Bärmann, Andreas; Burlacu, Robert; Hager, Lukas; Kleinert, Thomas

On piecewise linear approximations of bilinear terms: structural comparison of univariate and bivariate mixed-integer programming formulations

Journal of Global Optimization

Provided in Cooperation with:
Springer Nature

Suggested Citation: Bärmann, Andreas; Burlacu, Robert; Hager, Lukas; Kleinert, Thomas (2022) : On piecewise linear approximations of bilinear terms: structural comparison of univariate and bivariate mixed-integer programming formulations, Journal of Global Optimization, ISSN 1573-2916, Springer US, New York, NY, Vol. 85, Iss. 4, pp. 789-819, https://doi.org/10.1007/s10898-022-01243-y

This Version is available at:
https://hdl.handle.net/10419/312457

Journal of Global Optimization (2023) 85:789–819
https://doi.org/10.1007/s10898-022-01243-y

On piecewise linear approximations of bilinear terms: structural comparison of univariate and bivariate mixed-integer programming formulations

Andreas Bärmann[1] · Robert Burlacu[2,3] · Lukas Hager[1] · Thomas Kleinert[1]

Received: 5 August 2021 / Accepted: 12 September 2022 / Published online: 3 November 2022

Abstract
Bilinear terms naturally appear in many optimization problems. Their inherent non-convexity typically makes them challenging to solve. One approach to tackle this difficulty is to use bivariate piecewise linear approximations for each variable product, which can be represented via mixed-integer linear programming (MIP) formulations. Alternatively, one can reformulate the variable products as a sum of univariate functions. Each univariate function can again be approximated by a piecewise linear function and modelled via an MIP formulation. In the literature, heterogeneous results are reported concerning which approach works better in practice, but little theoretical analysis is provided. We fill this gap by structurally comparing bivariate and univariate approximations with respect to two criteria. First, we compare the number of simplices sufficient for an ε-approximation. We derive upper bounds for univariate approximations and compare them to a lower bound for bivariate approximations. We prove that for a small prescribed approximation error ε, univariate ε-approximations require fewer simplices than bivariate ε-approximations. The second criterion is the tightness of the continuous relaxations (CR) of corresponding sharp MIP formulations. Here, we prove that the CR of a bivariate MIP formulation describes the convex hull of a variable product, the so-called McCormick relaxation. In contrast, we show by a volume argument that the CRs corresponding to univariate approximations are strictly looser. This allows us to explain many of the computational effects observed in the literature and to give theoretical evidence on when to use which kind of approximation.

✉ Lukas Hager
lukas.hager@fau.de

Andreas Bärmann
andreas.baermann@math.uni-erlangen.de

Robert Burlacu
robert.burlacu@iis.fraunhofer.de

Thomas Kleinert
thomas.kleinert@fau.de

[1] Chair of Analytics and Mixed-Integer Optimization, Friedrich-Alexander-Universität Erlangen-Nürnberg, Cauerstr. 11, 91058 Erlangen, Germany

[2] Fraunhofer Institute for Integrated Circuits IIS, Nordostpark 93, 90411 Nuremberg, Germany

[3] Energie Campus Nürnberg, Fürther Str. 250, 90429 Nuremberg, Germany

Keywords Bilinear programming · Piecewise linear approximations · MIP formulations · Univariate reformulations · Convex relaxations

Mathematics Subject Classification 90C20 · 90C59 · 90C11 · 51A27 · 15A63

1 Introduction

Many real-world optimization problems contain bilinear terms. For example, the modelling of economic interactions quite often results in products of prices and (production) quantities in optimization models; see e.g. [11, 18]. Other applications of bilinear programming include water management [20], gas network optimization [13, 14, 31] or pooling problems [8, 33]. In practice, such bilinear terms of continuous variable products xy are often approximated by piecewise linear functions, because they can be modelled using mixed-integer linear formulations; see e.g. [6, 15, 17, 26, 30, 39, 44]. For any pre-specified $\varepsilon > 0$, this can be done in such a way that the *maximum approximation error*, given as the maximum absolute pointwise deviation between the pwl. approximation and the non-linear function, is at most ε for each term. One straightforward approach is to use mixed-integer programming (MIP) formulations for bivariate piecewise linear functions that approximate xy; see e.g. [16, 29, 47, 50]. At the same time, it is well known that xy can be reformulated as a sum of univariate functions using additional variables and constraints. For example, in [3, 28, 38, 49] the authors suggest to use the substitution $xy = p_1^2 - p_2^2$ with $p_1 := \frac{1}{2}(x + y)$ and $p_2 := \frac{1}{2}(x - y)$. The monomials p_1^2 and p_2^2 can then be approximated by two univariate piecewise linear functions, using a separate MIP formulation for each of these functions. This raises the main question of this article: which approach is more efficient in which situation?

In [36], it is suggested that there is no clear answer as to whether or not to reformulate products of variables by several univariate functions. This claim is supported by heterogeneous computational results from the literature. On the one hand, it is shown in [50] in a small computational study in the context of planning decentralized energy grids that a bivariate piecewise linear approximation may outperform a quadratic univariate formulation on certain instances. On the other hand, in [1] the authors obtain very good computational results with a quadratic univariate reformulation. Similarly, [21, 41] report good results for a univariate logarithmic reformulation. The authors of the latter articles suspect that this is due to the smaller number of simplices required by the MIP formulations they use. From the computational experience in the literature reviewed above, we conclude that the actual choice of univariate and bivariate piecewise linear functions used to approximate xy is crucial for their respective performance. From a theoretical point of view, the literature offers much fewer analysis of the two approaches. Firstly, the best choice of a bivariate piecewise linear approximation—uniquely determined by the given triangulation of the domain—is not straightforward. In particular, finding an explicit construction rule for the optimal triangulation (w.r.t. the number of triangles) of a rectangular domain in order to approximate xy is still an open problem. In [29], the author gives an implicit construction via a mixed-integer quadratically constrained quadratic program (MIQCQP). In the univariate case, there exist algorithms that can compute optimal piecewise linear approximations, for example for continuous functions (see [35]). However, these algorithms do not provide an algebraic expression of the approximation error. Further, [21] is the only theoretical analysis on the topic of univariate reformulations we are aware of. The authors derive an upper bound for the approximation error of a univariate logarithmic reformulation. They use it to construct ε-approximations that are more compact

than direct bivariate piecewise linear approximations on problem instances from the field of paper production. However, as the triangulations are chosen heuristically, their results are not sufficient to state that in general univariate reformulations require less simplices.

Altogether, there is no rigorous comparison up to now which allows a comparison of the two approximation approaches with respect to the required number of simplices. Apart from the mentioned studies, there is—to the best of our knowledge—no theoretical analysis that would give a recommendation under which circumstances any one of the approaches is preferable. Furthermore, we are not aware of any works which analyse the continuous relaxations of the corresponding MIP formulations. A tighter continuous relaxation results in a tighter root relaxation of the branch-and-bound tree and therefore helps to keep the tree small. Since the number of simplices determines the number of necessary binary variables, less simplices directly lead to a smaller branch-and-bound tree.

In this paper, we fill the observed gap in the literature concerning a theoretical comparison of univariate and bivariate MIP formulations for piecewise linear approximations of xy. We establish hierarchies among them with respect to the following two criteria:

(i) the number of simplices that are required to guarantee an approximation of xy with a given accuracy and
(ii) the tightness of the continuous relaxation of an MIP formulation with respect to the graph of xy in terms of the enclosed volume.

Naturally, both aspects are crucial for the efficient solution of optimization problems containing bilinear terms with branch-and-cut algorithms. In this respect, we will highlight two important findings. First, we prove that commonly used monomial univariate reformulations always require fewer simplices than any bivariate approximation, as long as the prescribed error is small. Second, we show that the continuous relaxations of bivariate approximations always equal the McCormick relaxations and are genuinely tighter than the continuous relaxations of univariate reformulations. In addition, we derive a hierarchy among the univariate reformulations with respect to both questions.

The remainder of this paper is structured as follows. In Sect. 2, we introduce the general notation and concepts that are used throughout the paper. Afterwards, we compare structural properties of the bivariate and univariate approximations in Sect. 3. In particular, in Sect. 3.1 we compare the number of required simplices, and in Sect. 3.2 the strength of the continuous relaxations of MIP formulations. In Sect. 3.3, we discuss how these results can be used for practical applications. We show why approximations with as few simplices as possible are advantageous for setting up good piecewise linear relaxations of xy and explain how to convert known cutting planes for quadratic expressions into univariately reformulated models. Finally, we draw our conclusions in Sect. 4.

2 Piecewise linear functions, approximations and MIP formulations

We start by collecting the relevant background needed for this work. We introduce piecewise linear functions, discuss their use in approximating non-linear functions and present the concept of MIP formulations to model piecewise linear functions.

2.1 Piecewise linear functions and approximations

A piecewise linear (pwl.) function is linear on each element of a given domain partition. In general, it is possible to use any family of polytopes to construct such a partition. However,

in practice most often triangulations are used, see e.g. [47]. Therefore, we limit ourselves to pwl. functions over triangulations. This is without loss of generality as a pwl. function defined on a polytopal partition can always be represented by a pwl. function over a triangulation, namely by triangulating each polytope.

In the following, we formally introduce the relevant definitions in this context. For the sake of simplicity, we restrict ourselves to continuous functions over compact domains. Further, we use the notation $V(P)$ for the vertex set of a polyhedron $P \subset \mathbb{R}^d$.

Definition 1 A *n-simplex* S is the convex hull of $n+1$ affinely independent points in $\mathbb{R}^d$. We call S a *full-dimensional simplex* if $n = d$ holds.

A triangulation is a partition consisting of full-dimensional simplices as defined next.

Definition 2 A set of full-dimensional simplices $\mathcal{T} := \{S_1, \ldots, S_k\}$, with $S_i \subset \mathbb{R}^d$ for $i = 1, \ldots, k$, is called a *triangulation* of a compact set $B \subseteq \mathbb{R}^d$ if both $B = \cup_{i=1}^k S_i$ holds and the intersection of the relative interiors $\text{int}(S_i)$, $\text{int}(S_j)$ of any two simplices $S_i, S_j \in \mathcal{T}$ is empty, i.e. $\text{int}(S_i) \cap \text{int}(S_j) = \emptyset$. Further, we denote the set of vertices of a triangulation $\mathcal{T}$ by $N(\mathcal{T}) := \cup_{i=1}^k V(S_i)$.

Using the above definition of a triangulation, we can define pwl. functions as follows.

Definition 3 Let $B \subset \mathbb{R}^d$ be a compact set, and let $\mathcal{T} := \{S_1, \ldots, S_k\}$, $k \in \mathbb{N}$, be a triangulation of B. A continuous function $g\colon B \to \mathbb{R}$ is called *piecewise linear* if there exist vectors $m_i \in \mathbb{R}^d$ and constants $c_i \in \mathbb{R}$ for $i = 1, \ldots, k$ such that

$$g(x) = m_i^\top x + c_i \quad \text{if } x \in S_i. \tag{1}$$

In particular, for univariate pwl. functions $g\colon [\underline{x}, \bar{x}] \to \mathbb{R}$ the simplices S_i in a triangulation of $[\underline{x}, \bar{x}]$ correspond to intervals $[x_{i-1}, x_i]$ with $x_{i-1} < x_i$, $x_0 = \underline{x} > -\infty$ and $x_k = \bar{x} < \infty$.

Piecewise linear functions can be used to approximate non-linear functions, as shown in the next definition.

Definition 4 Let $B \subset \mathbb{R}^d$ be a compact set, and let $\mathcal{T} := \{S_1, \ldots, S_k\}$, $k \in \mathbb{N}$, be a triangulation of B. We call a pwl. function $g\colon B \to \mathbb{R}$ a *pwl. approximation* of a continuous function $G\colon B \to \mathbb{R}$ if $g(x) = G(x)$ holds for all $x \in N(\mathcal{T})$.

Note that in this definition of a pwl. approximation, we restrict ourselves to interpolations. This is partly because some mixed-integer programming models of pwl. functions require continuity of the approximation, and partly because some of the results from the literature presented here have been developed specifically for interpolations (cf. [22, 29, 40]). Usually, the error of a pwl. approximation is measured by the maximum absolute pointwise deviation between the pwl. approximation itself and the non-linear function to be approximated; see e.g. [21, 22, 36, 50]. In the following, we also use this definition of the *approximation error* and extend it to separable functions by introducing the so-called *combined approximation error*. The latter reflects the cancellations between positive and negative local approximation errors of the individual univariate summands a separable function decomposes into.

Definition 5 Consider a triangulation $\mathcal{T}$ of a compact set $B \subset \mathbb{R}^d$ and let $g\colon B \to \mathbb{R}$ be a pwl. approximation of a continuous function $G\colon B \to \mathbb{R}$ w.r.t. $\mathcal{T}$. We call

$$E_{g,G}\colon B \to \mathbb{R}, \quad E_{g,G}(x) := g(x) - G(x)$$

the *error function* of g w.r.t. G and

$$\varepsilon_{g,G}(S) := \max_{x \in S} |E_{g,G}(x)|$$

the *approximation error* on a simplex $S \in \mathcal{T}$. Consequently, we define the *approximation error* of g (or, equivalently, of $\mathcal{T}$) w.r.t. G over the domain B as

$$\varepsilon_{g,G}(\mathcal{T}) := \max_{S \in \mathcal{T}} \varepsilon_{g,G}(S).$$

In the special case that $G(x) = \sum_{i=1}^{n} G_i(x_i)$ is a separable function and $g(x) = \sum_{i=1}^{n} g_i(x_i)$ is a separable pwl. approximation of G with pwl. approximations g_i of G_i, we define the *combined approximation error* as

$$\varepsilon_{g,G}((\mathcal{T}_i)_{i \in n}) := \max_{x \in B} \left| \sum_{i=1}^{n} E_{g_i,G_i}(x_i) \right|.$$

Given some $\varepsilon > 0$, we call g an *ε-approximation* and $\mathcal{T}$ (or $(\mathcal{T}_i)_{i \in n}$) an *$\varepsilon$-triangulation* (or an *$\varepsilon$-family of triangulations*) if the (combined) approximation error is less than or equal to ε.

For our results regarding the approximation error of univariate reformulations of non-linear functions, we use the following straightforward upper bound for the combined approximation error of a separable function.

Lemma 1 *Consider a compact set B and a separable continuous function $G\colon B \subset \mathbb{R}^d \to \mathbb{R}$, $G(x) = \sum_{i=1}^{d} G_i(x_i)$. Further, let $g\colon B \subset \mathbb{R}^d \to \mathbb{R}$, $g(x) = \sum_{i=1}^{d} g_i(x_i)$ be a separable pwl. approximation of G where each g_i is a pwl. approximation of G_i. Then the combined approximation error fulfils*

$$\varepsilon_{g,G}\left((\mathcal{T}_i)_{i=1,\dots,d}\right) \leq \max_{x \in B} \left\{ \sum_{i=1}^{d} \max\{0, E_{g_i,G_i}(x_i)\}, \left| \sum_{i=1}^{d} \min\{0, E_{g_i,G_i}(x_i)\} \right| \right\}.$$

2.2 Mixed-integer formulations of pwl. functions

Consider a continuous function $G\colon B \to \mathbb{R}$ and its pwl. approximation $g\colon B \to \mathbb{R}$. In the following, we focus on representations of the *graph* of g, defined as

$$\mathrm{gra}_{\bar{B}}(g) := \{(x, z) \in \bar{B} \times \mathbb{R} : z = g(x)\},$$

where we allow the restriction of g to a subset $\bar{B} \subseteq B$. When solving optimization problems where g occurs in the objective function or in the constraints, it is impractical to work with Definition 3 directly. Instead, we need an explicit representation of the "if"-condition in Eq. (1). Very often this is done by expressing g in terms of $\mathrm{gra}(g)$. For example, minimizing over g is equivalent to minimizing z subject to $(x, z) \in \mathrm{gra}(g)$. The graph of a pwl. function can be modelled with the help of additional auxiliary continuous and binary variables as well as linear constraints (cf. [24–27]).

Definition 6 Let $g\colon B \to \mathbb{R}$ be a pwl. function, with $B \subset \mathbb{R}^d$. We call the set $M_g \subseteq \mathbb{R}^{d+1} \times [0,1]^p \times \{0,1\}^q$ an *MIP formulation* of $\mathrm{gra}(g)$ if

$$(x, z) \in \mathrm{gra}(g) \iff \exists (\lambda, u) \in [0,1]^p \times \{0,1\}^q \quad \text{s.t.} \quad (x, z, \lambda, u) \in M_g.$$

Furthermore, we call the polyhedron

$$C(M_g) := \{(x, z, \lambda, v) \in \mathbb{R}^{d+1} \times [0,1]^p \times [0,1]^q : \exists (x, z, \lambda, u) \in M_g\}$$

the *continuous relaxation (CR)* of the MIP formulation M_g and

$$\mathrm{proj}_{(x,z)}\, C(M_g)$$

Table 1 Univariate reformulations of the bivariate product xy

Label	Substitution	Add. constraints	Refs.
Bin1	$xy = p_1^2 - p_2^2$	$p_1 = \frac{1}{2}(x+y)$, $p_2 = \frac{1}{2}(x-y)$	[3, 28, 38, 49]
Bin2	$xy = \frac{1}{2}(p^2 - x^2 - y^2)$	$p = x + y$	[1, 36]
Bin3	$xy = \frac{1}{2}(x^2 + y^2 - p^2)$	$p = x - y$	[50]
Ln	$xy = p$	$\ln(p) = \ln(x) + \ln(y)$	[21–23, 41]

its *projected continuous relaxation (PCR)*, where $\text{proj}_{(x,z)}\, C(M_g)$ is the projection of $C(M_g)$ onto the (x, z)-space.

Note that the dimensions p and q of the continuous and the binary auxiliary variables, respectively, do not necessarily coincide. In [46], several such MIP formulations for the graph of a pwl. function are presented, e.g. the incremental method or the multiple-choice method, with their respective sizes stated in Table 1. All MIP formulations mentioned there have the desirable property to be *sharp*. In order to define sharpness, we need some more notation. For this reason, we define the terms convex envelope and concave envelope, which we use to describe the convex hull of the graph.

Definition 7 Consider a continuous function $G\colon B \to \mathbb{R}$ over a compact set $B \subset \mathbb{R}^d$. We define the functions $\text{convenv}_{\bar{B}}(G)\colon \bar{B} \to \mathbb{R}$ and $\text{caveenv}_{\bar{B}}(G)\colon \bar{B} \to \mathbb{R}$ via

$$\text{convenv}_{\bar{B}}(G)(x) := \sup\{h(x) \mid h\colon B \to \mathbb{R} \text{ convex} \wedge h(x) \leq G(x)\, \forall x \in \bar{B}\},$$
$$\text{caveenv}_{\bar{B}}(G)(x) := \inf\{h(x) \mid h\colon B \to \mathbb{R} \text{ concave} \wedge h(x) \geq G(x)\, \forall x \in \bar{B}\},$$

as the *convex envelope* and the *concave envelope* of G with respect to $\bar{B} \subset B$.

We have

$$\text{conv}(\text{gra}_B(G)) = \{(x, z) \in B \times \mathbb{R} : \text{convenv}_B(g)(x) \leq z \leq \text{caveenv}_B(g)(x)\} \quad (2)$$

for the convex hull of $\text{gra}(g)$. For brevity, we use the notation $\text{gra}(g) := \text{gra}_B(g)$, $\text{convenv}(g) := \text{convenv}_B(g)$ and $\text{caveenv}(g) := \text{caveenv}_B(g)$.

An MIP formulation of a graph is called *sharp* if its PCR coincides with the convex hull of the graph.

Definition 8 Let $g\colon B \to \mathbb{R}$ be a continuous pwl. function. An MIP formulation M_g of $\text{gra}(g)$ is called *sharp* if

$$\text{conv}(\text{gra}(g)) = \text{proj}_{(x,z)}\, C(M_g).$$

To obtain a finer measure of the strength of an MIP formulation M_g, we further consider the volume of its PCR, namely $\text{vol}(\text{proj}_{(x,z)}\, C(M_g))$. The volume of an MIP formulation M_g for a corresponding pwl. function g is minimal if M_g is sharp, i.e. we have $\text{vol}(\text{conv}(\text{gra}(g))) = \text{vol}(\text{proj}_{(x,z)}\, C(M_g))$. If M_g is not sharp, the volume can be larger. We say that a MIP formulation is *looser* or *tighter* than another if the volume of its PCR is larger or smaller, respectively. These terms are suitable in the sense that the volume of the PCR is the integral over the maximum pointwise deviation to $\text{gra}(g)$. The volume can therefore be interpreted as an overall error measure of the continuous relaxation.

3 Structural properties of univariate and bivariate piecewise linear approximations

Our work focusses on the structural analysis of pwl. approximations of the non-linear function

$$F\colon D \to \mathbb{R}, \quad F(x, y) = xy,$$

where $D := [\underline{x}, \bar{x}] \times [\underline{y}, \bar{y}] \subset \mathbb{R}^2$ is a box domain with $\underline{x} < \bar{x}$ and $\underline{y} < \bar{y}$. It is a straightforward idea to approximate F via a bivariate pwl. function $f\colon D \to \mathbb{R}$. Using an MIP formulation M_f, we can then model $\mathrm{gra}(f)$ as

$$\mathrm{gra}(f) = \{(x, y, z) \in D \times \mathbb{R} \mid (x, y, z, \lambda, u) \in M_f\} \tag{3}$$

in order to obtain a mixed-integer linear representation of f.

Alternatively, we can equivalently reformulate F as a sum of univariate functions in order to approximate F by approximating each individual function in the sum. This reformulation can be done in various ways. Table 1 summarizes—to the best of our knowledge—all univariate reformulations of F used in the optimization literature. It shows the corresponding variable substitutions, the additionally required constraints as well as bibliographical references for the use of each reformulations in optimization.

Although we also list the logarithmic reformulation Ln in Table 1, we will not discuss it further in this work for various reasons. Firstly, the literature reports numerical difficulties in connection with the use of this reformulation in practice (see [10, 22, 50]), which is plausible given the asymptotic behavior of the logarithm for inputs close to zero. Secondly, Ln is only applicable in the case $\underline{x} > 0$ and $\underline{y} > 0$. Although this condition can always be fulfilled via a simple bound-shifting trick (see [21]), a shifted approximation does not retain its accuracy in general, as elementary examples show. Further, the upper bounds on the combined error of a pwl. approximation based on Ln stated in [21] deteriorate with increasing shift values as well.

In the following, we exemplarily derive an MIP formulation for a univariate approximation of $\mathrm{gra}(F)$ via reformulation Bin1 from Table 1. First, the graph of F can be stated as

$$\mathrm{gra}(F) = \left\{(x, y, p_1^2 - p_2^2) \in \mathbb{R}^3 \mid p_1 = \frac{1}{2}(x + y),\ p_2 = \frac{1}{2}(x - y),\ (x, y) \in D\right\}. \tag{4}$$

The domains of the additional variables p_1 and p_2 are consequently given by

$$D_1 := [\underline{p}_1, \bar{p}_1] := \left[\frac{1}{2}(\underline{x} + \underline{y}), \frac{1}{2}(\bar{x} + \bar{y})\right] \subset \mathbb{R},$$
$$D_2 := [\underline{p}_2, \bar{p}_2] := \left[\frac{1}{2}(\underline{x} - \bar{y}), \frac{1}{2}(\bar{x} - \underline{y})\right] \subset \mathbb{R}.$$

Now, let $f_1^{\mathrm{Bin1}}\colon D_1 \to \mathbb{R}$ and $f_2^{\mathrm{Bin1}}\colon D_2 \to \mathbb{R}$ be pwl. approximations of $F_1^{\mathrm{Bin1}}\colon D_1 \to \mathbb{R}$, $F_1^{\mathrm{Bin1}}(p_1) = p_1^2$ and $F_2^{\mathrm{Bin1}}\colon D_2 \to \mathbb{R}$, $F_2^{\mathrm{Bin1}}(p_2) = p_2^2$ respectively, with corresponding triangulations $\mathcal{T}_1^{\mathrm{Bin1}}$ and $\mathcal{T}_2^{\mathrm{Bin1}}$. We define $f^{\mathrm{Bin1}}\colon D \to \mathbb{R}$ via

$$f^{\mathrm{Bin1}}(x, y) = f_1^{\mathrm{Bin1}}(p_1) - f_2^{\mathrm{Bin1}}(p_2),$$
$$\text{with} \quad p_1 = \frac{1}{2}(x + y), \quad p_2 = \frac{1}{2}(x - y).$$

Further, let $M_1^{\mathrm{Bin1}} \subseteq D \times \mathbb{R} \times [0, 1]^{p_1^{\mathrm{Bin1}}} \times \{0, 1\}^{q_1^{\mathrm{Bin1}}}$ and $M_2^{\mathrm{Bin1}} \subseteq D \times \mathbb{R} \times [0, 1]^{p_2^{\mathrm{Bin1}}} \times \{0, 1\}^{q_2^{\mathrm{Bin1}}}$ be sharp MIP formulations of the graphs $\mathrm{gra}(f_1^{\mathrm{Bin1}})$ and $\mathrm{gra}(f_2^{\mathrm{Bin1}})$. We can then

model an approximation of $\text{gra}(F)$ as

$$\text{gra}(f^{\text{Bin1}}) = \{(x, y, z) \in D \times \mathbb{R} \mid (x, y, z, \lambda_1, u_1, \lambda_2, u_2) \in M_{f^{\text{Bin1}}}\}, \tag{5}$$

together with the MIP formulation

$$\begin{aligned} M_{f^{\text{Bin1}}} := \{&(x, y, z, \lambda_1, u_1, \lambda_2, u_2) \in D \times \mathbb{R} \\ &\times [0,1]^{p_1^{\text{Bin1}}} \times \{0,1\}^{q_1^{\text{Bin1}}} \times [0,1]^{p_2^{\text{Bin1}}} \times \{0,1\}^{q_2^{\text{Bin1}}} \quad \mid \\ &\exists (p_1, z_1, \lambda_1, u_1) \in M_1^{\text{Bin1}},\ (p_2, z_2, \lambda_2, u_2) \in M_2^{\text{Bin1}} \quad \text{s.t.} \\ &z = z_1 - z_2,\ p_1 = \frac{1}{2}(x + y),\ p_2 = \frac{1}{2}(x - y),\ (x, y) \in D\}. \end{aligned}$$

Corresponding MIP formulations for Bin2 and Bin3 are stated in "Appendix A".

In the remainder of this section, we will compare bivariate MIP formulations for the approximation of $\text{gra}(F)$ as given in (3) to univariate MIP formulations, such as (5), using two different metrics of efficiency. In Sect. 3.1, we analyse the number of simplices required in each case to construct an ε-approximation. We will show that using Bin1, Bin2 and Bin3, we can construct ε-families of triangulations with a smaller number of simplices than needed for any bivariate ε-triangulation if the prescribed approximation accuracy ε is sufficiently small. Furthermore, we will prove that a particular equidistant family of triangulations is ε-optimal for Bin1. In Sect. 3.2, we then investigate the tightness of the continuous relaxations of univariate and bivariate MIP formulations. On the one hand, we show that the PCR of any bivariate MIP formulation coincides with the convex hull of $\text{gra}(F)$, which is known as the *McCormick relaxation* [32]. On the other hand, we show how to compute the PCRs of the considered univariate MIP formulations and prove that these are indeed weaker relaxations of $\text{gra}(F)$ than the McCormick relaxation. Moreover, we show that using Bin1 yields the tightest continuous relaxation among the studied univariate reformulations. Finally, we indicate in Sect. 3.3 how to use these theoretical results in practice. In particular, we outline how to overcome the fact that univariate MIP formulations yield weaker continuous relaxations by adding the linear inequalities describing the convex hull, which are known as the *McCormick cuts*, to the univariate MIP formulations in a reformulated fashion, as done in [1]. Furthermore, we suggest under which circumstances which univariate reformulation should be chosen.

3.1 Number of simplices

We start our comparison between bivariate and univariate pwl. approximations of the bilinear function F by considering the size of the resulting MIP formulation. In this respect, the overall number of binary variables is a crucial factor for the computational complexity of the resulting optimization problem. This number, however, strongly depends on the specific modelling of the MIP formulation, see [47]. The number of binary variables can be reduced significantly, for example, by a logarithmic encoding of the simplices, compared to a straightforward modelling approach as shown in [29, 48]. Therefore, we will instead compare pwl. approximations by the number of simplices required to obtain a prescribed approximation guarantee, which directly impacts the number of binary variables in any modelling of the arising MIP formulation.

To this end, we introduce the concept of ε-optimal triangulations for the pwl. approximation of a non-linear function. We use the same definition as in [29, 41] and refer to [5] for more context on optimal triangulations and possible alternative definitions.

Definition 9 Let $B \subseteq \mathbb{R}^d$ be a compact set, and let $g\colon B \to \mathbb{R}$ be a pwl. ε-approximation of the continuous function $G\colon B \to \mathbb{R}$ w.r.t. the underlying ε-triangulation $\mathcal{T}$ of B. We say that $\mathcal{T}$ is *ε-optimal* if $|\mathcal{T}|$ is minimal among all ε-triangulations of B.

In the special case that $G(x) = \sum_{i=1}^n G_i(x_i)$ is a separable function and $g(x) = \sum_{i=1}^n g_i(x_i)$ is a pwl. approximation of G, such that each g_i is a pwl. approximation of G_i, we say that the corresponding family of triangulations $(\mathcal{T}_i)_{i=1,\dots,n}$ is *ε-optimal* if $\sum_{i=1}^n |\mathcal{T}_i|$ is minimal among all ε-families of triangulations.

It is not obvious how to determine ε-optimal triangulations in general. To the best of our knowledge, the complexity status of this problem is still open. The only related result we are aware of is the NP-hardness of finding minimum edge-weighted triangulations, where the aim is to minimize the sum of the edge weights, see [37]. However, finding an ε-optimal triangulation corresponds to minimizing the maximum edge weight in the chosen triangulation, as shown in [29]. Thus, we will mostly work with lower and upper bounds on the required number of simplices for a pwl. approximation. More precisely, we will show that for a sufficiently small prescribed approximation accuracy $\varepsilon > 0$ we can construct ε-families of triangulations for Bin1, Bin2 and Bin3, such that the corresponding number of simplices is smaller than that of any bivariate ε-triangulation.

3.1.1 Univariate pwl. approximations

We will now consider the construction of ε-approximations for univariate reformulations of F. For this purpose, we study equidistant triangulations for pwl. approximations of univariate quadratic functions. We then prove that a particular family of equidistant triangulations is ε-optimal for reformulation Bin1. Finally, we derive upper bounds for the size of ε-optimal triangulations in the reformulations Bin2 and Bin3 by using equidistant triangulations.

Finding ε-triangulations for univariate functions has been extensively covered in the literature under the term *minimax approximation*. For an overview, we refer to [35], where the author also provides an algorithm for finding an ε-optimal piecewise polynomial approximation of degree n for a given continuous univariate function. In particular, this algorithm can be used to find pwl. approximations. Another approach can be found in [42]. Here, the authors present a mixed-integer non-linear optimization program (MINLP) for computing an ε-optimal continuous pwl. approximation for a given univariate function. However, both approaches do not provide closed functional relations for the required number of simplices depending on ε. In contrast, our focus here will be on deriving functional relations for the number of simplices of ε-families of triangulations in Bin1, Bin2 and Bin3. We start with a relation for ε-optimal families of triangulations in reformulation Bin1. In order to do so, we make use of the following lemma about linear approximations of univariate quadratic functions, which is straightforward to prove via differential calculus.

Lemma 2 *Let $G\colon [\underline{x}, \bar{x}] \to \mathbb{R}$, $G(x) = \alpha x^2 + \beta x + \gamma$ with $\alpha, \beta, \gamma \in \mathbb{R}$ be a quadratic function, and let $L\colon [\underline{x}, \bar{x}] \to \mathbb{R}$ be the linear interpolant of G between $\underline{x}$ and $\bar{x}$. Then the maximum approximation error of L w.r.t. G over $[\underline{x}, \bar{x}]$ is given by*

$$\max_{x \in [\underline{x}, \bar{x}]} |L(x) - G(x)| = |\alpha| \frac{(\bar{x} - \underline{x})^2}{4}.$$

It is attained at the centre of the domain, i.e. at $x^* := \frac{\underline{x}+\bar{x}}{2}$.

The following result extends Lemma 2 to pwl. approximations of univariate quadratic functions. It says that an equidistant placement of vertices minimizes the approximation error.

Lemma 3 *Given $G\colon [\underline{x}, \bar{x}] \to \mathbb{R}, G(x) = x^2$, let $\mathcal{T}$ be the triangulation of $[\underline{x}, \bar{x}]$ formed by an equidistant placement of the $n+1$ vertices $x_0 := \underline{x} < \ldots < x_n := \bar{x} \in \mathbb{R}$. Further, let $g\colon [\underline{x}, \bar{x}] \to \mathbb{R}$ be the pwl. approximation of G w.r.t. $\mathcal{T}$. Then the corresponding approximation error is given by*

$$\varepsilon_{g,G}(\mathcal{T}) = \frac{(\bar{x} - \underline{x})^2}{4n^2}.$$

Furthermore, the approximation error of g is minimal among all pwl. approximations of G over n simplices.

Proof Let the triangulation $\mathcal{T} := \{S_0, S_1, \ldots, S_{n-1}\}$ be given by the simplices $S_i := [x_i, x_{i+1}] = [x_i, x_i + h_i]$ with respective diameters $h_i := x_{i+1} - x_i$, $i = 0, 1, \ldots, n-1$. As the corresponding pwl. approximation g is linear over each S_i and coincides with G at the vertices, its linear segments are given by functions $g_i\colon S_i \to \mathbb{R}$ with

$$g_i(x) = (2\mathrm{x}_i + h_i)x - (\mathrm{x}_i^2 + \mathrm{x}_i h_i).$$

Lemma 2 states that the approximation error over each simplex S_i is attained at the respective midpoint, with

$$\varepsilon_{g,G}(S_i) = \frac{1}{4}h_i^2.$$

Thus, the approximation error is minimized by an equidistant placement of the vertices, i.e. for $h_i := (\bar{x} - \underline{x})/n$, $i = 0, 1, \ldots, n-1$. □

Note that the approximation error for a univariate quadratic function only depends on the diameter of the domain and the number of simplices of the triangulation and is thus invariant under shifts of the domain itself.

We can now prove that particular equidistant families of triangulations are ε-optimal for reformulation Bin1.

Lemma 4 *Let $f^{\mathrm{Bin1}} = f_1^{\mathrm{Bin1}} - f_2^{\mathrm{Bin1}}$ be a pwl. approximation of F, with a corresponding family of triangulations $(\mathcal{T}_i^{\mathrm{Bin1}})_{i=1,2}$ defining f_1^{Bin1} and f_2^{Bin1}, and let $n_i := |\mathcal{T}_i^{\mathrm{Bin1}}|$, $i = 1, 2$. Then the combined approximation error of $(\mathcal{T}_i^{\mathrm{Bin1}})_{i=1,2}$ is at least*

$$\bar{\varepsilon} := \frac{1}{16}(\bar{x} - \underline{x} + \bar{y} - \underline{y})^2 \max\left\{\frac{1}{n_1^2}, \frac{1}{n_2^2}\right\}.$$

In particular, it is attained if $\mathcal{T}_1^{\mathrm{Bin1}}$ and $\mathcal{T}_2^{\mathrm{Bin1}}$ are equidistant triangulations.

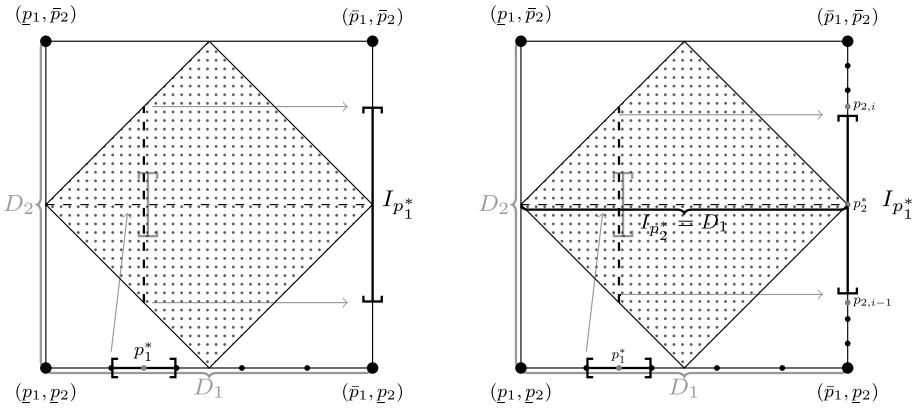
Conversely, an $\bar{\varepsilon}$-optimal family of triangulations $(\mathcal{T}_i^{\mathrm{Bin1}})_{i=1,2}$ is given by a pair of equidistant triangulations with

$$|\mathcal{T}_i^{\mathrm{Bin1}}| = \left\lceil \frac{\bar{x} - \underline{x} + \bar{y} - \underline{y}}{4\sqrt{\bar{\varepsilon}}} \right\rceil, \; i = 1, 2.$$

Proof First, note that $D_1 \times D_2$ is a quadratic box with a width of $(\bar{x} - \underline{x} + \bar{y} - \underline{y})/2$. Furthermore, the feasible domain of the variable substitution in Bin1, given by

$$I := \{(p_1, p_2) \in D_1 \times D_2 \mid \exists (x, y) \in D : p_1 = 0.5(x + y) \wedge p_2 = 0.5(x - y)\},$$

is a rhombus inscribed into this box. This situation is depicted in Fig. 1. Let $p_{i,0}, \ldots, p_{i,n_i}$ with $p_{i,j} < p_{i,j+1}$, $p_{i,0} = \underline{p}_i$ and $p_{i,n_i} = \bar{p}_i$ be the vertices in $N(\mathcal{T}_i)$, with $i = 1, 2$. W.l.o.g.,



(a) Case 1) of Lemma 4. **(b)** Case 2) of Lemma 4.

Fig. 1 Geometric arguments in the proof of Lemma 4

we assume $n_1 \leq n_2$. Further, for any $p_1^* \in D_1$ and $p_2^* \in D_2$, we define the projections of I onto the coordinate axes as

$$I_{p_1^*} := \{p_2 \in D_2 \mid \exists (x, y) \in D : p_1^* = 0.5(x + y) \wedge p_2 = 0.5(x - y)\},$$
$$I_{p_2^*} := \{p_1 \in D_2 \mid \exists (x, y) \in D : p_1 = 0.5(x + y) \wedge p_2^* = 0.5(x - y)\}$$

respectively. We consider now the following two exhaustive cases 1) and 2):
1) $p_{1,i} - p_{1,i-1} \leq \frac{\bar{x}-\underline{x}+\bar{y}-\underline{y}}{2n_1} \; \forall i = 1, \ldots, n_1 \wedge p_{2,j} - p_{2,j-1} \leq \frac{\bar{x}-\underline{x}+\bar{y}-\underline{y}}{2n_1} \; \forall j = 1, \ldots, n_2$: From this assumption it follows that $\mathcal{T}_1$ has to be equidistant. Moreover, we know from Lemma 3 that $\varepsilon_{f_1^{\text{Bin1}}, F_1^{\text{Bin1}}}(\mathcal{T}_1) = \bar{\varepsilon}$. By the same arguments, we also know that $\varepsilon_{f_2^{\text{Bin1}}, F_2^{\text{Bin1}}}(\mathcal{T}_2) \leq \bar{\varepsilon}$. Now let p_i^* be the midpoint of an arbitrary interval $[p_{1,i-1}, p_{1,i}]$. According to Lemma 2, we have $E_{f_1^{\text{Bin1}}, F_1^{\text{Bin1}}}(p_i^*) = \bar{\varepsilon}$. It is obvious by geometric reasoning that the diameter of the projection $I_{p_1^*}$ is longer than $(\bar{x} - \underline{x} + \bar{y} - \underline{y})/(2n_1)$, see Fig. 1a. As a result, there must be at least one vertex $p_{2,j}$ contained in $I_{p_1^*}$. As the approximation error at a vertex is always zero, it follows that the approximation error at $(p_1^*, p_{2,j})$ is

$$E_{f_1^{\text{Bin1}}, F_1^{\text{Bin1}}}(p_1^*) + E_{f_2^{\text{Bin1}}, F_2^{\text{Bin1}}}(p_{2,j}) = \bar{\varepsilon}.$$

In summary, we have

$$\varepsilon_{f^{\text{Bin1}}, F^{\text{Bin1}}}((\mathcal{T}_i)_{i=1,2}) = \bar{\varepsilon}.$$

2) $\exists 1 \leq i \leq n_1 : \; p_{1,i} - p_{1,i-1} > \frac{\bar{x}-\underline{x}+\bar{y}-\underline{y}}{2n_1} \vee \exists 1 \leq j \leq n_2 : p_{2,j} - p_{2,j-1} > \frac{\bar{x}-\underline{x}+\bar{y}-\underline{y}}{2n_1}$: W.l.o.g., we assume that the interval $[p_{1,i-1}, p_{1,i}]$ is longer than $(\bar{x} - \underline{x} + \bar{y} - \underline{y})/(2n_1)$. From Lemma 3, we know that $E_{f_1^{\text{Bin1}}, F_1^{\text{Bin1}}}(p_1^*) > \bar{\varepsilon}$, where p_1^* is the midpoint of $[p_{1,i-1}, p_{1,i}]$. Again, by geometric arguments, $I_{p_1^*}$ must be longer than $(\bar{x} - \underline{x} + \bar{y} - \underline{y})/2n_1$. However, due to the fact that the approximation error at a vertex is always zero, $I_{p_1^*}$ cannot contain any vertex $p_{2,j} \in N(\mathcal{T}_2)$ as this would imply that we have a point in $I_{p_1^*}$ at which the combined approximation error is greater than $\bar{\varepsilon}$, namely $(p_1^*, p_{2,j})$. Consequently, we have $I_{p_1^*} \subseteq [p_{2,j-1}, p_{2,j}]$. This means that at the midpoint p_2^* of D_2 (which is also the midpoint of $I_{p_1^*}$), $E_{f_2^{\text{Bin1}}, F_2^{\text{Bin1}}}(p_2^*) > \bar{\varepsilon}$ holds. Obviously, $I_{p_2^*} = D_1$, and therefore D_1 cannot contain

any points with an approximation error of zero, which is a contradiction to the fact that f_1^{Bin1} is a pwl. approximation (interpolation). □

It is not straightforward how to obtain a similar result as Lemma 4 for reformulations Bin2 and Bin3. The difficulty stems from the fact that in these two cases we have to approximate three functions simultaneously, instead of only two as in Bin1. However, we can still use equidistant triangulations to determine upper bounds on the number of simplices for Bin2 and Bin3.

Lemma 5 *Let $f^{\text{Bin2}} = 0.5(f_1^{\text{Bin2}} - f_2^{\text{Bin2}} - f_3^{\text{Bin2}})$ be a pwl. approximation of F as defined in Appendix A. Then for any $\varepsilon > 0$, there is an ε-family of equidistant triangulations $(\mathcal{T}_i^{\text{Bin2}})_{i=1,2,3}$ for the individual pwl. approximations f_1^{Bin2}, f_2^{Bin2} and f_3^{Bin2} with respective sizes*

$$|\mathcal{T}_1^{\text{Bin2}}| = \left\lceil \frac{(\bar{x}-\underline{x})+(\bar{y}-\underline{y})}{2\sqrt{2\varepsilon}} \right\rceil, \; |\mathcal{T}_2^{\text{Bin2}}| = \left\lceil \frac{(\bar{x}-\underline{x})}{2\sqrt{\varepsilon}} \right\rceil \text{ and } |\mathcal{T}_3^{\text{Bin2}}| = \left\lceil \frac{(\bar{y}-\underline{y})}{2\sqrt{\varepsilon}} \right\rceil.$$

Proof To obtain an ε-family of triangulations for Bin2, we use Lemma 3 to construct ε-triangulations for each of the two concave terms $-x^2$, approximated by $-f_2^{\text{Bin2}}$ and $-y^2$, approximated by $-f_3^{\text{Bin2}}$, as well as a 2ε-triangulation for the convex term $(x+y)^2$, approximated by f_1^{Bin2}. This directly yields the number of simplices stated in the claim. Taking into account the prefactor of 0.5 in the variable substitution, Lemma 1 then certifies that we have indeed found an ε-family of triangulations. □

The same result as above holds for Bin3, as it consists of the same quadratic terms, only with switched signs. The upper bounds for ε-families of triangulations derived so far are summarized in Table 2.

If we do not require ε-approximations for each of the terms $-x^2$ (or x^2) and $-y^2$ (or y^2) in Bin2 (or Bin3), but rather only require a 2ε-approximation for the combined approximation of these two functions, we can still apply Lemma 1 to obtain equidistant ε-families of triangulations, and it is possible in many cases to improve the bounds presented in Table 2. We can determine these improved bounds by solving a mixed-integer quadratically constrained quadratic program (MIQCQP) as follows.

Remark 1 Let $\varepsilon > 0$ be a prescribed maximum combined error for a pwl. approximation of F either via Bin2 or Bin3. Then we can compute the minimum possible number of simplices for any corresponding family of equidistant ε-triangulations as the optimal value n^* of the following optimization problem:

$$\begin{aligned} n^* := \min_{n_1,n_2,n_3} \; & n_1 + n_2 + n_3 \\ \text{s.t.} \; & \frac{(\bar{x}-\underline{x})^2}{4n_1^2} + \frac{(\bar{y}-\underline{y})^2}{4n_2^2} \le 2\varepsilon, \\ & \frac{(\bar{x}-\underline{x}+\bar{y}-\underline{y})^2}{4n_3^2} \le 2\varepsilon, \\ & n_1, n_2, n_3 \in \mathbb{N}. \end{aligned} \tag{6}$$

The variables n_1, n_2 and n_3 model the number of simplices used for the triangulations $\mathcal{T}_1^{\text{Bin2}}$, $\mathcal{T}_2^{\text{Bin2}}$ and $\mathcal{T}_3^{\text{Bin2}}$ (or $\mathcal{T}_1^{\text{Bin3}}$, $\mathcal{T}_2^{\text{Bin3}}$ and $\mathcal{T}_3^{\text{Bin3}}$) in the pwl. approximation of the terms $-x^2$, $-y^2$ and $+p^2$ (or x^2, y^2 and $-p^2$) respectively, see "Appendix A" for the complete models. The two inequality constraints of Problem (6) model the max-expression in the upper bound

Table 2 Upper bounds on the minimal number of simplices in an ε-family of triangulations in Bin1, Bin2 and Bin3. For Bin1, this is also the size of an ε-optimal family of triangulations

Reformulation	Max. required number of simplices
Bin1	$\left\lceil \frac{(\bar{x}-\underline{x})+(\bar{y}-\underline{y})}{4\sqrt{\varepsilon}} \right\rceil + \left\lceil \frac{(\bar{x}-\underline{x})+(\bar{y}-\underline{y})}{4\sqrt{\varepsilon}} \right\rceil$
Bin2, Bin3	$\left\lceil \frac{(\bar{x}-\underline{x})}{2\sqrt{\varepsilon}} \right\rceil + \left\lceil \frac{(\bar{y}-\underline{y})}{2\sqrt{\varepsilon}} \right\rceil + \left\lceil \frac{(\bar{x}-\underline{x})+(\bar{y}-\underline{y})}{2\sqrt{2\varepsilon}} \right\rceil$

on the combined approximation error provided by Lemma 1; the respective terms on the left-hand sides stem from Lemma 3. Note that Problem (6) can be equivalently reformulated as a non-convex MIQCQP:

$$
\begin{aligned}
n^* := \min_{n_1,n_2,n_3,\eta_1,\eta_2} \quad & n_1 + n_2 + n_3 \\
\text{s.t.} \quad & (\bar{x} - \underline{x})^2 \eta_2 + (\bar{y} - \underline{y})^2 \eta_1 \leq 8\varepsilon \cdot \eta_1 \eta_2, \\
& (\bar{x} - \underline{x} + \bar{y} - \underline{y})^2 \leq 8\varepsilon \cdot n_3^2, \\
& \eta_1 = n_1^2, \quad \eta_2 = n_2^2, \\
& n_1, n_2, n_3, \eta_1, \eta_2 \in \mathbb{N}.
\end{aligned}
$$

with auxiliary variables η_1 and η_2.

We cannot make a general hierarchical statement among the univariate reformulation Bin1, Bin2 and Bin3, since we do not know ε-optimal families of triangulations for Bin2 and Bin3. However, the simple fact that in Bin1 we only approximate two instead of three univariate functions suggests that ε-optimal families of triangulations for Bin2 and Bin3 consist of more simplices than those for Bin1.

xxx

3.1.2 Bivariate pwl. approximations

Finding a bivariate ε-optimal triangulation for the approximation of F over a rectangular domain is still an open problem, see the elaborations in [29] and the references therein. However, it will be sufficient for us to determine a lower bound on the number of simplices in an ε-optimal triangulation to see that in essence bivariate pwl. approximations of F require more simplices than univariate ones. In order to derive this lower bound, we first prove the following rather general lemma, which has been presented in the dissertation [12] of the second author. It gives sufficient conditions under which the maximum approximation error between a non-linear function and its pwl. approximation is attained at a facet of one of the simplices of the triangulation.

Lemma 6 *Let $G\colon B \to \mathbb{R}$ be a continuous function over a compact set $B \subset \mathbb{R}^d$, and let $g\colon B \to \mathbb{R}$ be a pwl. approximation of G defined by a triangulation $\mathcal{T}$ of B. If for each $x \in B$ there is a line $L_x \subseteq \mathbb{R}^d$ containing x such that the function G is linear along $B \cap L_x$, then for each simplex $S \in \mathcal{T}$ there is a point on one of the facets of S where $\varepsilon_{g,G}(S)$ is attained.*

Proof Let $S \in \mathcal{T}$, and let g_S be the linear approximation of G over the simplex S. Furthermore, let $x \in S$ be a point in the interior of the simplex S, and let L_x be a line such that G is linear along $S \cap L_x$. Naturally, g_S is also linear along $S \cap L_x$, which therefore also holds for the function $g_S - G$. Thus, $g_S - G$ attains its minimum on one end point of the line segment

$S \cap L_x$ and its maximum on the other end point. Therefore, the error function $|g_S - G|$ over $S \cap L_x$ attains its maximum, i.e. the maximal approximation error, on one of the facets of S. As $S \in \mathcal{T}$ and $x \in S$ were chosen arbitrarily, this finishes the proof. □

With the help of the above lemma, we can now characterize the approximation error of a bivariate pwl. approximation of F. Note that the following result is well known in the literature. We show it again in order to demonstrate the utility of Lemma 6 in delivering a concise proof.

Lemma 7 ([4, 29, 40, 50]) *Let f be a pwl. approximation f of F and $\mathcal{T}$ its underlying triangulation of D. Then the approximation error $\varepsilon_{f,F}(S)$ over any simplex $S \in \mathcal{T}$ is attained at the centre of one of its facets. Further, if (x_0, y_0) and (x_1, y_1) are the endpoints of a facet over which the approximation error is attained, we have*

$$\varepsilon_{f,F}(S) = \left| E_{f,F}\left(\frac{x_1 - x_0}{2}, \frac{y_1 - y_0}{2}\right)\right| = \frac{1}{4}\left|(x_1 - x_0)(y_1 - y_0)\right|.$$

Proof It is obvious that the prerequisites of Lemma 6 apply to F. In particular, for each point in some simplex $S \in \mathcal{T}$, F is linear along each of the two coordinate axes. Consequently, the approximation error is attained over a facet e of S. We can now parametrize the functions $f_{|e}$ and $F_{|e}$, i.e. the restrictions of f and F onto e, using the convex combination of its endpoints (x_0, y_0) and (x_1, y_1). By writing each point $(x, y) \in e$ as $(x, y) = (x_0, y_0) + (1 - \lambda)(x_1, y_1)$ for some $\lambda \in [0, 1]$, we can express $f_{|e}$, $F_{|e}$ and $E_{f_{|e},F_{|e}}$ as functions in λ:

$$\begin{aligned} f_{|e}(\lambda) &:= \lambda(x_1 y_1 - x_0 y_0) + x_0 y_0, \\ F_{|e}(\lambda) &:= (\lambda(x_1 - x_0) + x_0)(\lambda(y_1 - y_0) + y_0), \\ E_{f_{|e},F_{|e}}(\lambda) &:= F_{|e}(\lambda) - f_{|e}(\lambda) - (\lambda(x_1 - x_0) + x_0)(\lambda(y_1 - y_0) + y_0) \\ &= (-\lambda^2 + \lambda)(x_1 - x_0)(y_1 - y_0). \end{aligned}$$

Lemma 2 implies that the approximation error, i.e. the maximum of the quadratic error function $E_{f_{|e},F_{|e}}$, has a value of

$$\varepsilon_{f,F}(S) = |E_{f_{|e},F_{|e}}(\lambda^*)| = \frac{1}{4}|(x_1 - x_0)(y_1 - y_0)|$$

and is attained at $\lambda^* = 0.5$, corresponding to the centre of e. □

Lemma From 7, we can conclude that the (maximum) error of a bivariate pwl. approximation of F corresponding to a given triangulation of D is always attained at the centre of a facet of one of its simplices. In [29], the author uses this property to formulate the problem to find ε-optimal triangulations as an MIQCQP. To the best of our knowledge, this is the only work considering provably ε-optimal triangulations of the rectangular domain D for the approximation of F. Unfortunately, due to the size of the resulting optimization model, this approach is computationally intractable even for small instances. However, in order to prove that univariate ε-families of triangulations require fewer simplices than any bivariate ε-triangulation for a sufficiently small approximation error ε, it suffices to derive a suitable lower bound for the size of an ε-triangulation. The following lemma gives such a lower bound by using so-called ε-optimal triangles. An *ε-optimal triangle* satisfies a prescribed approximation error bound of ε while taking a maximum possible area. The idea of the following lower bound is to assume that there exists a triangulation consisting exclusively of ε-optimal triangles.

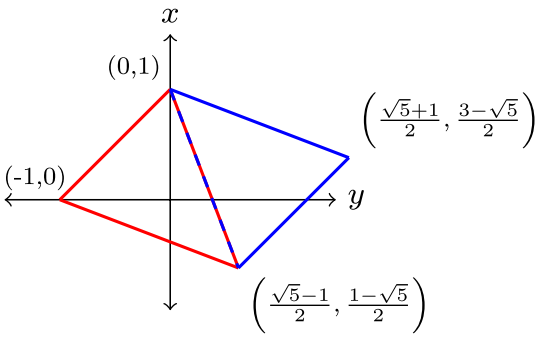


Fig. 2 Optimal triangles with an approximation error of 0.25 which can tile $\mathbb{R}^2$

Lemma 8 ([29]) *An ε-optimal triangulation $\mathcal{T}$ of D for the approximation of F requires at least $\left\lceil \frac{(\bar{x}-\underline{x})(\bar{y}-\underline{y})}{2\sqrt{5}\varepsilon} \right\rceil$ simplices.*

Proof In [40], the authors show with the help of a version of Lemma 7 that the area of an ε-optimal triangle is $2\sqrt{5}\varepsilon$. The area of the rectangular domain D is $(\bar{x}-\underline{x})(\bar{y}-\underline{y})$. Assuming that we can triangulate D solely by ε-optimal triangles, we obtain the indicated lower bound. □

Figure 2 shows two different 0.25-optimal triangles as an example. Together they form a parallelogram. Therefore, copies of the two triangles can be arranged to obtain a triangulation of the plane $\mathbb{R}^2$. However, it is unclear if or how we can use ε-optimal triangles to triangulate polyhedral domains, such as boxes. The problem with using only ε-optimal triangles is their orientation in the plane. Since we want to triangulate an axis-parallel box domain, we have at least four edges that are axis-parallel. However, there is no ε-optimal triangle that has an axis-parallel edge. If a triangle has at least one axis-parallel edge, its maximal area can be at most 4ε instead of $2\sqrt{5}\varepsilon$, as shown in [29]. For more information about ε-optimal triangles, we refer the reader to [4, 40]. For an overview of actual triangulations of box domains to approximate variable products, see [7].

Furthermore, it is easy to see that the lower bound from Lemma 8 is not always tight. From Monsky's Theorem in [34], we know that we cannot triangulate a rectangle with an odd number of simplices such that all simplices have the same area. As a consequence, at least for all values of ε for which the lower bound is an odd number, we need at least one additional simplex than the lower bound suggests.

3.1.3 Comparison of univariate and bivariate approximations

We close Sect. 3.1 by comparing univariate and bivariate approaches with respect to the required number of simplices. Our main result concerning ε-approximations of F then says the following: Via the reformulations Bin1, Bin2 and Bin3 we can always obtain ε-families of triangulations with fewer simplices than any bivariate ε-triangulation, if the approximation accuracy ε is sufficiently small. This finding is formally stated in Theorem 1.

Theorem 1 *For each univariate reformulation* Bin1, Bin2 *and* Bin3*, there exists corresponding thresholds $\varepsilon^{\text{Bin1}}$, $\varepsilon^{\text{Bin2}}$ and $\varepsilon^{\text{Bin3}} > 0$ such that there are $\varepsilon^{\text{Bin1}}$-, $\varepsilon^{\text{Bin2}}$- and $\varepsilon^{\text{Bin3}}$-families of triangulations consisting of fewer simplices than those of any bivariate $\varepsilon^{\text{Bin1}}$-, $\varepsilon^{\text{Bin2}}$- and $\varepsilon^{\text{Bin3}}$-triangulation, respectively.*

Table 3 Comparison of the number of simplices in univariate reformulations and the bivariate lower bound ($D = [0, 2] \times [0, 6]$)

Triangulation	ε	$\lvert\mathcal{T}\rvert$	$\varepsilon_{f,F}(\mathcal{T})$	Triangulation	ε	$\lvert\mathcal{T}\rvert$	$\varepsilon_{f,F}(\mathcal{T})$
Bin1	1.00	4	1.0000	Bivariate	1.00	3	–
(equidistant	0.5	6	0.4444	(Lower bound)	0.5	6	–
triangulations)	0.25	8	0.2500	Lemma 8	0.25	11	–
Lemma 4	0.1	14	0.0816		0.1	27	–
	0.05	18	0.0494		0.05	54	–
Bin2	1.00	10	0.8889	Bin3	1.00	10	0.8889
(equidistant	0.5	14	0.5000	(equidistant	0.5	14	0.5000
triangulations)	0.25	19	0.2500	triangulations)	0.25	19	0.2500
Remark 1	0.1	31	0.0987	Remark 1	0.1	31	0.0987
	0.05	43	0.0473		0.05	43	0.0473

Proof On the one hand, we have established upper bounds on the number of simplices for univariate ε-families of triangulations stated in Lemmas 4 and 5. For ε-families of triangulations in Bin1, Bin2 and Bin3, these bounds grow with $\mathcal{O}(1/\varepsilon)$, cf. Table 2. On the other hand, Lemma 8 gives a lower bound on the number of simplices in any bivariate ε-triangulation. This lower bound in turn increases with a higher rate in $\mathcal{O}(1/\sqrt{\varepsilon})$. Therefore, the desired thresholds $\varepsilon^{\text{Bin1}}$, $\varepsilon^{\text{Bin2}}$ and $\varepsilon^{\text{Bin3}}$ exist. □

For any given ε, we can compare the bounds stated in Table 2 and Lemma 8 respectively in order to determine if univariate or bivariate approximation yields smaller triangulations.

To illustrate Theorem 1, we provide some exemplary numerical results for the concrete domain $D = [0, 2] \times [0, 6]$ in Table 3. We list the numbers of simplices in the triangulations constructed via Lemma 4 for Bin1 and Remark 1 for Bin2 and Bin3 together with the actual approximation error in the columns entitled $|\mathcal{T}|$ and $\varepsilon_{f,F}(\mathcal{T})$, respectively. For the bivariate approximation, we list the lower bounds from Lemma 8.

For all approximation accuracies lower than 0.25, the equidistant pair of triangulations in Bin1 dominates all other triangulations. Further, for the smallest considered approximation accuracy $\varepsilon = 0.05$, all univariate numbers fall below the bivariate lower bound. In particular, Bin1 requires three times less simplices than the bivariate lower bound postulates. This demonstrates the advantage of univariate reformulations for pwl. approximations most clearly.

3.2 Envelopes and strength of the continuous relaxations

An important property of any MIP formulation is the tightness of its continuous relaxation (CR), i.e. the set obtained by relaxing the integrality constraints. Very often, MIP formulations of pwl. functions are used to represent or approximate the non-linear parts of an optimization problem. The usual solution method is then a branch-and-cut approach, in which a continuous relaxation of that problem is solved at each node in the branch-and-bound tree to compute bounds on the objective function value of the optimization problem. In general, a tighter relaxation is more desirable as it yields a smaller branch-and-bound tree, which in turn often leads to shorter computation times. Thus, when comparing MIP formulations for the approximation of $\text{gra}(F)$ it is relevant to study the quality of the respective CRs.

In the following, we compare the bivariate MIP formulation (3) with the univariate MIP formulations (5), (10) and (12), where the latter two are stated explicitly in "Appendix A". Since these MIP formulations require additional auxiliary variables, we compare the quality of their respective continuous relaxation based on the volume of their PCRs, i.e. after projection to the surrounding space of $\mathrm{gra}(F)$. This will lead to two main results. Firstly, we show that the PCR of any bivariate MIP formulation equals $\mathrm{conv}(\mathrm{gra}(F))$. Secondly, we show that the PCRs of univariate MIP formulations are strict relaxations of $\mathrm{conv}(\mathrm{gra}(F))$.

3.2.1 Continuous relaxations of bivariate pwl. approximations

According to Definition 8, the PCR of a sharp MIP formulation actually coincides with the convex hull of the modelled pwl. graph. This means that in this sense, all sharp MIP formulations of a graph are equivalent. Sharpness is a property many well-known MIP formulations fulfil, such as the convex-combination method, the multiple-choice method and the incremental method (see [46]).

In the following, we consider sharp MIP formulations M_f for $\mathrm{gra}(f)$, where f is a bivariate pwl. approximations of F. For these, we show that the PCR $\mathrm{proj}_{(\mathrm{x,y,z})}(C(M_f))$ is not only independent of the chosen MIP formulation, but also independent of the underlying triangulation that defines f. For this purpose, we first recall some important notions concerning the convex and the concave envelope of a given function; see [45] for a more extensive treatment of the subject.

Definition 10 Let $B \subset \mathbb{R}^n$ be a polytope with vertices $V(B)$. We say that a continuous function $G\colon B \to \mathbb{R}$ has a *vertex polyhedral* convex envelope if

$$\mathrm{convenv}_B(G)(x) = \mathrm{convenv}_{V(B)}(G)(x)$$

holds for every $x \in B$. In this case, we also call the function G itself *convex polyhedral*. Analogously, the function G has a *vertex polyhedral* concave envelope if

$$\mathrm{caveenv}_B(G)(x) = \mathrm{caveenv}_{V(B)}(G)(x)$$

holds for every $x \in B$; the function G is then called *concave polyhedral*.

For functions that are convex or concave polyhedral, we can show that this property also carries over to their pwl. approximations. This new result allows us to directly give an algebraic representation of $\mathrm{proj}_{(\mathrm{x,z})} C(M_f)$ from the convex and concave envelope of F.

Lemma 9 *Let $B \subset \mathbb{R}^n$ be a polytope and $G\colon B \to \mathbb{R}$ be a convex (concave) polyhedral function. Further, let g be a pwl. approximation of G over B, defined by a triangulation $\mathcal{T}$. Then* $\mathrm{convenv}_B(g)$ ($\mathrm{caveenv}_B(g)$) *is convex (concave) polyhedral as well and* $\mathrm{convenv}_B(g) = \mathrm{convenv}_B(G)$ *holds.*

Proof It suffices to show the statement for the convex polyhedral case as the concave polyhedral one is analogous. As g is a pwl. approximation of G, we have $g(x) = G(x)$ for all $x \in N(\mathcal{T})$. Since $V(B) \subseteq N(\mathcal{T})$, this implies $\mathrm{convenv}_{V(B)}(G)(x) = \mathrm{convenv}_{V(B)}(g)(x)$ for all $x \in B$.

It remains to show that $g(x) \geq \mathrm{convenv}_{V(B)}(g)(x)$ for all $x \in B$. To this end, let $x \in B$, and let $S \in \mathcal{T}$ be a simplex with vertices $s_0, \ldots, s_n$, chosen such that $x \in S$ holds. Then there exist $\lambda_i \geq 0$, $i = 0, \ldots, n$, such that $x = \sum_{i=0}^n \lambda_i s_i$, with $\sum_{i=0}^n \lambda_i = 1$. Thus, it follows

$$g(x) = \sum_{i=0}^n \lambda_i G(s_i) \geq \sum_{i=0}^n \lambda_i \,\mathrm{convenv}_{V(B)}(G)(s_i)$$

$$\begin{aligned}
&\geq \text{convenv}_{V(B)}(G)\left(\sum_{i=0}^{n}\lambda_i s_i\right)\\
&= \text{convenv}_{V(B)}(G)(x)\\
&= \text{convenv}_{V(B)}(g)(x).
\end{aligned}$$

This results in $\text{convenv}_{V(B)}(g)(x) \leq \text{convenv}_B(g)(x)$. By definition it holds that $\text{convenv}_{V(B)}(g)(x) \geq \text{convenv}_B(g)(x)$, which proves the claim that $\text{convenv}_{V(B)}(g)(x) = \text{convenv}_B(g)(x)$. □

This leads to the following central result for pwl. approximations f of F. It says that the PCR of (3) is (i) independent of the actual choice of f and (ii) independent of the MIP formulation modelling $\text{gra}(f)$ as long as the MIP formulation is sharp.

Theorem 2 *Let f be a pwl. approximation of F, and let $M_f \subset \mathbb{R}^{2+1} \times [0,1]^p \times \{0,1\}^q$ be a sharp MIP formulation for* $\text{gra}(f)$. *Then we have*

$$\text{proj}_{(x,z)}\, C(M_f) = \text{conv}(\text{gra}(F)).$$

Proof In [43, Remark 1.3], it is shown that multi-linear functions on boxes are both convex and concave polyhedral. Thus, F has a vertex polyhedral convex and concave envelope. By Lemma 9, every pwl. approximation f of F is also convex and concave polyhedral. In addition, $F(x, y) = f(x, y) = xy$ holds for all $(x, y) \in V(D)$. It follows that

$$\text{convenv}(F) = \text{convenv}_{V(D)}(F) = \text{convenv}_{V(D)}(f) = \text{convenv}(f)$$

and

$$\text{caveenv}(F) = \text{caveenv}_{V(D)}(F) = \text{caveenv}_{V(D)}(f) = \text{caveenv}(f),$$

and therefore

$$\text{conv}(\text{gra}(F)) = \text{conv}(\text{gra}(f)).$$

From the sharpness of M_f for $\text{gra}(f)$, we can conclude that

$$\text{proj}_{(x,z)}\, C(M_f) = \text{conv}(\text{gra}(f)) = \text{conv}(\text{gra}(F)),$$

which completes the proof. □

From the literature, $\text{conv}(\text{gra}(F))$ is known as the *McCormick relaxation* of F (cf. [32]). It is defined by the two functions $C^L\colon D \to \mathbb{R}$ and $C^U\colon D \to \mathbb{R}$ with

$$\begin{aligned}
C^L(x, y) &:= \text{convenv}(F)(x, y) = \max\{\underline{y}x + \underline{x}y - \underline{x}\,\underline{y},\ \bar{y}x + \bar{x}y - \bar{x}\bar{y}\},\\
C^U(x, y) &:= \text{caveenv}(F)(x, y) = \min\{\underline{y}x + \bar{x}y - \bar{x}\,\underline{y},\ \bar{y}x + \underline{x}y - \underline{x}\bar{y}\}.
\end{aligned}$$

The McCormick relaxation is the tightest relaxation of $\text{gra}(F)$ that any MIP formulation can obtain. In the following remark, we discuss how the relaxation of bivariate MIP formulations can be tightened when additional restrictions are added for x and y.

Remark 2 We consider the special case where D is intersected with a compact set $Z \in \mathbb{R}^2$. This might be the case if F occurs as a term in the objective function or a constraint of an optimization problem. For this case, the set Z can model a large variety of possible constraints involving the variables x and y. We know the following:

$$\text{conv}(\text{gra}_{D\cap Z}(f)) \subseteq \text{conv}(\text{gra}(f)) \cap (Z \times \mathbb{R})$$

Table 4 Envelopes and PCRs in univariate reformulations Bin1, Bin2 and Bin3

Model	Convex envelopes as functions $D \to \mathbb{R}$
Bin1	$C_1^L(x,y) = \frac{1}{4}((x+y)^2 - (\bar{x}+\underline{x}-\bar{y}-\underline{y})(x-y) + (\underline{x}-\bar{y})(\bar{x}-\underline{y}))$
Bin2	$C_2^L(x,y) = \frac{1}{2}((x+y)^2 - (\bar{x}+\underline{x})x + \bar{x}\underline{x} - (\bar{y}+\underline{y})y + \bar{y}\underline{y})$
Bin3	$C_3^L(x,y) = \frac{1}{2}(x^2+y^2 - (\bar{x}+\underline{x}-\bar{y}-\underline{y})(x-y) + (\underline{x}-\bar{y})(\bar{x}-\underline{y}))$
	Concave envelopes as functions $D \to \mathbb{R}$
Bin1	$C_1^U(x,y) = \frac{1}{4}((\underline{x}+\underline{y}+\bar{x}+\bar{y})(x+y) - (\underline{x}+\underline{y})(\bar{x}+\bar{y}) - (x-y)^2)$
Bin2	$C_2^U(x,y) = \frac{1}{2}((\underline{x}+\underline{y}+\bar{x}+\bar{y})(x+y) - (\underline{x}+\underline{y})(\bar{x}+\bar{y}) - x^2 - y^2)$
Bin3	$C_3^U(x,y) = \frac{1}{2}((\underline{x}+\bar{x})x - \underline{x}\bar{x} + (\underline{y}+\bar{y})y - \underline{y}\bar{y} - (x-y)^2)$
	PCRs
Bin1	$\text{proj}_{(x,y,z)}\, C(M_{f^{\text{Bin1}}}) = \{(x,y,z) \in D \times \mathbb{R} : C_1^L(x,y) \le z \le C_1^U(x,y)\}$
Bin2	$\text{proj}_{(x,y,z)}\, C(M_{f^{\text{Bin2}}}) = \{(x,y,z) \in D \times \mathbb{R} : C_2^L(x,y) \le z \le C_2^U(x,y)\}$
Bin3	$\text{proj}_{(x,y,z)}\, C(M_{f^{\text{Bin3}}}) = \{(x,y,z) \in D \times \mathbb{R} : C_3^L(x,y) \le z \le C_3^U(x,y)\}$

$$= \text{proj}_{(x,z)}\, C(M_f) \cap (Z \times \mathbb{R})$$
$$= \{(x,z) \in \mathbb{R}^{n+1} : (x,z,\lambda,u) \in C(M_f), x \in Z\}.$$

This means that the PCR of M_f restricted to $D \cap Z$ can potentially be tightened by adding additional constraints. See [2], where the authors consider the set $Z := \{(x,y) \in \mathbb{R}^2 \mid xy \le u\}$ for some $u \in \mathbb{R}$ and derive $\text{conv}(\text{gra}_{D\cap Z}(f))$ by adding additional linear and conic constraints to $\text{conv}(\text{gra}(f)) \cap (Z \times \mathbb{R})$.

3.2.2 Continuous relaxations of univariate pwl. approximations

We now turn to the PCRs of sharp univariate MIP formulations as in (5), (10) and (12). We point out that univariate reformulations are described by separable functions over rectangular domains. Such functions are known to be *sum decomposable*; see [45]. This means that the envelopes of separable functions are determined by the sum of the envelopes of their univariate summands; see also [19]. As a consequence of this, the convex and concave envelopes of pwl. univariate approximations of F, and thus the PCRs of the corresponding MIP formulations, depend on both the choice of the univariate reformulation and the chosen triangulations defining the pwl. approximations. The dependency on the triangulations is in contrast to the result we had in the bivariate case. The consequence is that the tightness of the PCR is influenced by the approximation error and thus depends on the number and placement of the vertices of the triangulations. For further details we refer to [9], where the effects of the approximation error on PCRs are discussed, and neglect the approximation error in the following. We rather assume that the approximation error is sufficiently small so that it does not interfere with the comparison of the PCRs. Consequently, we focus on the envelopes that we obtain from the non-linear univariate reformulations Bin1, Bin2 and Bin3, i.e. (4), (9) and (11).

Note that each of the univariate reformulation Bin1, Bin2, and Bin3 is a sum of quadratic functions which are all either convex or concave. The convex (concave) envelope of each convex (concave) summand is the convex (concave) function itself. In contrast, a convex (concave) function is vertex polyhedral; its concave (convex) envelope is therefore given as the linear interpolant which uses the domain bounds as vertices. In Table 4, we list the

convex and concave envelopes of the pwl. approximations $f^{\text{Bin1}}\colon D \to \mathbb{R}$, $f^{\text{Bin2}}\colon D \to \mathbb{R}$ and $f^{\text{Bin3}}\colon D \to \mathbb{R}$ of F that we obtain by exploiting sum decomposability as explained above.

We emphasize that these envelopes are strict under- resp. overestimators of F and thus only give a relaxation of $\text{conv}(\text{gra}(F))$ in the sense of Eq. (2). Further, we also state the respective PCRs in Table 4. The following proposition compares the volumes of these PCRs. It states that among the three univariate reformulations, the PCR $\text{proj}_{(\text{x,z})}\, C(M_{f^{\text{Bin1}}})$ is a strictly tighter relaxation of $\text{gra}(F)$ than $\text{proj}_{(\text{x,z})}\, C(M_{f^{\text{Bin2}}})$ and $\text{proj}_{(\text{x,z})}\, C(M_{f^{\text{Bin3}}})$, which coincide in terms of volume.

Lemma 10 *The volumes* V^D_{Bin1}, V^D_{Bin2} *and* V^D_{Bin3} *of the projections* $\text{proj}_{(\text{x,y,z})}\, C(M_{f^{\text{Bin1}}})$, $\text{proj}_{(\text{x,y,z})}\, C(M_{f^{\text{Bin2}}})$ *and* $\text{proj}_{(\text{x,y,z})}\, C(M_{f^{\text{Bin3}}})$, *respectively, form the following hierarchy:*

$$V^D_{\text{Bin1}} < V^D_{\text{Bin2}} = V^D_{\text{Bin3}}.$$

Proof For the volumes of the two projections $\text{proj}_{(\text{x,y,z})}\, C(M_{f^{\text{Bin2}}})$ and $\text{proj}_{(\text{x,y,z})}\, C(M_{f^{\text{Bin3}}})$, we have

$$\begin{aligned} V^D_{\text{Bin2}} := V(\text{proj}_{(\text{x,y,z})}\, C(M_{f^{\text{Bin2}}})) &= \int_{\underline{y}}^{\bar{y}} \int_{\underline{x}}^{\bar{x}} C_2^U(x,y) - C_2^L(x,y)\, dx\, dy \\ &= \int_{\underline{y}}^{\bar{y}} \int_{\underline{x}}^{\bar{x}} C_3^U(x,y) - C_3^L(x,y)\, dx\, dy \\ &= V(\text{proj}_{(\text{x,y,z})}\, C(M_{f^{\text{Bin3}}})) =: V^D_{\text{Bin3}}. \end{aligned}$$

Both volumes of V^D_{Bin2} and V^D_{Bin3} are given by

$$V^D_{\text{Bin2}} = \frac{1}{12}(\bar{x}-\underline{x})(\bar{y}-\underline{y})\left(2(\bar{x}-\underline{x})^2 + 3(\bar{x}-\underline{x})(\bar{y}-\underline{y}) + 2(\bar{y}-\underline{y})^2\right). \tag{7}$$

The volume of the projection $\text{proj}_{(\text{x,y,z})}\, C(M_{f^{\text{Bin1}}})$ is given as

$$\begin{aligned} V^D_{\text{Bin1}} := V(\text{proj}_{(\text{x,y,z})}\, C(M_{f^{\text{Bin1}}})) &= \int_{\underline{y}}^{\bar{y}} \int_{\underline{x}}^{\bar{x}} C_1^U(x,y) - C_1^L(x,y)\, dx\, dy \\ &= \frac{1}{12}(\bar{x}-\underline{x})(\bar{y}-\underline{y})\left((\underline{x}-\underline{x}) + 3(\underline{x}^2-\underline{x})(\bar{y}-\underline{y}) + (\bar{y}-\underline{y})^2\right). \end{aligned}$$

Together with (7), we obtain

$$V^D_{\text{Bin2}} - V^D_{\text{Bin1}} = V^D_{\text{Bin3}} - V^D_{\text{Bin1}} = \frac{1}{12}(\underbrace{(\bar{x}-\underline{x})^2}_{>0} + \underbrace{(\bar{y}-\underline{y})^2}_{>0}) > 0,$$

which completes the proof. □

3.2.3 Comparison of the univariate and bivariate continuous relaxations

We now compare the PCRs that result from the univariate and bivariate MIP formulations. The following theorem says that the PCRs of the univariate MIP formulations always yield looser relaxations of $\text{gra}(F)$ than the PCR of a bivariate MIP formulation.

Theorem 3 *The PCRs of the MIP formulations in the reformulations* Bin1, Bin2 *and* Bin3 *are looser relaxations of* $\text{gra}(F)$ *than the PCR of a bivariate MIP formulation. In particular, the following applies:*

The volumes V^D_{Bin1}, V^D_{Bin2} *and* V^D_{Bin3} *of the PCRs* $\text{proj}_{(x,y,z)}\, C(M_{f^{\text{Bin1}}})$, $\text{proj}_{(x,y,z)}\, C(M_{f^{\text{Bin2}}})$ *and* $\text{proj}_{(x,y,z)}\, C(M_{f^{\text{Bin3}}})$ *are larger than the volume* V^D_{McC} *of the PCR* $\text{proj}_{(x,y,z)}\, C(M_f)$.

Proof From Theorem 2, we know that $\text{proj}_{(x,y,z)}\, C(M_f)$ of a bivariate MIP formulation M_f is equivalent to the McCormick relaxation. The volume of the McCormick relaxation is given by

$$V^D_{McC} := \int_{\underline{y}}^{\bar{y}} \int_{\underline{x}}^{\bar{x}} C^U(x, y)\, dx\, dy - \int_{\underline{y}}^{\bar{y}} \int_{\underline{x}}^{\bar{x}} C^L(x, y)\, dx\, dy = \frac{1}{6}(\bar{x} - \underline{x})^2(\bar{y} - \underline{y})^2.$$

Further, we know from Lemma 10 that Bin1 provides the tightest CR among the univariate reformulations. It now holds that the difference of these two volumes is always greater than zero, i.e.

$$V^D_{\text{Bin1}} - V^D_{McC} = \frac{1}{12}(\underbrace{(\bar{x} - \underline{x})^2}_{>0} + \underbrace{(\bar{y} - \underline{y})^2}_{>0} + \underbrace{(\bar{x} - \underline{x})(\bar{y} - \underline{y})}_{>0}) > 0.$$

Thus,

$$V^D_{\text{Bin2}} = V^D_{\text{Bin3}} > V^D_{McC}.$$

also holds. □

To quantify this downside of the univariate MIP formulations, we calculate the ratio between the volume of their PCRs to the volume of $\text{conv}(\text{gra}(F))$. We denote the ratios by

$$R^D_{\text{Bin1}} := \frac{V^D_{\text{Bin1}}}{V^D_{McC}} = \frac{(\bar{x} - \underline{x})^2 + (\bar{y} - \underline{y})^2}{2(\bar{x} - \underline{x})(\bar{y} - \underline{y})} + \frac{3}{2},$$

$$R^D_{\text{Bin2}} := \frac{V^D_{\text{Bin2}}}{V^D_{McC}} = \frac{(\bar{x} - \underline{x})^2 + (\bar{y} - \underline{y})^2}{(\bar{x} - \underline{x})(\bar{y} - \underline{y})} + \frac{3}{2},$$

$$R^D_{\text{Bin3}} := \frac{V^D_{\text{Bin3}}}{V^D_{McC}} = \frac{(\bar{x} - \underline{x})^2 + (\bar{y} - \underline{y})^2}{(\bar{x} - \underline{x})(\bar{y} - \underline{y})} + \frac{3}{2}.$$

Obviously, the ratios R^D_{Bin1}, R^D_{Bin2} and R^D_{Bin3} are invariant under axial shifts of the domain D. This means that the ratios depend only on the length of the axes $(\bar{x} - \underline{x})$ and $(\bar{y} - \underline{y})$. In Fig. 3, we plot R^D_{Bin1}, R^D_{Bin2} and R^D_{Bin3} with respect to the elongation and scaling of the domain by varying $(\bar{x} - \underline{x})$ and $(\bar{y} - \underline{y})$. In accordance with Theorem 3, Bin1 always yields a better ratio than either of Bin2 or Bin3. Furthermore, it is noteworthy that the more rectangularly stretched D is, the worse the ratios of the univariate reformulations become. The ratios start from 2.5 (Bin1) and 3.5 (Bin2, Bin3) on the quadratic domain $D = [0, 1] \times [0, 1]$ and then increase towards infinity as the domain becomes more rectangular.

To illustrate the shapes of the different PCRs, we have plotted them exemplarily for the quadratic domain $D = [0, 1] \times [0, 1]$ in Fig. 4. Although the volumes V^D_{Bin2} and V^D_{Bin3} are the same, it can be shown that C^L_2 is a tighter convex underestimator for F over D than C^L_3. The opposite is true for the concave overestimators, where C^U_3 is a tighter convex overestimator than C^U_2. These observations are of particular interest in the context of an optimization problem. If for example F appears in the objective function of a minimization problem, Bin2 gives a tighter convex underestimator, while Bin3 gives a tighter convex overestimator if F instead appears in the objective function of a maximization problem. However, this clear hierarchy does not hold for Bin1, which yields tighter or less tight relaxations than Bin2 or

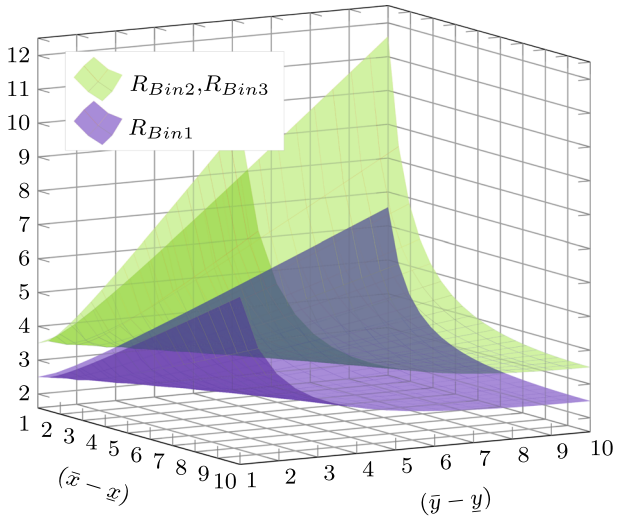


Fig. 3 Volume ratios between univariate PCRs and the McCormick relaxation of $F(x, y) = xy$ over $D = [\underline{x}, \bar{x}] \times [\underline{y}, \bar{y}]$

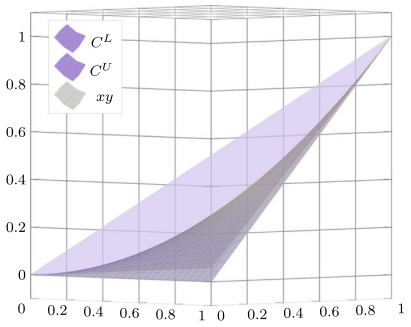
Bin3 depending on the elongation of the domain and the optimization sense. Formal proofs of these hierarchical observations are given in Section A.1.

3.3 Discussion and guidelines for practice

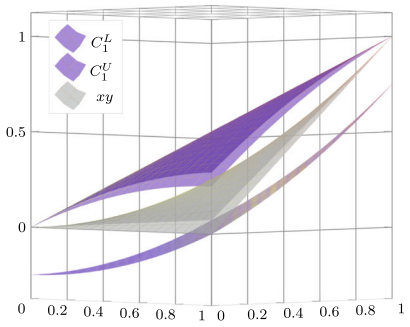
In Sect. 3.1, we have shown that univariate MIP formulations are superior to bivariate MIP formulations when it comes to the size of the underlying triangulation required to attain a certain high approximation accuracy for F. However, this is in part bought by the fact that their corresponding PCRs are looser, as we showed in Sect. 3.2. In this section, we discuss some consequences of these observations for the practical use of pwl. approximations in the modelling of optimization problems.

On the one hand, a bivariate MIP formulation is favourable if we are interested in obtaining good dual bounds for a pwl. approximation of a given non-convex MIQCQPs early in the solution process, for example. This is mainly because in the root node it yields the best possible linear-programming (LP) bound as its PCR equals the McCormick envelope, independent of the number of simplices used, as we showed in Theorem 2. In contrast, in Theorem 3 we have proved that the PCR of any univariate MIP formulation is looser than the bivariate PCR. Therefore, the initial LP bound at the root node is weaker.

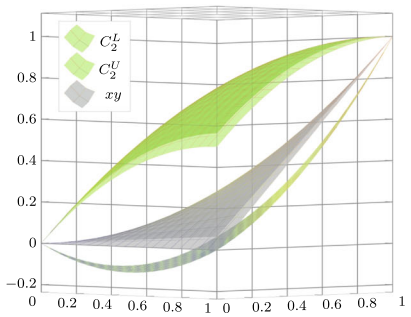
On the other hand, if instead the optimal solution of a high-accuracy MIP approximation of a certain MIQCQP is required, the results of Sect. 3.1 suggest to pursue a univariate reformulation scheme, as it requires less simplices to obtain an ε-approximations for some prescribed guarantee ε. To compensate for the disadvantage of looser PCRs in this case, we can easily tighten the univariate reformulation by incorporating a univariate variant of the well-known *McCormick cuts*, which are known to completely describe the convex hull of F. To this end, we can simply replace the term xy in the corresponding univariate reformulation of the constraint at hand. We exemplarily state the resulting version of the McCormick cuts



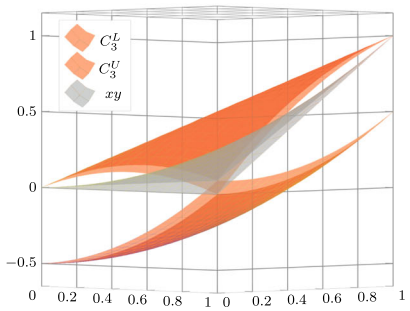
(a) McCormick relaxation over quadratic domain.



(b) Bin1: PCR over quadratic domain.



(c) Bin2: PCR over quadratic domain.



(d) Bin3: PCR over quadratic domain.

Fig. 4 PCRs of univariate and bivariate MIP formulations

for the reformulation Bin1, which uses the substitution $z = p_1^2 - p_2^2$:

$$\begin{aligned} z &\geq \underline{x}y + x\underline{y} - \underline{x}\underline{y}, \\ z &\geq \bar{x}y + x\bar{y} - \bar{x}\bar{y}, \\ z &\leq \bar{x}y + x\underline{y} - \bar{x}\underline{y}, \\ z &\leq \underline{x}y + x\bar{y} - \underline{x}\bar{y}. \end{aligned} \tag{8}$$

For Bin2 and Bin3, the corresponding McCormick cuts are straightforward to compute as well. With an increasing prescribed accuracy of a pwl. approximation, a bivariate approach requires unproportionally more simplices and consequently binary variables. Hence, a univariate reformulation approach together with the addition of the four inequalities (8) quickly becomes the cheaper alternative in terms of complexity. This recommendation is in line with the results of [1], where pwl. approximations are utilized to solve MINLPs arising in the context of alternating current optimal power flow. The authors reformulate the bilinear terms in their original model for the problem by the univariate reformulation Bin2. Additionally, they add the reformulated McCormick cuts shown in (8). It turns out that the resulting univariate model is solved much faster than the bivariate one, while the solutions of both models are of the same approximation quality. To the best of our knowledge, the authors of [1] are the first who use such a univariate reformulation enhanced with additional cutting planes.

Although the figures stated in Table 3 suggest that Bin1 compares favourably to Bin2 and Bin3 in terms of the number of required simplices, the structure of the constraint set of the considered optimization problem is crucial. If, for instance, bounds for the term $x - y$ are known a priori, for example inferred from the problem data, using Bin3 can be advantageous (cf. [50]). The same holds for Bin2, if bounds for the term $x + y$ are available. Moreover, in case that for a subset $x_1, x_2, \ldots, x_n$ of the variables at hand many of the bilinear terms $x_i x_j$ with $i, j \in \{1, 2, \ldots, n\}$ occur in the constraints of the problem, using Bin2 or Bin3 can again be beneficial. The reason for this is the following general observation. If the same non-linear function G occurs multiple times in an optimization problem (except for linear factors), we can replace this function with the same variable $\tilde{g}$ everywhere in the model and add the constraint $\tilde{g} = G$ only once. This way, we need only one pwl. approximation for all occurrences of G. Thus, if we reformulate the terms $x_i x_j$ via Bin2 or Bin3, for each of the $\mathcal{O}(n)$ many quadratic monomials x_i^2 and x_j^2 only one pwl. approximation has to be constructed. Apart from this, we only need one pwl. approximation for each of the $\mathcal{O}(n^2)$-many $p_{i,j}^2 = (x_i + x_j)^2$. In case of Bin1, however, we need two different pwl. approximations for each of the $\mathcal{O}(n^2)$-many $p_{1,i,j,}^2 = (\frac{1}{2}(x_i + x_j))^2$ and $p_{2,i,j}^2 = (\frac{1}{2}(x_i - x_j))^2$.

4 Conclusion and discussion

In this paper, we studied MIP formulations for pwl. approximations of bilinear terms in optimization models. More precisely, we compared MIP formulations for direct bivariate pwl. approximations of variable products to MIP formulations for pwl. approximations after univariate reformulations with respect to two different metrics of efficiency. First, we proved that for a sufficiently small prescribed approximation error ε, all considered univariate reformulations allow more compact ε-approximations than any bivariate ε-approximation requires – as measured by the number of simplices in the underlying triangulation. In this sense, concerning the size of the resulting pwl. approximations, and consequently the required number of binary variables, our results are a strong indication for using univariate reformulations in optimization problems. Second, we showed that, in contrast, all univariate reformulations lead to genuinely weaker continuous relaxations than bivariate MIP formulations. These two opposing characteristics of the respective MIP formulations explain many of the mixed computational results found in the literature. Finally, we discussed our theoretical results with regard to their application in practice. Notably, the looser relaxations of the univariate reformulation approaches can be improved to equal those of a bivariate pwl. approximation by adding linear cutting planes, the so-called McCormick cuts. A first algorithmic approach constructed in this fashion can already be found in the literature ([1]), reporting very good computational results for the considered application. In this way, the authors profit from compact MIP formulations as well as from tight relaxations at the same time. Both our theoretical results and these first empirical evidence indicate that it would be promising to study generic algorithms for MIQCQPs based on univariate reformulations as part of future research on the topic.

Acknowledgements This research was supported by the Bavarian Ministry of Economic Affairs, Regional Development and Energy through the Center for Analytics – Data – Applications (ADA-Center) within the framework of "BAYERN DIGITAL II" (20-3410-2-9-8). Furthermore, this research has been performed as part of the Energie Campus Nürnberg and is supported by funding of the Bavarian State Government. Moreover, we thank the DFG for their support within Projects B07 and B08 in CRC TRR 154. Last but not least, we

thank the anonymous referees for their insightful comments, which led to a substantial improvement of the paper.

Funding Open Access funding enabled and organized by Projekt DEAL.

Appendix A: MIP formulations

In this part, we derive the MIP formulations of the pwl. approximation in the reformulations Bin2, Bin3, and Ln. We proceed analogously to reformulation Bin1 in Sect. 3.

We start with reformulation Bin2:

$$\operatorname{gra}(F) = \left\{(x, y, p^2 - x^2 - y^2) \in \mathbb{R}^3 \mid p = x + y,\ (x, y) \in D\right\}. \tag{9}$$

Now, let $f_1^{\text{Bin2}}\colon [\underline{x}, \bar{x}] \to \mathbb{R}$ be a pwl. approximation of x^2 with triangulation $\mathcal{T}_1^{\text{Bin2}}$, $f_2^{\text{Bin2}}\colon [\underline{y}, \bar{y}] \to \mathbb{R}$ a pwl. approximation of y^2 with triangulation $\mathcal{T}_2^{\text{Bin2}}$, and $f_3^{\text{Bin2}}\colon [\underline{x} + \underline{y}, \bar{x} + \bar{y}] \to \mathbb{R}$ a pwl. approximation of p^2 with triangulation $\mathcal{T}_3^{\text{Bin2}}$.

We can model an approximation of $\operatorname{gra}(F)$ by $f^{\text{Bin2}}\colon D \to \mathbb{R}$,

$$\begin{aligned} f^{\text{Bin2}}(x, y) &= \frac{1}{2}(f_3^{\text{Bin2}}(p) - f_1^{\text{Bin2}}(x) - f_2^{\text{Bin2}}(y)), \\ p &= x + y. \end{aligned}$$

Further, let $M_1^{\text{Bin2}} \subseteq D \times \mathbb{R} \times [0,1]^{p_1^{\text{Bin2}}} \times \{0,1\}^{q_1^{\text{Bin2}}}$, $M_2^{\text{Bin2}} \subseteq D \times \mathbb{R} \times [0,1]^{p_2^{\text{Bin2}}} \times \{0,1\}^{q_2^{\text{Bin2}}}$ and $M_3^{\text{Bin2}} \subseteq D \times \mathbb{R} \times [0,1]^{p_3^{\text{Bin2}}} \times \{0,1\}^{q_3^{\text{Bin2}}}$ be sharp MIP formulations of the graphs $\operatorname{gra}(f_1^{\text{Bin2}})$, $\operatorname{gra}(f_2^{\text{Bin2}})$ and $\operatorname{gra}(f_3^{\text{Bin2}})$. We can then model an approximation of $\operatorname{gra}(F)$ as:

$$\operatorname{gra}(f^{\text{Bin2}}) = \{(x, y, z) \in \mathbb{R}^3 \mid (x, y, z, \lambda_1, u_1, \lambda_2, u_2, \lambda_3, u_3) \in M_{f^{\text{Bin2}}}\} \tag{10}$$

together with the MIP formulation.

$$\begin{aligned} M_{f^{\text{Bin2}}} := \{&(x, y, z, \lambda_1, u_1, \lambda_2, u_2, \lambda_3, u_3) \in D \times \mathbb{R} \times [0,1]^{p_1^{\text{Bin2}}} \times \{0,1\}^{q_1^{\text{Bin2}}} \\ &\times [0,1]^{p_2^{\text{Bin2}}} \times \{0,1\}^{q_2^{\text{Bin2}}} \times [0,1]^{p_3^{\text{Bin2}}} \times \{0,1\}^{q_3^{\text{Bin2}}} \mid \\ &(p_1, z_1, \lambda_1, u_1) \in M_1^{\text{Bin2}},\ (p_2, z_2, \lambda_2, u_2) \in M_2, \\ &(p_3, z_3, \lambda_3, u_3) \in M_3^{\text{Bin2}}, \\ &z = \frac{1}{2}(z_1 - z_2 - z_3),\ p = x + y,\ (x, y) \in D\} \end{aligned}$$

Next, we apply Bin3:

$$\operatorname{gra}(F) = \left\{(x, y, x^2 + y^2 - p^2) \in \mathbb{R}^3 \mid p = x - y,\ (x, y) \in D\right\}. \tag{11}$$

Now, let $f_1^{\text{Bin3}}\colon [\underline{x}, \bar{x}] \to \mathbb{R}$ be a pwl. approximation of x^2 with triangulation $\mathcal{T}_1^{\text{Bin3}}$, $f_2^{\text{Bin3}}\colon [\underline{y}, \bar{y}] \to \mathbb{R}$ a pwl. approximation of y^2 with triangulation $\mathcal{T}_2^{\text{Bin3}}$, and $f_3^{\text{Bin3}}\colon [\underline{x} - \bar{y}, \bar{x} - \underline{y}] \to \mathbb{R}$ a pwl. approximation of p^2 with triangulation $\mathcal{T}_3^{\text{Bin3}}$.

We can model an approximation of (4) by $f^{\text{Bin3}}\colon D \to \mathbb{R}$,

$$\begin{aligned} f^{\text{Bin3}}(x, y) &= \frac{1}{2}(f_1^{\text{Bin3}}(x) + f_2^{\text{Bin3}}(y) - f_3^{\text{Bin3}}(p)), \\ p &= x - y. \end{aligned}$$

Further, let $M_1^{\text{Bin3}} \subseteq D \times \mathbb{R} \times [0,1]^{p_1^{\text{Bin3}}} \times \{0,1\}^{q_1^{\text{Bin3}}}$, $M_2^{\text{Bin3}} \subseteq D \times \mathbb{R} \times [0,1]^{p_2^{\text{Bin3}}} \times \{0,1\}^{q_2^{\text{Bin3}}}$ and $M_3^{\text{Bin3}} \subseteq D \times \mathbb{R} \times [0,1]^{p_3^{\text{Bin3}}} \times \{0,1\}^{q_3^{\text{Bin3}}}$ be sharp MIP formulations of the graphs $\text{gra}(f_1^{\text{Bin3}})$, $\text{gra}(f_2^{\text{Bin3}})$ and $\text{gra}(f_3^{\text{Bin3}})$. We can model an approximation of $\text{gra}(F)$ as:

$$\text{gra}(f^{\text{Bin3}}) = \{(x, y, z) \in \mathbb{R}^3 \mid (x, y, z, \lambda_1, u_1, \lambda_2, u_2, \lambda_3, u_3) \in M_{f^{\text{Bin3}}}\} \tag{12}$$

together with the MIP formulation.

$$\begin{aligned} M_{f^{\text{Bin3}}} := \{&(x, y, z, \lambda_1, u_1, \lambda_2, u_2, \lambda_3, u_3) \in D \times \mathbb{R} \times [0,1]^{p_1^{\text{Bin3}}} \times \{0,1\}^{q_1^{\text{Bin3}}} \\ &\times [0,1]^{p_2^{\text{Bin3}}} \times \{0,1\}^{q_2^{\text{Bin3}}} \times [0,1]^{p_3^{\text{Bin3}}} \times \{0,1\}^{q_3^{\text{Bin3}}} \mid \\ &(p_1, z_1, \lambda_1, u_1) \in M_1^{\text{Bin3}},\ (p_2, z_2, \lambda_2, u_2) \in M_2^{\text{Bin3}}, \\ &(p_3, z_3, \lambda_3, u_3) \in M_3^{\text{Bin3}}, \\ &z = \frac{1}{2}(z_1 + z_2 - z_3),\ p = x - y,\ (x, y) \in D\} \end{aligned}$$

Finally, we apply Ln:

$$\text{gra}(F) = \left\{(x, y, p) \in \mathbb{R}^3 \mid \ln(p) = \ln(x) + \ln(y),\ (x, y) \in D\right\}.$$

Now, let $f_1^{\text{Ln}}\colon [\underline{x}, \bar{x}] \to \mathbb{R}$ be a pwl. approximation of $\ln(x)$ with triangulation $\mathcal{T}_1^{\text{Ln}}$, $f_2^{\text{Ln}}\colon [\underline{y}, \bar{y}] \to \mathbb{R}$ a pwl. approximation of $\ln(y)$ with triangulation $\mathcal{T}_2^{\text{Ln}}$, and $f_3^{\text{Ln}}\colon [\underline{x}\,\underline{y}, \bar{x}\,\bar{y}] \to \mathbb{R}$ a pwl. approximation of $\ln(p)$ with triangulation $\mathcal{T}_3^{\text{Ln}}$.

We can model an approximation of $\text{gra}(F)$ by $f^{\text{Ln}}\colon D \to \mathbb{R}$,

$$\begin{aligned} f^{\text{Ln}}(x, y) &= p, \\ f_3^{\text{Ln}}(p) &= f_1^{\text{Ln}}(x) + f_2^{\text{Ln}}(y),\ (x, y) \in D. \end{aligned}$$

Further, let $M_1^{\text{Ln}} \subseteq D \times \mathbb{R} \times [0,1]^{p_1^{\text{Bin2}}} \times \{0,1\}^{q_1^{\text{Bin2}}}$, $M_2^{\text{Ln}} \subseteq D \times \mathbb{R} \times [0,1]^{p_2^{\text{Bin2}}} \times \{0,1\}^{q_2^{\text{Bin2}}}$ and $M_3^{\text{Ln}} \subseteq D \times \mathbb{R} \times [0,1]^{p_3^{\text{Bin2}}} \times \{0,1\}^{q_3^{\text{Bin2}}}$ be sharp MIP formulations of the graphs $\text{gra}(f_1^{\text{Ln}})$, $\text{gra}(f_2^{\text{Ln}})$ and $\text{gra}(f_3^{\text{Ln}})$. We can model an approximation of $\text{gra}(F)$ as:

$$\text{gra}(f^{\text{Ln}}) = \{(x, y, z) \in \mathbb{R}^3 \mid (x, y, z, \lambda_1, u_1, \lambda_2, u_2, \lambda_3, u_3) \in M_{f^{\text{Ln}}}\}$$

together with the MIP formulation.

$$\begin{aligned} M_{f^{\text{Ln}}} := \{&(x, y, z, \lambda_1, u_1, \lambda_2, u_2, \lambda_3, u_3) \in D \times \mathbb{R} \times [0,1]^{p_1^{\text{Ln}}} \times \{0,1\}^{q_1^{\text{Ln}}} \times \\ &[0,1]^{p_2^{\text{Ln}}} \times \{0,1\}^{q_2^{\text{Ln}}} \times [0,1]^{p_3^{\text{Ln}}} \times \{0,1\}^{q_3^{\text{Ln}}} \mid \\ &(p_1, z_1, \lambda_1, u_1) \in M_1^{\text{Ln}},\ (p_2, z_2, \lambda_2, u_2) \in M_2^{\text{Ln}}, \\ &(p_3, z_3, \lambda_3, u_3) \in M_3^{\text{Ln}},\ z_3 = z_1 + z_2,\ (x, y) \in D\}. \end{aligned}$$

A.1: A hierarchy of convex underestimators

In the following, we derive a hierarchy for the convex underestimators that result from the continuous relaxations of the univariate reformulations (see Table 4). The following results are useful, for example, if F occurs as a term in the objective function to be minimized in some optimization problem. This is because the choice of convex underestimators determines the tightness of the resulting continuous relaxation (while the overestimators of F are not relevant due to the optimization sense).

We start by comparing the convex underestimators C_1^L with C_3^L, belonging to Bin1 and Bin3 respectively.

Proposition 1 *The convex envelope $C_1^L : D \to \mathbb{R}$ resulting from the univariate reformulation* Bin1 *is a tighter convex underestimator of F over D than the convex envelope $C_3^L : D \to \mathbb{R}$ resulting from the univariate reformulation* Bin3*, i.e. we have*

$$C_1^L(x, y) - C_3^L(x, y) \geq 0 \quad \forall (x, y) \in D,$$

and there exists a point $(x, y) \in D$ with

$$C_1^L(x, y) - C_3^L(x, y) > 0.$$

Proof We note that the first condition is equivalent to proving that the optimal objective value of the maximization problem

$$\max_{(x,y)\in D} C_{31}(x, y), \tag{13}$$

with $C_{31} : D \to \mathbb{R}$ and

$$\begin{aligned} C_{31}(x, y) &:= 4(C_3^L(x, y) - C_1^L(x, y)) \\ &= (x - y)^2 - (\bar{x} + \underline{x} - \bar{y} - \underline{y})(x - y) + (\underline{x} - \bar{y})(\bar{x} - \underline{y}), \end{aligned}$$

is less than or equal to 0, which we do in the following.

In Problem (13), we maximize a univariate convex quadratic function in $x - y$, which means that the maximum is attained at one of the two bounds of the domain of $x - y$ over D, i.e. at either at $(\underline{x}, \bar{y})$ or at $(\bar{x}, \underline{y})$. Evaluating C_{31} at these two points yields

$$\begin{aligned} C_{31}(\underline{x}, \bar{y}) &= (\underline{x} - \bar{y})^2 - (\bar{x} + \underline{x} - \bar{y} - \underline{y})(\underline{x} - \bar{y}) + (\underline{x} - \bar{y})(\bar{x} - \underline{y}) \\ &= (\underline{x} - \bar{y})(\underline{x} - \bar{y} - \bar{x} - \underline{x} + \bar{y} + \underline{y} + \bar{x} - \underline{y}) \\ &= 0 \end{aligned}$$

and

$$\begin{aligned} C_{31}(\bar{x}, \underline{y}) &= (\bar{x} - \underline{y})^2 - (\bar{x} + \underline{x} - \bar{y} - \underline{y})(\bar{x} - \underline{y}) + (\underline{x} - \bar{y})(\bar{x} - \underline{y}) \\ &= (\bar{x} - \underline{y})(\bar{x} - \underline{y} - \bar{x} - \underline{x} + \bar{y} + \underline{y} + \underline{x} - \bar{y}) \\ &= 0 \end{aligned}$$

This means that the optimal objective value of Problem (13) is indeed 0. Now consider the point $(\underline{x}, \underline{y})$. We have

$$\begin{aligned} C_{31}(\underline{x}, \underline{y}) &= (\underline{x} - \underline{y})(\underline{x} - \underline{y} - \bar{x} - \underline{x} + \bar{y} + \underline{y}) + (\underline{x} - \bar{y})(\bar{x} - \underline{y}) \\ &= \underline{x}\bar{y} + \underline{y}\bar{x} - \underline{x}\underline{y} - \bar{y}\bar{x} = (\bar{x} - \underline{x})(\underline{y} - \bar{y}) \end{aligned}$$

$$< 0.$$

Thus, C_1^L is strictly tighter than C_3^L. □

The same results as above also holds with respect to C_2^L and C_3^L, belonging to Bin2 and Bin3 respectively.

Proposition 2 *The convex envelope $C_2^L : D \to \mathbb{R}$ resulting from the univariate reformulation* Bin2 *is a tighter convex underestimator of F over D than the convex envelope $C_3^L(x, y)$ resulting from the univariate reformulation* Bin3*, i.e. we have*

$$C_2^L(x, y) - C_3^L(x, y) \geq 0 \quad \forall (x, y) \in D,$$

and there exists a point $(x, y) \in D$ with

$$C_2^L(x, y) - C_3^L(x, y) > 0.$$

Proof Consider the optimization problem

$$\min_{(x,y)\in D} C_{23}(x, y), \tag{14}$$

with $C_{23} : D \to \mathbb{R}$ and

$$\begin{aligned} C_{23}(x, y) &:= 2(C_2^L(x, y) - C_3^L(x, y)) \\ &= 2xy - (\bar{y} + \underline{y})x - (\bar{x} + \underline{x})y + \underline{x}\,\underline{y} + \bar{x}\,\bar{y}. \end{aligned}$$

Problem (14) minimizes a bilinear function over a box. It is obvious that C_{23} is linear along both the x-axis and the y-axis, i.e. along the edges of the box. This means that C_{23} is edge-concave, and therefore the minimum of C_{23} over D is attained at one of the vertices $V_D = \{(\underline{x}, \bar{y}), (\bar{x}, \underline{y}), (\underline{x}, \underline{y}), (\bar{x}, \bar{y})\}$ of the box. By evaluation, we obtain:

$$\begin{aligned} C_{23}(\underline{x}, \bar{y}) &= 2\underline{x}\,\bar{y} - (\bar{y} + \underline{y})\underline{x} - (\bar{x} + \underline{x})\bar{y} + \underline{x}\,\underline{y} + \bar{x}\,\bar{y} \\ &= 2\underline{x}\,\bar{y} - \underline{x}\,\bar{y} - \underline{x}\,\underline{y} - \bar{x}\,\bar{y} - \underline{x}\,\bar{y} + \underline{x}\,\underline{y} + \bar{x}\,\bar{y} \\ &= 0, \\ C_{23}(\bar{x}, \underline{y}) &= 2\bar{x}\,\underline{y} - (\bar{y} + \underline{y})\bar{x} - (\bar{x} + \underline{x})\underline{y} + \underline{x}\,\underline{y} + \bar{x}\,\bar{y} \\ &= 2\bar{x}\,\underline{y} - \bar{x}\,\bar{y} - \bar{x}\,\underline{y} - \bar{x}\,\underline{y} - \underline{x}\,\underline{y} + \underline{x}\,\underline{y} + \bar{x}\,\bar{y} \\ &= 0 \end{aligned}$$

and

$$\begin{aligned} C_{23}(\underline{x}, \underline{y}) &= 2\underline{x}\,\underline{y} - (\bar{y} + \underline{y})\underline{x} - (\bar{x} + \underline{x})\underline{y} + \underline{x}\,\underline{y} + \bar{x}\,\bar{y} \\ &= \underline{x}\,\underline{y} - \underline{x}\,\bar{y} + \underline{x}\,\underline{y} + \bar{x}\,\bar{y} = (\bar{x} - \underline{x})(\bar{y} - \underline{y}) \\ &> 0, \\ C_{23}(\bar{x}, \bar{y}) &= 2\bar{x}\,\bar{y} - (\bar{y} + \underline{y})\bar{x} - (\bar{x} + \underline{x})\bar{y} + \underline{x}\,\underline{y} + \bar{x}\,\bar{y} \\ &= \bar{x}\,\bar{y} - \underline{x}\,\bar{y} - \bar{x}\,\underline{y} + \underline{x}\,\underline{y} = (\bar{x} - \underline{x})(\bar{y} - \underline{y}) \\ &> 0, \end{aligned}$$

which proves the claim. □

Between C_1^L and C_2^L, belonging to Bin1 and Bin2 respectively, C_2^L is the tighter convex underestimator; however, this only holds over square-shaped domains.

Proposition 3 *The convex envelope $C_2^L : D \to \mathbb{R}$ resulting from the univariate reformulation* Bin2 *is a tighter convex underestimator of F over D than the convex envelope $C_1^L(x, y)$ resulting from the univariate reformulation* Bin1 *if D is a square. In this case, we have*

$$C_2^L(x, y) - C_1^L(x, y) \geq 0 \quad \forall (x, y) \in D,$$

and there exists a point $(x, y) \in D$ with

$$C_2^L(x, y) - C_1^L(x, y) > 0.$$

Proof Consider the optimization problem

$$\min_{(x,y)\in D} C_{21}(x, y), \tag{15}$$

with $C_{21} : D \to \mathbb{R}$ and

$$C_{21} := 4(C_2^L - C_1^L) = (x + y)^2 - (\bar{x} + \underline{x} + \bar{y} + \underline{y})(x + y) + (\underline{x} + \bar{y})(\bar{x} + \underline{y}).$$

Since we assume that D is a square, we have $\bar{x} - \underline{x} = \bar{y} - \underline{y}$ and equivalently $\bar{x} + \underline{y} = \underline{x} + \bar{y}$. Therefore, we can simplify the minimization problem (15) to

$$\min_{(x,y)\in D} -2(\bar{x} + \underline{y})(x + y) + (\bar{x} + \underline{y})^2 + (x + y)^2.$$

This means that Problem (15) minimizes a convex quadratic univariate function in $x + y$. Using a first-order argument, the minimum is attained at a point $(x^*, y^*) \in D$ that fulfils $x^* + y^* = \bar{x} + \underline{y}$. It is straightforward to see that $C_{21}(x^*, y^*) = 0$, i.e. the minimum objective value of Problem (15) is 0.

Finally, we obtain

$$\begin{aligned} C_{21}(\underline{x}, \underline{y}) &= (\underline{x} + \underline{y})^2 - 2(\bar{x} + \underline{y})(\underline{x} + \underline{y}) + (\bar{x} + \underline{y})^2 \\ &= (\bar{x} - \underline{x})^2 + (2\underline{y})^2 \\ &> 0. \end{aligned}$$

In other words, there exists a point $(x, y) \in D$ with $C_2^L(x, y) - C_1^L(x, y) > 0$. □

References

1. Aigner, K.-M., Burlacu, R., Liers, F., Martin, A.: Solving AC optimal power flow with discrete decisions to global optimality (2020). http://www.optimization-online.org/DB_HTML/2020/08/7981.html
2. Anstreicher, K.M., Burer, S., Park, K.: Convex Hull representations for bounded products of variables (2020). https://arxiv.org/pdf/2004.07233.pdf
3. Appa, G.M., Pitsoulis, L., Williams, H.P.: Handbook on Modelling for Discrete Optimization, vol. 88, Springer (2006)
4. Atariah, D., Rote, G., Wintraecken, M.: Optimal triangulation of saddle surfaces. In: Beiträge zur algebra und geometrie/contributions to algebra and geometry **59**(1), 113–126 (2018)
5. Aurenhammer, F., Xu, Y.-F.: Optimal triangulations. In: Encyclopedia of Optimization, Springer, pp. 2757–2764 (2008)
6. Balakrishnan, A., Graves, S.C.: A composite algorithm for a concave-cost network flow problem. Networks **19**(2), 175–202 (1989)
7. Bärmann, A., Burlacu, R., Hager, L., Kutzer, K.: A p5/2-approximation algorithm for optimal piecewise linear approximations of bounded variable products (2022). https://optimization-online.org/2022/03/8831/
8. BÄrmann A., Martin, A., Schneider, O.: The bipartite boolean quadric polytope with multiple-choice constraints (2022). https://arxiv.org/abs/2009.11674

9. Beach, B., Hildebrand, R., Huchette, J.: Compact mixed-integer programming relaxations in quadratic optimization (2021). https://arxiv.org/pdf/2011.08823.pdf
10. Belotti, P., Kirches, C., Leyffer, S., Linderoth, J., Luedtke, J., Mahajan, A.: Mixed-integer nonlinear optimization. Acta Numer. **22**, 1–131 (2013)
11. BÖttger, T., Grimm, V., Kleinert, T., Schmidt, M.: The cost of decoupling trade and transport in the European entry-exit gas market with linear physics modeling. Eur. J. Oper. Res. **297**(3), 1095–1111 (2022). https://doi.org/10.1016/j.ejor.2021.06.034
12. Burlacu, R.: Adaptive Mixed-Integer Refinements for Solving Nonlinear Problems with Discrete Decisions. PhD Thesis (2020)
13. Burlacu, R., Geißler, B., Schewe, L.: Solving mixed-integer nonlinear programmes using adaptively refined mixed-integer linear programmes. Optim. Methods Softw. **35**(1), 37–64 (2020)
14. Correa-Posada, C.M., Sánchez-Martín, P.: Gas network optimization: a comparison of piecewise linear models. In: Optimization (2014)
15. Croxton, K.L., Gendron, B., Magnanti, T.L.: A comparison of mixed-integer programming models for nonconvex piecewise linear cost minimization problems. Manag. Sci. **49**(9), 1268–1273 (2003)
16. D'Ambrosio, C., Lodi, A., Martello, S.: Piecewise linear approximation of functions of two variables in MILP models. Oper. Res. Lett. **38**(1), 39–46 (2010)
17. Dantzig, G.B.: On the significance of solving linear programming problems with some integer variables. Econom. J. Econom. Soci. **28**, 30–44 (1960)
18. Egerer, J., Grimm, V., Kleinert, T., Schmidt, M., Zöttl, G.: The impact of neighboring markets on renewable locations, transmission expansion, and generation investment. Eur. J. Oper. Res. **292**(2), 696–713 (2021). https://doi.org/10.1016/j.ejor.2020.10.055
19. Falk, J.E.: Lagrange multipliers and nonconvex programs. SIAM J. Control **7**(4), 534–545 (1969)
20. Faria, D.C., Bagajewicz, M.J.: Novel bound contraction procedure for global optimization of bilinear MINLP problems with applications to water management problems. Comput. Chem. Eng. **35**(3), 446–455 (2011)
21. Fügenschuh, A., Hayn, C., Michaels, D.: Mixed-integer linear methods for layout-optimization of screening systems in recovered paper production. Optim. Eng. **15**(2), 533–573 (2014)
22. Geißler, B.: Towards Globally Optimal Solutions for MINLPs by Discretization Techniques with Applications in Gas Network Optimization. PhD Thesis (2011)
23. Geißler, B., Martin, A., Morsi, A., Schewe, L.: Using piecewise linear functions for solving minlps. In: Mixed Integer Nonlinear Programming, Springer, pp. 287–314 (2012)
24. Jeroslow, R.G.: Representability in mixed integer programming, I: characterization results. Discrete Appl. Math. **17**(3), 223–243 (1987)
25. Jeroslow, R.G.: Representability of functions. Discrete Appl. Math. **23**(2), 125–137 (1989)
26. Jeroslow, R.G., Lowe, J.K.: Experimental results on the new techniques for integer programming formulations. J. Oper. Res. Soc. **36**(5), 393–403 (1985)
27. Jeroslow, R.G., Lowe, J.K.: Modeling with integer variables. Math. Program. Study **22**, 167–184 (1984)
28. Knight, U.G.: Power Systems Engineering and Mathematics: International Series of Monographs in Electrical Engineering, vol. 3, Elsevier (2017)
29. Kutzer, K.: Using Piecewise Linear Approximation Techniques to Handle Bilinear Constraints. PhD Thesis (2020)
30. Markowitz, H.M., Manne, A.S.: On the solution of discrete programming problems. Econom. J. Econom. Soc. **25**, 84–110 (1957)
31. Martin, A., Möller, M., Moritz, S.: Mixed integer models for the stationary case of gas network optimization. Math. Program. **105**(2–3), 563–582 (2006)
32. McCormick, G.P.: Computability of global solutions to factorable nonconvex programs: part I—convex underestimating problems. Math. Program. **10**(1), 147–175 (1976)
33. Misener, R., Floudas, C.A.: Advances for the pooling problem: modeling, global optimization, and computational studies. Appl. Comput. Math. **8**(1), 3–22 (2009)
34. Monsky, P.: On dividing a square into triangles. Am. Math. Mon. **77**(2), 161–164 (1970)
35. Morsi, A.: Solving MINLPs on Loosely-Coupled Networks with Applications in Water and Gas Network Optimization. PhD Thesis (2013)
36. Morsi, A., Geißler, B., Martin, A.: Mixed integer optimization of water supply networks. In: Mathematical Optimization of Water Networks, vol. 162, Springer, pp. 35–54 (2012)
37. Mulzer, W., Rote, G.: Minimum-weight triangulation is NP-hard. J. ACM (JACM) **55**(2), 1–29 (2008)
38. Nowatzki, T., Ferris, M., Sankaralingam, K., Estan, C., Vaish, N., Wood, D.: Optimization and mathematical modeling in computer architecture. Synth. Lect. Comput. Archit. **8**(4), 1–144 (2013)
39. Padberg, M.: Approximating separable nonlinear functions via mixed zero-one programs. Oper. Res. Lett. **27**(1), 1–5 (2000)

40. Pottmann, H., Krasauskas, R., Hamann, B., Joy, K., Seibold, W.: On piecewise linear approximation of quadratic functions. J. Geom. Gr. **4**(1), 31–53 (2000)
41. Rebennack, S., Kallrath, J.: Continuous piecewise linear delta approximations for bivariate and multivariate functions. J. Optim. Theory Appl. **167**(1), 102–117 (2015)
42. Rebennack, S., Kallrath, J.: Continuous piecewise linear delta approximations for univariate functions: computing minimal breakpoint systems. J. Optim. Theory Appl. **167**(2), 617–643 (2015)
43. Rikun, A.D.: A convex envelope formula for multilinear functions. J. Global Optim. **10**(4), 425–437 (1997)
44. Sherali, H.D.: On mixed-integer zero-one representations for separable lower-semicontinuous piecewise-linear functions. Oper. Res. Lett. **28**(4), 155–160 (2001)
45. Tardella, F.: On the existence of polyhedral convex envelopes. In: Floudas, C.A., Pardalos, P. (eds.) Frontiers in Global Optimization, pp. 563–573. Springer, Boston (2004)
46. Vielma, J.P.: Mixed integer linear programming formulation techniques. SIAM Rev. **57**(1), 3–57 (2015)
47. Vielma, J.P., Ahmed, S., Nemhauser, G.: Mixed-integer models for nonseparable piecewise-linear optimization: unifying framework and extensions. Oper. Res. **58**(2), 303–315 (2010)
48. Vielma, J.P., Keha, A.B., Nemhauser, G.L.: Nonconvex, lower semicontinuous piecewise linear optimization. Discrete Optim. **5**(2), 467–488 (2008)
49. Wei, W., Wang, J.: Modeling and Optimization of Interdependent Energy Infrastructures, Springer (2019)
50. Zelmer, A.: Designing Coupled Energy Carrier Networks By Mixed-Integer Programming Methods. PhD Thesis (2010)