0.5 of the way and $1/\sqrt{3} \approx 0.5773503$ of the way through the interval is essentially sharp.

By identifying and analyzing the appropriate piecewise functions for the other possible cases described by Figure 5.7, we expect to be able to obtain an algorithm, in the spirit of Algorithm 1, that would output the optimal branching point for x_1 when using the relaxation $\mathcal{P}_3$.

5.4 Concluding remarks and future work

In this section we have produced a simple algorithm for obtaining the optimal branching point when using the convex-hull convexification 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.

We have begun the analysis for obtaining the optimal branching point when using the best double-McCormick convexification and branching on variable x_1, but this case is far more complex. Alongside completing the double-McCormick analysis for x_1, it would be a natural next step to obtain results for branching on x_2 or x_3 using either convexification.

5.5 Technical propositions, lemmas, and theorems

In this section we state the propositions, lemmas, and theorems used for the analysis, along with their proofs.

5.5.1 Convex-hull convexification

Proposition 5.4. Given that the upper and lower bound parameters respect the labeling Ω, 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 $V_2\left(\frac{a_1 b_2}{a_2}\right) = V_3\left(\frac{a_1 b_2}{a_2}\right)$.

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

where

$$\begin{aligned}
p = &\Big(-3a_2 a_3 - a_2 b_3 + b_2 a_3 + 3b_2 b_3\Big)\times \\
&\Big(-3a_2^3 a_3 - a_2^3 b_3 + 13 a_2^2 b_2 a_3 + 7 a_2^2 b_2 b_3 - 12 a_2 b_2^2 a_3 - 20 a_2 b_2^2 b_3 + 16 b_2^3 b_3\Big) \\
&= \Big(-3a_2 a_3 - a_2 b_3 + b_2 a_3 + 3b_2 b_3\Big)\times \\
&\Big((-3a_2^3 + 13 a_2^2 b_2 - 12 a_2 b_2^2) a_3 + (-a_2^3 + 7 a_2^2 b_2 - 20 a_2 b_2^2 + 16 b_2^3) b_3\Big),
\end{aligned}$$

$$q = 4a_2 b_1 (2a_2^2 a_3 - 3a_2 b_2 a_3 - 3a_2 b_2 b_3 + 4 b_2^2 b_3)(3a_2 a_3 + a_2 b_3 - b_2 a_3 - 3b_2 b_3),$$

$$r = 4a_2^2 b_1^2 (a_2 a_3 + a_2 b_3 - 2b_2 b_3)^2.$$

To show that $V_2\left(\frac{a_1 b_2}{a_2}\right) - V_1(q_1)$ is non-negative for all parameters satisfying Ω, we will show that $p a_1^2 + q a_1 + r \geq 0$ for all parameters satisfying Ω.

We observe:

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

where

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

and

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

Therefore, by Lemma 3.12 we have that $b_3Y + a_3Z$ is non-negative and so p is non-negative. 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

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

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

$$\frac{4a_2^2b_1^2(b_3 - a_3)(b_2 - a_2)^3(3a_2a_3 + a_2b_3 - 4b_2b_3)^2}{(-3a_2^3a_3 - a_2^3b_3 + 13a_2^2b_2a_3 + 7a_2^2b_2b_3 - 12a_2b_2^2a_3 - 20a_2b_2^2b_3 + 16b_2^3b_3)}.$$

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 Ω. Therefore $pa_1^2 + qa_1 + r \geq 0$, and consequently, $V_2\left(\frac{a_1b_2}{a_2}\right) - V_1(q_1) \geq 0$ as required.

□

Lemma 5.5. Given that the upper and lower bound parameters respect the labeling Ω, and $\frac{b_1a_2}{b_2} \leq \frac{a_1b_2}{a_2}$,

$$V_1\left(\frac{b_1a_2}{b_2}\right) = V_2\left(\frac{b_1a_2}{b_2}\right) \geq V_2\left(\frac{a_1b_2}{a_2}\right) = V_3\left(\frac{a_1b_2}{a_2}\right)$$

Proof. It is easy to check that $V_1\left(\frac{b_1a_2}{b_2}\right) = V_2\left(\frac{b_1a_2}{b_2}\right)$ and $V_2\left(\frac{a_1b_2}{a_2}\right) = V_3\left(\frac{a_1b_2}{a_2}\right)$.

Furthermore,

$$\begin{aligned}
&V_2\left(\frac{b_1a_2}{b_2}\right) - V_2\left(\frac{a_1b_2}{a_2}\right) \\
&= \frac{(b_3 - a_3)(b_2 - a_2)^2(b_1a_2 - a_1b_2)(a_1b_2^2 - a_2^2b_1)(-3a_2a_3 - a_2b_3 + b_2a_3 + 3b_2b_3)}{12a_2^2b_2^2} \\
&\geq 0,
\end{aligned}$$

as required.

□

Proposition 5.6. Given that the upper and lower bound parameters respect the labeling Ω, and $\frac{b_1a_2}{b_2} > \frac{a_1b_2}{a_2}$,

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

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

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

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

$$r = 16b_1^2(-3a_2^3a_3-a_2^3b_3+3a_2^2b_2a_3+5a_2^2b_2b_3-a_2b_2^2a_3-4a_2b_2^2b_3+b_2^3b_3)(b_2b_3-a_2a_3).$$

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

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:

$$\frac{16b_1^2(b_3-a_3)(b_2-a_2)^3(b_2b_3-a_2a_3)(3a_2a_3+a_2b_3-4b_2b_3)^2}{(3a_2a_3+a_2b_3+b_2a_3-5b_2b_3)^2},$$

which is non-negative for all parameters satisfying Ω. 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.

□

Lemma 5.7. Given that the upper and lower bound parameters respect the labeling Ω, 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 $V_1\left(\frac{a_1b_2}{a_2}\right) = V_4\left(\frac{a_1b_2}{a_2}\right)$ and $V_4\left(\frac{b_1a_2}{b_2}\right) = V_3\left(\frac{b_1a_2}{b_2}\right)$.

Furthermore,

$$V_4\left(\frac{a_1b_2}{a_2}\right) - V_4\left(\frac{b_1a_2}{b_2}\right) = \frac{(b_3-a_3)(b_2-a_2)^2(b_1a_2^2 - a_1b_2^2)(b_1a_2 - a_1b_2)(b_2b_3 - a_2a_3)}{3a_2^2b_2^2} \geq 0,$$

as required.

□

Lemma 5.8. Given that the parameters satisfy the conditions Ω, and furthermore, $\frac{b_1a_2}{b_2} \leq \frac{a_1b_2}{a_2}$, we have

$$q_1 \geq \frac{b_1a_2}{b_2}.$$

Proof. From the proof of Proposition 5.2, we know that the midpoint, q_2, cannot be smaller than *both* $\frac{a_1b_2}{b_1}$ and $\frac{b_1a_2}{b_2}$. Therefore we have:

$$q_2 \geq \min\left\{\frac{a_1b_2}{b_1}, \frac{b_1a_2}{b_2}\right\},$$

and because we saw in 5.5 that $q_1 \geq q_2$ we also have

$$q_1 \geq \min\left\{\frac{a_1b_2}{b_1}, \frac{b_1a_2}{b_2}\right\}.$$

Therefore, under the conditions of the lemma, $q_1 \geq \frac{b_1a_2}{b_2}$ as required.

□

Lemma 5.9. Given that the parameters satisfy the conditions Ω, and furthermore, $\frac{b_1a_2}{b_2} \geq \frac{a_1b_2}{a_2}$, we have

$$q_1 \geq \frac{a_1b_2}{a_2}.$$

Proof. We saw in the proof of Lemma 5.8 that

$$q_1 \geq \min\left\{\frac{a_1b_2}{b_1}, \frac{b_1a_2}{b_2}\right\}.$$

Therefore, under the conditions of the lemma, $q_1 \geq \frac{a_1b_2}{a_2}$ as required.

□

5.5.2 Double-McCormick convexification

Theorem 5.10. Given that the upper and lower bound parameters respect the labeling Ω, the total volume function:

$$TV_{DM}(c_1) = V_{DM}(a_1, c_1, a_2, b_2, a_3, b_3) + V_{DM}(c_1, b_1, a_2, b_2, a_3, b_3)$$

is a convex function in the branching point c_1, on the domain $[a_1, b_1]$.

Proof. It is natural to try and show that each summand in the theorem statement is convex in c_1 on the domain $[a_1, b_1]$; but this is not the case. Instead, we let

$$V_{DM}(c_1, b_1, a_2, b_2, a_3, b_3) =: s_1 + s_2,$$

and

$$V_{DM}(a_1, c_1, a_2, b_2, a_3, b_3) =: s_3 + s_4,$$

where:

$$s_1 := (b_1 - c_1)(b_2 - a_2)(b_3 - a_3) \times \frac{b_1(5b_2b_3 - a_2b_3 - b_2a_3 - 3a_2a_3) + c_1(5a_2a_3 - b_2a_3 - a_2b_3 - 3b_2b_3)}{24},$$

$$s_2 := \frac{(b_1 - c_1)(b_2 - a_2)^2(b_3 - a_3)^2(3a_2b_1^2 - 5a_2b_1c_1 + 5b_1b_2c_1 - 3b_2c_1^2)}{24(b_1b_2 - c_1a_2)},$$

$$s_3 := (c_1 - a_1)(b_2 - a_2)(b_3 - a_3) \times \frac{c_1(5b_2b_3 - a_2b_3 - b_2a_3 - 3a_2a_3) + a_1(5a_2a_3 - b_2a_3 - a_2b_3 - 3b_2b_3)}{24},$$

$$s_4 := \frac{(c_1 - a_1)(b_2 - a_2)^2(b_3 - a_3)^2(-3a_1^2b_2 - 5a_1a_2c_1 + 5a_1b_2c_1 + 3a_2c_1^2)}{24(b_2c_1 - a_1a_2)}.$$

Now, to show that $TV_{DM}(c_1) = V_{DM}(a_1, c_1, a_2, b_2, a_3, b_3) + V_{DM}(c_1, b_1, a_2, b_2, a_3, b_3)$ is convex in c_1, we will show that $s_1 + s_2 + s_3$ is convex in c_1, and that s_4 is convex in c_1, both over the domain $[a_1, b_1]$.

Proof of convexity of $s_1 + s_2 + s_3$

Taking the second derivative of $s_1 + s_2 + s_3$ with respect to c_1 we obtain:

$$\frac{(b_2-a_2)(b_3-a_3)}{12(b_1b_2-c_1a_2)^3}\times\big((8a_2^4a_3+3a_2^3a_3b_2-11a_2^3b_2b_3-3a_2^2a_3b_2^2+3a_2^2b_2^2b_3)c_1^3+(-24a_2^3a_3b_1b_2-9a_2^2a_3b_1b_2^2+33a_2^2b_1b_2^2b_3+9a_2a_3b_1b_2^3-9a_2b_1b_2^3b_3)c_1^2+(24a_2^2a_3b_1^2b_2^2+9a_2a_3b_1^2b_2^3-33a_2b_1^2b_2^3b_3-9a_3b_1^2b_2^4+9b_1^2b_2^4b_3)c_1+3a_2^4a_3b_1^3-3a_2^4b_1^3b_3-11a_2^3a_3b_1^3b_2+11a_2^3b_1^3b_2b_3+18a_2^2a_3b_1^3b_2^2-18a_2^2b_1^3b_2^2b_3-26a_2a_3b_1^3b_2^3+18a_2b_1^3b_2^3b_3+8a_3b_1^3b_2^4\big).$$

It is clear that the first multiplicand is non-negative for all $c_1 \in [a_1, b_1]$ (and parameters satisfying Ω). The second multiplicand is a cubic in c_1. By taking the first derivative of this cubic and setting it equal to zero, we can solve for c_1 and obtain the stationary points. In doing so, we find that there is one repeated root and therefore one stationary point at $c_1 = b_1b_2/a_2$. Because this cubic has only one stationary point, we know that it is monotone in c_1.

Consider substituting $c_1 = b_1$ into this cubic. In doing so, we obtain the quantity

$$b_1^3(b_2-a_2)^3\left((b_2+3a_2)(b_3-a_3)+8(b_2b_3-a_2a_3)\right),$$

which is non-negative for all values of the parameters satisfying Ω.

Substituting $c_1 = 0$ into this cubic, we obtain

$$b_1^3\bigg(b_3(11a_2^3b_2+18a_2b_2^3-18a_2^2b_2^2-3a_2^4)+a_3(3a_2^4+8b_2^4+18a_2^2b_2^2-26a_2b_2^3-11a_2^3b_2)\bigg). \quad (5.7)$$

We use Lemma 3.12, with:

$$A = b_3,$$
$$B = a_3,$$
$$C = 11a_2^3b_2 + 18a_2b_2^3 - 18a_2^2b_2^2 - 3a_2^4 = (18a_2b_2^2 + 3a_2^3)(b_2 - a_2) + 8a_2^3b_2, \text{ and}$$
$$D = 3a_2^4 + 8b_2^4 + 18a_2^2b_2^2 - 26a_2b_2^3 - 11a_2^3b_2,$$

which implies: $C + D = 8b_2^3(b_2 - a_2) \geq 0$, to see that 5.7 is non-negative for all values of the parameters satisfying Ω.

Because the cubic is monotone, we conclude that it is non-negative for all $c_1 \in [0, b_1]$ and therefore for all $c_1 \in [a_1, b_1]$. Consequently, the second derivative of $s_1 + s_2 + s_3$ with respect to c_1 is non-negative on the domain $c_1 \in [a_1, b_1]$, and $s_1 + s_2 + s_3$ is convex.

Proof of convexity of s_4

Taking the second derivative of s_4 with respect to c_1 we obtain

$$\frac{(b_2-a_2)^2(b_3-a_3)^2}{12(c_1b_2-a_1a_2)^3}\times$$
$$\left((3a_2b_2^2)c_1^3-(9a_1a_2^2b_2)c_1^2+(9a_1^2a_2^3)c_1-8a_1^3a_2^3+10a_1^3a_2^2b_2-8a_1^3a_2b_2^2+3a_1^3b_2^3\right).$$

It is clear that the first multiplicand is non-negative for all $c_1 \in [a_1, b_1]$. The second multiplicand is a cubic in c_1; by taking the first derivative of this cubic and setting it equal to zero, we can solve for c_1 and obtain the stationary points. In doing so, we find that there is one repeated root and therefore one stationary point at $c_1 = a_1a_2/b_2$. Because this cubic has only one stationary point, we know that it is monotone in c_1. By considering the leading term of this cubic and seeing that it is non-negative, we know that this function is non-decreasing.

Substituting $c_1 = a_1$ into this cubic, we obtain

$$a_1^3(3b_2+a_2)(b_2-a_2)^2,$$

which is non-negative for all values of the parameters satisfying Ω.

Because the cubic is non-decreasing, we conclude that it is non-negative for all $c_1 \in [a_1, \infty)$ and therefore for all $c_1 \in [a_1, b_1]$. Consequently, the second derivative of s_4 with respect to c_1 is non-negative over the domain $[a_1, b_1]$, and therefore s_4 is convex.

Consequently $(s_1+s_2+s_3)+s_4 = V_{DM}(a_1, c_1, a_2, b_2, a_3, b_3)+V_{DM}(c_1, b_1, a_2, b_2, a_3, b_3)$ is a convex function in c_1 over the domain $[a_1, b_1]$ as required. □

Proposition 5.11. Given that the upper and lower bound parameters respect the labeling Ω, the minimum of the convex function

$$TV_{DM}(c_1) = V_{DM}(a_1, c_1, a_2, b_2, a_3, b_3) + V_{DM}(c_1, b_1, a_2, b_2, a_3, b_3)$$

over the domain $c_1 \in [a_1, b_1]$, occurs at some value of $c_1 > (a_1 + b_1)/2$.

Proof. Consider the value of the first derivative at the midpoint $(a_1 + b_1)/2$:

$$\frac{(b_1-a_1)^3(b_2-a_2)^3(b_3-a_3)^2(3a_2^2-2a_2b_2+3b_2^2)((a_1^2a_2-b_1^2b_2)+3a_1b_1(a_2-b_2))}{24(b_1b_2+a_1b_2-2a_1a_2)^2(2b_1b_2-a_1a_2-b_1a_2)^2}.$$

We can see that for all parameters satisfying Ω this quantity is negative (we can claim negativity and not just non-positivity because we require $b_i > a_i$). Therefore at the midpoint the convex function is still decreasing, and hence the minimum occurs when $c_1 > (a_1 + b_1)/2$.

□

Theorem 5.12. Given that the upper and lower bound parameters respect the labeling Ω, the minimum of the convex function

$$TV_{DM}(c_1) = V_{DM}(a_1, c_1, a_2, b_2, a_3, b_3) + V_{DM}(c_1, b_1, a_2, b_2, a_3, b_3),$$

over the domain $c_1 \in [a_1, b_1]$, occurs at some value of $c_1 \leq a_1 + (b_1 - a_1)/\sqrt{3}$. Moreover, this interval cannot be tightened any further from the right.

Proof. Consider the value of the first derivative at the point $c_1 = a_1 + (b_1 - a_1)/\sqrt{3}$:

$$\frac{(2\sqrt{3} - 3)(b_1 - a_1)(b_2 - a_2)(b_3 - a_3) \times g}{12(b_2 a_1\sqrt{3} - b_2\sqrt{3}b_1 + 3a_1a_2 - 3b_2a_1)^2(a_2\sqrt{3}b_1 - b_2\sqrt{3}b_1 + 2a_1a_2 + a_2b_1 - 3b_1b_2)^2}, \tag{5.8}$$

where g is a complicated function of the parameters a_1, b_1, a_2, b_2, a_3, b_3.

We now think of g as a polynomial in a_2, parameterized by a_1, b_1, b_2, a_3 and b_3.

$$g(a_2) = \gamma_5 a_2^5 + \gamma_4 a_2^4 + \gamma_3 a_2^3 + \gamma_2 a_2^2 + \gamma_1 a_2 + \gamma_0,$$

where:

$$\begin{aligned}
\gamma_5 := \Big(& -54\sqrt{3}a_1^3a_3b_1 + 6\sqrt{3}a_1^3b_1b_3 + 9\sqrt{3}a_1^2a_3b_1^2 - 33\sqrt{3}a_1^2b_1^2b_3 \\
& - 63\sqrt{3}a_1a_3b_1^3 + 63\sqrt{3}a_1b_1^3b_3 - 54a_1^4a_3 + 6a_1^4b_3 - 54a_1^3a_3b_1 \\
& + 6a_1^3b_1b_3 - 48a_1^2b_1^2b_3 - 108a_1a_3b_1^3 + 108a_1b_1^3b_3\Big), \\
\gamma_4 := \Big(& -24\sqrt{3}a_1^4a_3b_2 - 8\sqrt{3}a_1^4b_2b_3 + 120\sqrt{3}a_1^3a_3b_1b_2 + 72\sqrt{3}a_1^3b_1b_2b_3 \\
& + 3\sqrt{3}a_1^2a_3b_1^2b_2 + 165\sqrt{3}a_1^2b_1^2b_2b_3 + 273\sqrt{3}a_1a_3b_1^3b_2 - 241\sqrt{3}a_1b_1^3b_2b_3 \\
& + 24\sqrt{3}a_3b_1^4b_2 - 24\sqrt{3}b_1^4b_2b_3 + 78a_1^4a_3b_2 + 66a_1^4b_2b_3 \\
& + 156a_1^3a_3b_1b_2 + 36a_1^3b_1b_2b_3 + 54a_1^2a_3b_1^2b_2 + 282a_1^2b_1^2b_2b_3 \\
& + 462a_1a_3b_1^3b_2 - 414a_1b_1^3b_2b_3 + 42a_3b_1^4b_2 - 42b_1^4b_2b_3\Big),
\end{aligned}$$

$$\begin{aligned}
\gamma_3 := \Big(& -8\sqrt{3}a_1^4a_3b_2^2 + 72\sqrt{3}a_1^4b_2^2b_3 + 40\sqrt{3}a_1^3a_3b_1b_2^2 - 248\sqrt{3}a_1^3b_1b_2^2b_3 \\
& + 78\sqrt{3}a_1^2a_3b_1^2b_2^2 - 486\sqrt{3}a_1^2b_1^2b_2^2b_3 - 526\sqrt{3}a_1a_3b_1^3b_2^2 \\
& + 366\sqrt{3}a_1b_1^3b_2^2b_3 - 88\sqrt{3}a_3b_1^4b_2^2 + 80\sqrt{3}b_1^4b_2^2b_3 + 14a_1^4a_3b_2^2 \\
& - 174a_1^4b_2^2b_3 - 128a_1^3a_3b_1b_2^2 - 144a_1^3b_1b_2^2b_3 + 168a_1^2a_3b_1^2b_2^2 - 888a_1^2b_1^2b_2^2b_3 \\
& - 908a_1a_3b_1^3b_2^2 + 636a_1b_1^3b_2^2b_3 - 154a_3b_1^4b_2^2 + 138b_1^4b_2^2b_3 \Big), \\
\gamma_2 := \Big(& 56\sqrt{3}a_1^4a_3b_2^3 - 88\sqrt{3}a_1^4b_2^3b_3 - 108\sqrt{3}a_1^3a_3b_1b_2^3 + 140\sqrt{3}a_1^3b_1b_2^3b_3 \\
& - 366\sqrt{3}a_1^2a_3b_1^2b_2^3 + 774\sqrt{3}a_1^2b_1^2b_2^3b_3 + 490\sqrt{3}a_1a_3b_1^3b_2^3 - 218\sqrt{3}a_1b_1^3b_2^3b_3 \\
& + 144\sqrt{3}a_3b_1^4b_2^3 - 104\sqrt{3}b_1^4b_2^3b_3 - 78a_1^4a_3b_2^3 + 142a_1^4b_2^3b_3 \\
& + 84a_1^3a_3b_1b_2^3 + 140a_1^3b_1b_2^3b_3 - 660a_1^2a_3b_1^2b_2^3 + 1236a_1^2b_1^2b_2^3b_3 \\
& + 840a_1a_3b_1^3b_2^3 - 328a_1b_1^3b_2^3b_3 + 246a_3b_1^4b_2^3 - 182b_1^4b_2^3b_3 \Big), \\
\gamma_1 := \Big(& -24\sqrt{3}a_1^4a_3b_2^4 + 24\sqrt{3}a_1^4b_2^4b_3 - 34\sqrt{3}a_1^3a_3b_1b_2^4 + 66\sqrt{3}a_1^3b_1b_2^4b_3 \\
& + 393\sqrt{3}a_1^2a_3b_1^2b_2^4 - 537\sqrt{3}a_1^2b_1^2b_2^4b_3 - 195\sqrt{3}a_1a_3b_1^3b_2^4 + 3\sqrt{3}a_1b_1^3b_2^4b_3 \\
& - 104\sqrt{3}a_3b_1^4b_2^4 + 48\sqrt{3}b_1^4b_2^4b_3 + 24a_1^4a_3b_2^4 - 24a_1^4b_2^4b_3 \\
& + 6a_1^3a_3b_1b_2^4 - 102a_1^3b_1b_2^4b_3 + 456a_1^2a_3b_1^2b_2^4 - 648a_1^2b_1^2b_2^4b_3 \\
& - 216a_1a_3b_1^3b_2^4 - 120a_1b_1^3b_2^4b_3 - 198a_3b_1^4b_2^4 + 102b_1^4b_2^4b_3 \Big), \\
\gamma_0 := \Big(& 36a_3b_2^5a_1^3b_1\sqrt{3} - 36b_2^5b_3a_1^3b_1\sqrt{3} - 117a_3b_2^5a_1^2b_1^2\sqrt{3} + 117b_2^5b_3a_1^2b_1^2\sqrt{3} \\
& + 21a_3b_2^5a_1b_1^3\sqrt{3} + 27b_2^5b_3a_1b_1^3\sqrt{3} + 24a_3b_2^5b_1^4\sqrt{3} - 114a_3b_2^5a_1^2b_1^2 \\
& + 162b_2^5b_3a_1^2b_1^2 - 6a_3b_2^5a_1b_1^3 + 54b_2^5b_3a_1b_1^3 + 48a_3b_2^5b_1^4 \Big).
\end{aligned}$$

It is clear from 5.8 that the derivative at the point $c_1 = a_1 + (b_1 - a_1)/\sqrt{3}$ is non-negative if and only if g is non-negative. We will show that for all values of the parameters satisfying Ω, we have $g(a_2) \geq 0$ on the interval $[0, b_2]$.

To do this, we first construct a mapping of the interval $[0, b_2]$ to the interval $[0, \infty)$, (see [47]). This mapping results in a second polynomial in a_2 (we refer to this as $\hat{g}(a_2)$), also parameterized by a_1, b_1, b_2, a_3 and b_3. Furthermore, if $\hat{g}(a_2) \geq 0$ over the interval $[0, \infty)$ then we know that $g(a_2)$ will be non-negative over $[0, b_2]$. Using Descartes' Rule of Signs, we establish that if each of the *coefficients* in the polynomial

$\hat{g}(a_2)$ are non-negative then there is no real root of $\hat{g}(a_2)$ contained in $(0, \infty)$. In what follows, we show that the coefficients in $\hat{g}(a_2)$ are in fact non-negative and furthermore, $\hat{g}(0)$ is non-negative. Therefore $\hat{g}(a_2)$ is non-negative over $[0, \infty)$, and hence $g(a_2)$ is non-negative over $[0, b_2]$.

As described in [47], the mapping that we require is:

$$\hat{F}(x) := (x+1)^n F\left(\frac{lx+u}{x+1}\right),$$

where n is the degree of the highest term in F and l (respectively u) is the lower (upper) bound on the variable x.

Therefore, in our notation, we consider the mapping

$$\hat{g}(a_2) := (a_2+1)^5 \times g\left(\frac{b_2}{a_2+1}\right).$$

This gives us the polynomial

$$\hat{g}(a_2) = \delta_5 a_2^5 + \delta_4 a_2^4 + \delta_3 a_2^3 + \delta_2 a_2^2 + \delta_1 a_2 + \delta_0,$$

where:

$$\begin{aligned}
\delta_5 &:= \sqrt{3} b_2^5 b_1 \Big(36 a_1 (b_1^2 - a_1^2)(b_3 - a_3) + (38\sqrt{3} + 117) a_1^2 b_1 (b_3 - a_3) + 16\sqrt{3} a_1^2 b_1 b_3 \\
&\qquad + (57 - \sqrt{3}) a_1 b_1^2 a_3 + (18\sqrt{3} - 9) a_1 b_1^2 b_3 + (16\sqrt{3} + 24) b_1^3 a_3 \Big), \\
\delta_4 &:= (\sqrt{3} - 1) b_2^5 \Big(24 a_1^4 (b_3 - a_3) + b_1 \Big((123 + 75\sqrt{3})(b_1^3 - a_1^3)(b_3 - a_3) \\
&\qquad + a_1 (99 + \sqrt{3})(b_1^2 - a_1^2)(b_3 - a_3) + (32\sqrt{3} a_1^2 b_3 + (104\sqrt{3} + 168) b_1^2 a_3)(b_1 - a_1) \\
&\qquad + (a_1^2 b_1 (73\sqrt{3} + 153) + a_1 b_1^2 (63\sqrt{3} - 9))(b_3 - a_3) \\
&\qquad + ((80\sqrt{3} + 192) a_1 b_1)(b_1 b_3 - a_1 a_3) \Big) \Big), \\
\delta_3 &:= \frac{40\sqrt{3} - 18}{1119} b_2^5 (b_1 - a_1) \Big(1119 (b_1^3 - a_1^3)(b_3 - a_3) + (496\sqrt{3} - 672) a_1^2 b_3 (b_1 - a_1) \\
&\qquad + (a_1^2 b_1 (111 + 1490\sqrt{3}) + a_1 b_1^2 (1635 + 152\sqrt{3}) + b_1^3 (138 + 804\sqrt{3}))(b_3 - a_3) \\
&\qquad + (a_1 b_1 (2736 + 112\sqrt{3}) + b_1^2 (2400 + 1852\sqrt{3}))(b_1 b_3 - a_1 a_3) \Big), \\
\delta_2 &:= \frac{76 - 16\sqrt{3}}{313} b_2^5 (b_1 - a_1)^2 \Big((b_1^2 (709 + 314\sqrt{3}) + a_1 b_1 (292\sqrt{3} - 178)
\end{aligned}$$

$$+ a_1^2(369 - 120\sqrt{3}))(b_3 - a_3) + (b_1(428 + 156\sqrt{3}) + a_1(120\sqrt{3} - 56))(b_1 b_3 - a_1 a_3)\Big),$$

$$\delta_1 := 8(\sqrt{3} - 1)b_2^5(b_1 - a_1)^3\Big(((8 + 6\sqrt{3})b_1 + (3 - \sqrt{3})a_1)(b_3 - a_3)$$
$$+ (\sqrt{3} + 1)(b_1 b_3 - a_1 a_3)\Big),$$

$$\delta_0 := 16b_2^5(b_1 - a_1)^4(b_3 - a_3).$$

By assumption, the conditions Ω hold. Because of this, (and the way we have chosen to factor them), it is easy to check that $\delta_i \geq 0$, for $i = 0, \ldots, 5$. Therefore, by Descartes' Rule of Signs, we have that $\hat{g}(a_2)$ has no real root over the interval $[0, \infty)$. Furthermore,

$$\hat{g}(0) = 16b_2^5(b_1 - a_1)^4(b_3 - a_3) \geq 0.$$

Therefore, $\hat{g}(a_2)$ is non-negative over the interval $[0, \infty)$, and consequently, from the definition of our mapping ([47]), g is non-negative over the interval $a_2 \in [0, b_2]$. From this we have that the derivative of $TV_{DM}(c_1) = V_{DM}(a_1, c_1, a_2, b_2, a_3, b_3) + V_{DM}(c_1, b_1, a_2, b_2, a_3, b_3)$ is non-negative at the point $c_1 = a_1 + (b_1 - a_1)/\sqrt{3}$, and therefore the minimizer of this convex function cannot be to the right of this point. Furthermore, we observed in §5.3.2 that when $a_i = 0$ and $b_i = 1$ for all i, the minimum of $TV_{DM}(c_1) = V_{DM}(a_1, c_1, a_2, b_2, a_3, b_3) + V_{DM}(c_1, b_1, a_2, b_2, a_3, b_3)$ occurs exactly at $c_1 = 1/\sqrt{3}$. Therefore this result is sharp. □

BIBLIOGRAPHY

[1] C.S. Adjiman, S. Dallwig, C.A. Floudas, and A. Neumaier. A global optimization method, αBB, for general twice-differentiable constrained NLPs: I. Theoretical advances. *Computers & Chemical Engineering*, 22(9):1137–1158, 1998.

[2] S. Aktürk, A. Atamtürk, and S. Gürel. A strong conic quadratic reformulation for machine-job assignment with controllable processing times. *Operations Research Letters*, 37(3):187–191, 2009.

[3] F. Al-Khayyal and J. Falk. Jointly constrained biconvex programming. *Mathematics of Operations Research*, 8(2):273–286, 1983.

[4] F. Ardila, C. Benedetti, and J. Doker. Matroid polytopes and their volumes. *Discrete & Computational Geometry*, 43(4):841–854, 2010.

[5] X. Bao, A. Khajavirad, N.V. Sahinidis, and M. Tawarmalani. Global optimization of nonconvex problems with multilinear intermediates. *Mathematical Programming Computation*, 7(1):1–37, 2015.

[6] P. Belotti, J. Lee, L. Liberti, F. Margot, and A. Wächter. Branching and bounds tightening techniques for non-convex MINLP. *Optimization Methods & Software*, 24(4-5):597–634, 2009.

[7] R.R. Boorstyn and H. Frank. Large-scale network topological optimization. *IEEE Transactions on Communications*, 25(1):29–47, 1977.

[8] G. Brightwell and P. Winkler. Counting linear extensions is #P-complete. In *Proceedings of the Twenty-third Annual ACM Symposium on Theory of Computing*, STOC '91, pages 175–181, New York, NY, USA, 1991. ACM.

[9] S. Burer and A. N. Letchford. Non-convex mixed-integer nonlinear programming: A survey. *Surveys in Operations Research and Management Science*, 17:97–106, 2012.

[10] K. Burggraf, J. De Loera, and M. Omar. *On Volumes of Permutation Polytopes*, pages 55–77. Springer International Publishing, Heidelberg, 2013.

[11] S. Cafieri, J. Lee, and L. Liberti. On convex relaxations of quadrilinear terms. *Journal of Global Optimization*, 47:661–685, 2010.

[12] A. Costa and L. Liberti. Relaxations of multilinear convex envelopes: Dual is better than primal. In R. Klasing, editor, *Experimental Algorithms*, volume 7276 of *Lecture Notes in Computer Science*, pages 87–98. Springer Berlin Heidelberg, 2012.

[13] S. Dey, M. Molinaro, and Q. Wang. Approximating polyhedra with sparse inequalities. *Mathematical Programming*, 154(1):329–352, 2015.

[14] E.D. Dolan and J.J. Moré. Benchmarking optimization software with performance profiles. *Mathematical Programming*, 91(2):201–213, 2002.

[15] M. Dyer, A. Frieze, and R. Kannan. A random polynomial-time algorithm for approximating the volume of convex bodies. *J. ACM*, 38(1):1–17, January 1991.

[16] A. Fügenschuh, H. Homfeld, H. Schülldorf, and S. Vigerske. Mixed-integer nonlinear problems in transportation applications. In H. Rodrigues, editor, *Proceedings of the 2nd International Conference on Engineering Optimization (+CD-rom)*, 2010.

[17] B. Grünbaum. *Convex polytopes*. Springer-Verlag, New York, 2nd edition, 2003.

[18] B. Grünbaum and V. P. Sreedharan. An enumeration of simplicial 4-polytopes with 8 vertices. *Journal of Combinatorial Theory*, 2(4):437–465, June 1967.

[19] O. Günlük, J. Lee, and R. Weismantel. MINLP strengthening for separable convex quadratic transportation-cost UFL. Technical report RC24213 (W0703-042), IBM Research Division, March 2007.

[20] O. Günlük and J. Linderoth. Perspective relaxation of mixed integer nonlinear programs with indicator variables. In *Proceedings of the 13th International Conference on Integer Programming and Combinatorial Optimization*, IPCO'08, pages 1–16. Springer-Verlag, 2008.

[21] B.A. Hendrickson and M.H. Wright. Mathematical research challenges in optimization of complex systems department of energy workshop report, December 7-8 2006.

[22] M. Jach, D. Michaels, and R. Weismantel. The convex envelope of (n-1)-convex functions. *SIAM Journal on Optimization*, 19(3):1451–1466, 2008.

[23] C.-W. Ko, J. Lee, and E. Steingrímsson. The volume of relaxed Boolean-quadric and cut polytopes. *Discrete Mathematics*, 163(1-3):293–298, 1997.

[24] J. Lee. Mixed integer nonlinear programming: Some modeling and solution issues. *IBM Journal of Research and Development*, 51(3/4):489–497, 2007.

[25] J. Lee and W. Morris. Geometric comparison of combinatorial polytopes. *Discrete Applied Mathematics*, 55:163–182, 1994.

[26] J. Lee and D. Skipper. Volume computation for sparse boolean quadric relaxations. arXiv:1703.02444, `https://arxiv.org/abs/1703.02444`, 2017.

[27] M.A. Lejeune and F. Margot. Solving chance-constrained optimization problems with stochastic quadratic inequalities. *Operations Research*, 64(4):939–957, 2016.

[28] X. Li, E. Armagan, A. Tomasgard, and P.I. Barton. Stochastic pooling problem for natural gas production network design and operation under uncertainty. *AIChE Journal*, 57(8):2120–2135, 2010.

[29] L. Liberti, S. Cafieri, and D. Savourey. The reformulation-optimization software engine. In K. Fukuda, J. Hoeven, M. Joswig, and N. Takayama, editors, *Mathematical Software — ICMS 2010*, volume 6327 of *Lecture Notes in Computer Science*, pages 303–314. Springer Berlin Heidelberg, 2010.

[30] L. Liberti, S. Cafieri, and F. Tarissan. Reformulations in mathematical programming: A computational approach. In A. Abraham, A. Hassanien, P. Siarry, and A. Engelbrecht, editors, *Foundations of Computational Intelligence Volume 3*, volume 203 of *Studies in Computational Intelligence*, pages 153–234. Springer Berlin Heidelberg, 2009.

[31] J. Luedtke, M. Namazifar, and J. Linderoth. Some results on the strength of relaxations of multilinear functions. *Mathematical Programming*, 136:325–351, 2012.

[32] G.P. McCormick. Computability of global solutions to factorable nonconvex programs: Part I. Convex underestimating problems. *Mathematical Programming*, 10:147–175, 1976.

[33] C.A. Meyer and C.A. Floudas. Trilinear monomials with mixed sign domains: Facets of the convex and concave envelopes. *Journal of Global Optimization*, 29:125–155, 2004.

[34] C.A. Meyer and C.A. Floudas. Trilinear monomials with positive or negative domains: Facets of the convex and concave envelopes. *Frontiers in Global Optimization*, pages 327–352, 2004.

[35] C.A. Meyer and C.A. Floudas. Convex envelopes for edge-concave functions. *Mathematical Programming*, 103(2, Ser. B):207–224, 2005.

[36] R. Misener. private communication, August 2016.

[37] R. Misener and C. A. Floudas. ANTIGONE: Algorithms for coNTinuous / Integer Global Optimization of Nonlinear Equations. *Journal of Global Optimization*, 2014. DOI: 10.1007/s10898-014-0166-2.

[38] R. Misener and C.A. Floudas. Advances for the pooling problem: modeling, global optimization, and computational studies *survey*. *Applied Computational Mathematics*, 8(1):3–22, 2009.

[39] R. Misener and C.A. Floudas. Glomiqo: Global mixed-integer quadratic optimizer. *Journal of Global Optimization*, 57(1):3–50, 2013.

[40] H. Nagarajan, M. Lu, E. Yamangil, and R. Bent. Tightening McCormick relaxations for nonlinear programs via dynamic multivariate partitioning. arXiv:1606.05806, `https://arxiv.org/abs/1606.05806`, 2016.

[41] A. Perold. Large-scale portfolio optimization. *Management Science*, 30:1143–1160, 1984.

[42] A. Rikun. A convex envelope formula for multilinear functions. *Journal of Global Optimization*, 10:425–437, 1997.

[43] R.T. Rockafellar. Lagrange multipliers and optimality. *SIAM Review*, 35(2), June 1993.

[44] A.E. Rosenbluth, S.J. Bukofsky, M.S. Hibbs, K. Lai, A.F. Molless, R.N. Singh, and A.K.K. Wong. Optimum mask and source patterns to print a given shape. In *Proceedings of the XIV SPIE Conference on Optical Microlithography*, volume 4346, pages 486–502, 2001.

[45] H.S. Ryoo and N.V. Sahinidis. A branch-and-reduce approach to global optimization. *Journal of Global Optimization*, 8(2):107–138, 1996.

[46] H.S. Ryoo and N.V. Sahinidis. Analysis of bounds for multilinear functions. *Journal of Global Optimization*, 19(4):403–424, 2001.

[47] M. Sagraloff. On the complexity of the Descartes method when using approximate arithmetic. *Journal of Symbolic Computation*, 65:79–110, 2014.

[48] N.V. Sahinidis. private communication, August 2016.

[49] N.V. Sahinidis. *BARON 15.6.5: Global Optimization of Mixed-Integer Nonlinear Programs,* User's Manual, 2015.

[50] H. Schichl and A. Neumaier. Interval analysis on directed acyclic graphs for global optimization. *Journal of Global Optimization*, 33(4):541–562, 2005.

[51] R. Schneider. *Convex bodies: the Brunn-Minkowski theory*, volume 44 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, 1993.

[52] R. Schneider. *Convex bodies: the Brunn-Minkowski theory*, volume 151 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, expanded edition, 2014.

[53] E.M.B. Smith and C.C. Pantelides. A symbolic reformulation/spatial branch-and-bound algorithm for the global optimisation of nonconvex MINLPs. *Computers & Chemical Engineering*, 23:457–478, 1999.

[54] A.M. Sofi, M. Mamat, S.Z. Mohid, M.A.H. Ibrahim, and N. Khalid. Performance profile comparison using matlab. In *Proceedings of International Conference on Information Technology & Society*, 2015.

[55] E. Speakman and J. Lee. On sBB branching for trilinear monomials. In A. Rocha, M. Costa, and E. Fernandes, editors, *Proceedings of the XIII Global Optimization Workshop (GOW16)*, pages 81–84, 2016. ISBN: 978-989-20-6764-3.

[56] E. Speakman and J. Lee. Quantifying double McCormick. *To appear in: Mathematics of Operations Research*, 2017.

[57] E. Speakman, H. Yu, and J. Lee. Experimental validation of volume-based comparison for double McCormick. *To appear in: The proceedings of the Fourteenth International Conference on Integration of Artificial Intelligence and Operations Research Techniques in Constraint Programming (CPAIOR)*, 2017.

[58] R.P. Stanley. Two poset polytopes. *Discrete & Computational Geometry*, 1(1):9–23, 1986.

[59] E. Steingrímsson. A decomposition of 2-weak vertex-packing polytopes. *Discrete Computational Geometry*, 12(4):465–479, 1994.

[60] M. Tawarmalani and N.V. Sahinidis. *Convexification and global optimization in continuous and mixed-integer nonlinear programming: theory, algorithms, software and applications*, volume 65 of *Nonconvex Optimization and Its Applications*. Kluwer Academic Publishers, Dordrecht, 2002.

[61] S. Vigerske and A. Gleixner. `SCIP`: Global optimization of mixed-integer nonlinear programs in a branch-and-cut framework. Technical Report 16-24, ZIB, Takustr.7, 14195 Berlin, 2016.

[62] X. Vu, H. Schichl, and D. Sam-Haroud. Interval propagation and search on directed acyclic graphs for numerical constraint solving. *Computational Optimization and Applications*, 45(4):499–531, 2009.